Stochastic calculus methods in machine learning

Samy Tindel

Purdue University

August 12, 2026

Contents

CHAPTER 1

Basics on Brownian motion

Brownian motion is the founding object of stochastic calculus, and it is also, in disguise, everywhere in modern machine learning: it is the noise source of stochastic gradient algorithms, the driving process of the stochastic differential equations that we use to model continuous-time dynamics, and, as we shall see in Chapter **??**, the building block of score-based generative models. Before any of this can be made precise we need a solid construction of the object itself, together with the vocabulary (processes, filtrations, modifications) needed to state its properties rigorously. This is the purpose of the present chapter.

We first recall, in Section 1.1, the two measure-theoretic notions, σ-algebra and probability measure, on which everything that follows rests. We then proceed in two further steps, corresponding to the next two sections. Section 1.2 introduces the general language of stochastic processes, indexed families of random variables, and the bookkeeping device of a filtration, which records the information available as time elapses. Section 1.3 then specializes this language to define the Wiener process, states its existence, and proves it through the classical Lévy–Ciesielski construction. Later sections of this chapter (first properties, the martingale and Markov properties, pathwise regularity) will build on this construction to develop the calculus needed for the rest of the course; our main reference throughout is the monograph of Karatzas and Shreve [6], and we occasionally lean on Durrett [2] for background probabilistic facts.

1.1 Basic measure theoretical background

Probability, as we use it throughout the course, is built on top of two measure-theoretic notions: the σ-algebra, which fixes which subsets of an underlying set Ω can be assigned

a probability, and the probability measure itself, which assigns these probabilities consistently. We recall both before putting them to use.

Definition 1.1.1

Let Ω be a set. A *σ-algebra* on Ω is a collection $\mathcal{F}$ of subsets of Ω, called *events*, such that:

(i) $\Omega \in \mathcal{F}$;

(ii) if $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$;

(iii) if $(A_k)_{k\geq 1} \subset \mathcal{F}$, then $\bigcup_{k\geq 1} A_k \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is then called a *measurable space*.

A σ-algebra alone only names the events; a probability measure is what makes them measurable in size.

Definition 1.1.2

Let $(\Omega, \mathcal{F})$ be a measurable space (Definition 1.1.1). A *probability measure* on $(\Omega, \mathcal{F})$ is a map $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that:

(i) $\mathbb{P}(\Omega) = 1$;

(ii) (σ-additivity) for every sequence $(A_k)_{k\geq 1}$ of pairwise disjoint sets in $\mathcal{F}$,

$$\mathbb{P}\Big(\bigcup_{k\geq 1} A_k\Big) = \sum_{k\geq 1} \mathbb{P}(A_k). \tag{1.1}$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is then called a *probability space*.

With these two definitions in hand, we fix once and for all the notation for the underlying probability space and related objects that will be used throughout the course.

Notation 1.1.3

Throughout the course $(\Omega, \mathcal{F}, \mathbb{P})$ denotes an underlying probability space (Definitions 1.1.1 and 1.1.2), and $\mathbb{E}$ the expectation with respect to $\mathbb{P}$. For two random variables (or events) X, Y we write $X \perp\!\!\!\perp Y$ to mean that X and Y are independent. For a real random variable $X \in L^2(\Omega)$ we write $\mathrm{Var}(X) := \mathbb{E}[X^2] - \mathbb{E}[X]^2$, and for $X, Y \in L^2(\Omega)$ we write $\mathrm{Cov}(X, Y) := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, so that $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$. For $s, t \in \mathbb{R}$ we write $s \wedge t := \min(s, t)$ and $s \vee t := \max(s, t)$.

1.2 Stochastic processes

We start by fixing the vocabulary of stochastic processes that will be used throughout the course: a process is simply a time-indexed family of random variables, and its *path* is

the function obtained by freezing the outcome $\omega \in \Omega$ and letting time run.

Definition 1.2.1

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (Notation 1.1.3), let $I \subset \mathbb{R}_+$ be an interval, and let $\{X_t\,;\, t \in I\}$ be a family of $\mathbb{R}^n$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) If the map $\omega \mapsto X_t(\omega)$ is measurable for every $t \in I$, the family $X = \{X_t\,;\, t \in I\}$ is called a *stochastic process*.

(ii) For a fixed $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called a *path* of X.

(iii) The process X is *continuous* if its paths are continuous for $\mathbb{P}$-almost every $\omega \in \Omega$.

Two processes built from the same family of random variables can still differ on how they agree with each other: agreement can hold at every fixed time, or simultaneously along every path. The next definition separates these two notions.

Definition 1.2.2

Let X and Y be two stochastic processes indexed by the same interval I, on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(i) We say that X is a *modification* of Y if

$$\mathbb{P}\Big(X_t = Y_t\Big) = 1, \qquad \text{for all} \quad t \in I. \tag{1.2}$$

(ii) We say that X and Y are *indistinguishable* if

$$\mathbb{P}\Big(X_t = Y_t \text{ for all } t \in I\Big) = 1. \tag{1.3}$$

Remark 1.2.3. (i) *Relation (1.3) implicitly requires the event*

$$\Big(X_t = Y_t \text{ for all } t \in I\Big)$$

to belong to $\mathcal{F}$, which is automatic when I is countable but is a genuine requirement (typically met through continuity of the paths) when I is an interval of $\mathbb{R}_+$.

(ii) *Indistinguishability (1.3) is a much stronger requirement than being a modification (see (1.2)): the latter only asks the two processes to agree, time by time, outside of a $\mathbb{P}$-negligible set that is allowed to depend on t, while the former asks for a single $\mathbb{P}$-negligible exceptional set that works simultaneously for every $t \in I$.*

(iii) *When X and Y are both continuous processes, and I admits a countable dense subset $D \subset I$ (for instance $I = \mathbb{R}_+$, with $D = \mathbb{Q}_+$), relations (1.3) and (1.2) are equivalent: two continuous processes that agree almost surely at every $t \in D$ agree, outside a single negligible set, on the whole of I.*

In order to formalize the flow of information carried by a stochastic process, we now introduce filtrations. This device will be indispensable in the martingale and Markov analysis of Brownian motion carried out later in the chapter.

Definition 1.2.4

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $I \subset \mathbb{R}_+$ be an interval.

(i) A *filtration* on $(\Omega, \mathcal{F}, \mathbb{P})$ is an increasing family $(\mathcal{F}_t)_{t\in I}$ of sub-σ-algebras of $\mathcal{F}$, that is $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $s < t$ in I. We interpret $\mathcal{F}_t$ as the information available to an observer at time t.

(ii) The collection $\mathcal{N} := \{F \in \mathcal{F}\,;\, \mathbb{P}(F) = 0\}$ of *negligible sets* is fixed once and for all. A filtration $(\mathcal{F}_t)_{t\in I}$ is *complete* if $\mathcal{N} \subset \mathcal{F}_t$ for every $t \in I$.

(iii) A *stochastic basis* is a quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$ in which $(\mathcal{F}_t)_{t\in I}$ is a complete filtration.

Remark 1.2.5. *Any filtration $(\mathcal{F}_t)_{t\in I}$ can be completed at no cost: replacing $\mathcal{F}_t$ by $\widehat{\mathcal{F}}_t := \sigma(\mathcal{F}_t, \mathcal{N})$ produces a complete filtration $(\widehat{\mathcal{F}}_t)_{t\in I}$ containing the original one. In what follows we always work with a completed filtration, and drop the hat from the notation.*

The last notion we need measures how a process interacts with a given filtration: a process is adapted to $(\mathcal{F}_t)_{t\in I}$ when its value at time t is determined by the information available at that time.

Definition 1.2.6

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$ be a stochastic basis (Definition 1.2.4) and let $X = \{X_t\,;\, t \in I\}$ be an $\mathbb{R}^n$-valued stochastic process (Definition 1.2.1). We say that X is *$\mathcal{F}_t$-adapted* if, for every $t \in I$, the map

$$X_t \,:\, (\Omega, \mathcal{F}_t) \longrightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \tag{1.4}$$

is measurable.

Remark 1.2.7. *Let $\mathcal{F}_t^X := \sigma\{X_s\,;\, s \le t\}$ denote the* natural filtration *generated by a process X, that is the smallest filtration to which X is adapted.*

(i) *Any process X is adapted to its own natural filtration $(\mathcal{F}_t^X)_{t\in I}$.*

(ii) *More generally, X is $\mathcal{F}_t$-adapted, in the sense of Definition 1.2.6, if and only if $\mathcal{F}_t^X \subset \mathcal{F}_t$ for every $t \in I$: a process is adapted to a filtration exactly when that filtration carries at least as much information, at each time, as the process itself generates.*

With this vocabulary in hand we can now turn to the central object of the chapter: the Wiener process.

1.3 Definition and construction of the Wiener process

This section has two goals. We first give the defining properties of the Wiener process (Definition 1.3.2) and state its existence (Theorem 1.3.4). We then prove existence through the classical Lévy-Ciesielski construction, which expands the process along an orthonormal system of $L^2([0,1])$ built from the Haar functions. Beyond its historical importance, this construction has the merit of making the almost-sure continuity of the paths, part (iv) of Definition 1.3.2, transparent: it is inherited from the uniform convergence of the defining series.

Throughout, for a function $f : I \to \mathbb{R}^n$ and $s, t \in I$ we abbreviate the increment of f between s and t as follows.

Notation 1.3.1

For a function f defined on an interval $I \subset \mathbb{R}_+$ and $s, t \in I$, we write

$$\delta f_{st} := f_t - f_s. \tag{1.5}$$

The same notation is used for a stochastic process $f = X$, in which case δX_{st} is itself a random variable.

Definition 1.3.2

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{W_t\,;\, t \geq 0\}$ be an $\mathbb{R}$-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ (Definition 1.2.1). We say that W is a (standard, real-valued) *Wiener process*, or *Brownian motion*, if

(i) $W_0 = 0$ almost surely.

(ii) For every $n \geq 1$ and every $0 = t_0 < t_1 < \cdots < t_n$, the increments (Notation 1.3.1)

$$\delta W_{t_0 t_1},\ \delta W_{t_1 t_2},\ \ldots,\ \delta W_{t_{n-1} t_n} \tag{1.6}$$

are independent.

(iii) For every $0 \leq s < t$,

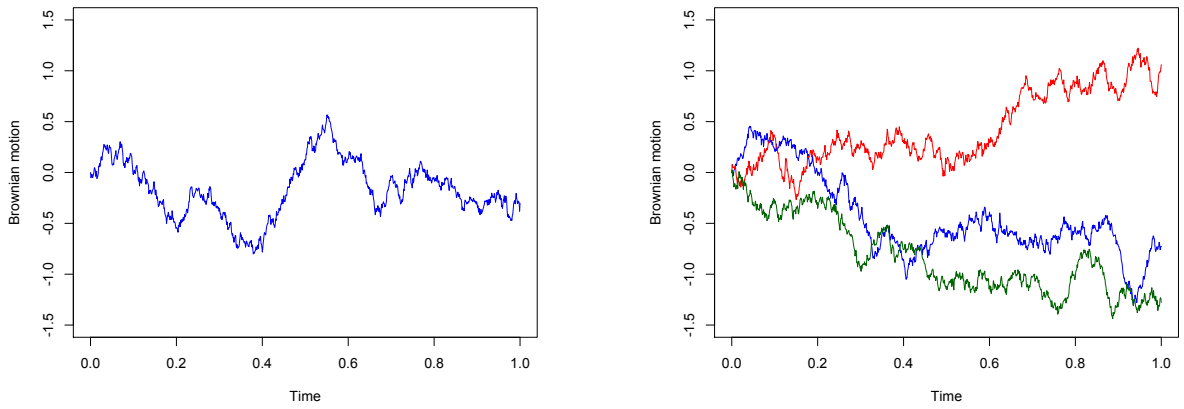
$$\delta W_{st} \sim \mathcal{N}(0,\, t - s). \tag{1.7}$$

(iv) W has continuous paths almost surely.

Remark 1.3.3. *Properties (i)–(iii) alone only prescribe the finite-dimensional distributions of W, that is the law of $(W_{t_1}, \ldots, W_{t_n})$ for every finite set of times: by Kolmogorov's extension theorem, they determine a consistent family of such laws, hence a process on some probability space, but they say nothing about the regularity of its paths. Many references state continuity of the paths as a separate theorem (an existence-of-a-continuous-modification result, typically proved through Kolmogorov's continuity criterion,*

as in route (i) of Remark 1.3.5) rather than as part of the definition of a Wiener process. We include it as property (iv) directly in Definition 1.3.2, purely for convenience: the construction of Section 1.3.1 (Proposition 1.3.11) produces a process satisfying (i)–(iv) simultaneously, with no extra argument needed, so nothing is lost by assuming continuity from the outset. Throughout the course, "Wiener process" always refers to this continuous version.

Properties (ii) and (iii) together say that W has *stationary independent Gaussian increments*: the law of an increment only depends on the length of the time interval, and increments over disjoint intervals carry no information about one another. Figures 1.1a and 1.1b display two simulated paths of a Wiener process: their irregularity is already a visible sign that, as we shall see in Section 1.7 later in this chapter, the paths are almost surely nowhere differentiable.



(a) A single simulated path. **(b)** Three independent simulated paths.

Figure 1.1 – Simulated paths of a Wiener process on $[0, 1]$, illustrating the randomness of the trajectories.

Definition 1.3.2 describes the properties we want W to have, but it gives no reason to believe that such a process exists: it is not a priori clear that a single probability space can support an uncountable family of Gaussian random variables satisfying simultaneously the independence and covariance constraints (ii)–(iii), together with the path regularity of (iv). This existence question is settled by the following theorem.

Theorem 1.3.4

There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which one can construct a Wiener process, in the sense of Definition 1.3.2.

Remark 1.3.5. *Theorem 1.3.4 can be proved in several classical ways, among which*

(i) *an application of Kolmogorov's extension theorem to the family of finite-dimensional Gaussian distributions prescribed by* (1.6)–(1.7)*, followed by a separate argument (Kolmogorov's continuity criterion) to upgrade the resulting process to one with continuous paths;*

(ii) *an invariance-principle argument, obtaining W as the limit, after diffusive rescaling, of a simple random walk;*

(iii) *the* Lévy–Ciesielski construction*, which builds W directly as a random series expansion along an orthonormal basis of $L^2([0,1])$.*

We follow route (iii), which we now develop in full.

1.3.1 The Lévy–Ciesielski construction on $[0,1]$

The construction proceeds by expanding a Wiener process on $[0,1]$ along a well-chosen orthonormal basis of $L^2([0,1])$, the *Haar functions*.

Definition 1.3.6

The family of *Haar functions* $\{h_k : [0,1] \to \mathbb{R}\,;\, k \geq 0\}$ is defined by $h_0 \equiv 1$,

$$h_1(t) \;=\; \mathbf{1}_{[0,1/2]}(t) - \mathbf{1}_{(1/2,1]}(t), \tag{1.8}$$

and, for $n \geq 1$ and $2^n \leq k < 2^{n+1}$,

$$h_k(t) \;=\; 2^{n/2}\,\mathbf{1}_{\left[\frac{k-2^n}{2^n},\,\frac{k-2^n+1/2}{2^n}\right]}(t) \;-\; 2^{n/2}\,\mathbf{1}_{\left(\frac{k-2^n+1/2}{2^n},\,\frac{k-2^n+1}{2^n}\right]}(t). \tag{1.9}$$

Figure 1.2 displays the first Haar functions $h_1, \ldots, h_7$: each h_k, for $2^n \leq k < 2^{n+1}$, is a rescaled step function supported on a single dyadic interval of length 2^{-n}, taking the value $2^{n/2}$ on its left half and $-2^{n/2}$ on its right half.

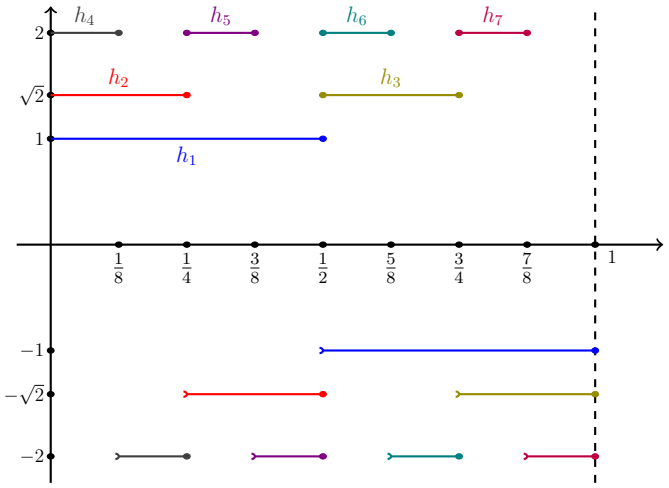


Figure 1.2 – The Haar functions $h_1, \ldots, h_7$ (Definition 1.3.6) on $[0,1]$.

Lemma 1.3.7

The Haar functions $\{h_k : [0,1] \to \mathbb{R}\,;\, k \geq 0\}$ of Definition 1.3.6 form an orthonormal basis of $L^2([0,1])$.

Proof. We check successively that the family is normalized, that it is orthogonal, and that it is complete in $L^2([0,1])$.

Step 1: Norm. For $2^n \leq k < 2^{n+1}$, using definition (1.9),

$$\int_0^1 h_k^2(t)\,dt \;=\; 2^n \int_{\frac{k-2^n}{2^n}}^{\frac{k-2^n+1}{2^n}} dt \;=\; 1, \tag{1.10}$$

and likewise $\int_0^1 h_0^2 = \int_0^1 h_1^2 = 1$. Hence every h_k has unit $L^2([0,1])$-norm.

Step 2: Orthogonality. Let $k < l$, both in the range $k \geq 1$, and write $\mathrm{supp}(h_k)$ for the support of h_k. Two situations occur.

(i) If $\mathrm{supp}(h_k) \cap \mathrm{supp}(h_l) = \varnothing$, then trivially $\langle h_k, h_l\rangle_{L^2([0,1])} = 0$.

(ii) If instead $\mathrm{supp}(h_l) \subset \mathrm{supp}(h_k)$, then observe that h_k is constant equal to $\pm 2^{n/2}$ (where $2^n \leq k < 2^{n+1}$) on each half of its support. Moreover, h_l has mean zero. Therefore we get

$$\langle h_k, h_l\rangle_{L^2([0,1])} \;=\; \pm 2^{n/2} \int_0^1 h_l(t)\,dt \;=\; 0. \tag{1.11}$$

Since two distinct dyadic supports are always either disjoint or nested, this covers every pair $k \neq l$, and $\{h_k\,;\, k \geq 0\}$ is an orthonormal system.

Step 3: Completeness. It remains to show that the orthonormal system $\{h_k\,;\, k \geq 0\}$ spans a dense subspace of $L^2([0,1])$. Let $f \in L^2([0,1])$ be such that $\langle f, h_k\rangle = 0$ for every $k \geq 0$; we show that $f = 0$ almost everywhere. Unwinding the relations $\langle f, h_k\rangle = 0$ level by level in n shows that

$$\int_s^t f(u)\,du \;=\; 0, \qquad \text{for every pair of dyadic numbers } s,t \in [0,1]. \tag{1.12}$$

By density of the dyadic numbers and continuity of $t \mapsto \int_0^t f$, relation (1.12) in fact holds for every $s,t \in [0,1]$. In particular $F(t) := \int_0^t f(u)\,du$ vanishes identically, so by Lebesgue's differentiation theorem

$$f(t) \;=\; \partial_t F(t) \;=\; 0, \qquad \text{for almost every } t \in [0,1].$$

Hence the only element of $L^2([0,1])$ orthogonal to every h_k is 0, which proves completeness and finishes the proof. ∎

The Wiener process itself will not be expanded along the Haar functions directly, but along their primitives, called Schauder functions.

Definition 1.3.8

The family of *Schauder functions* $\{s_k : [0,1] \to \mathbb{R}\,;\, k \geq 0\}$ is defined from the Haar functions of Definition 1.3.6 by

$$s_k(t) := \int_0^t h_k(u)\, du. \tag{1.13}$$

Lemma 1.3.9

For $2^n \leq k < 2^{n+1}$, the Schauder function s_k of Definition 1.3.8 satisfies

$$\operatorname{supp}(s_k) = \operatorname{supp}(h_k) = \left[\frac{k-2^n}{2^n}, \frac{k-2^n+1}{2^n}\right], \tag{1.14}$$

and

$$\|s_k\|_\infty = \frac{1}{2^{n/2+1}}. \tag{1.15}$$

Figure 1.3 shows the first three Schauder functions: each s_k is a triangular tent supported on the same dyadic interval as h_k, with height controlled by (1.15); this decay of the sup-norm as $k \to \infty$ is exactly what will make the random series defining the Wiener process converge uniformly.

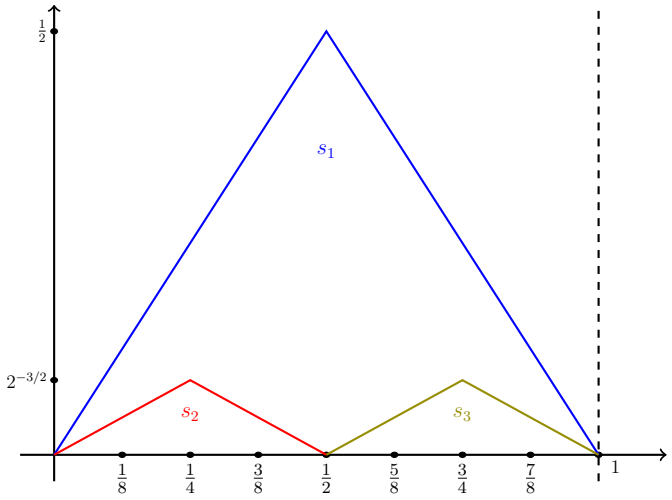


Figure 1.3 – The Schauder functions s_1, s_2, s_3 (Definition 1.3.8) on $[0,1]$.

We record a classical almost-sure growth bound for the maximum of n i.i.d. standard Gaussian variables, which will control the coefficients of the random series defining the Wiener process.

Lemma 1.3.10

Let $\{X_k\,;\,k\geq 1\}$ be an i.i.d. sequence of $\mathcal{N}(0,1)$ random variables and set $M_n := \sup\{|X_k|\,;\,1\leq k\leq n\}$. Then

$$M_n \;=\; \mathrm{O}\Big(\sqrt{\ln(n)}\Big), \qquad \text{almost surely.} \tag{1.16}$$

Proof. We first bound the Gaussian tail, then invoke Borel–Cantelli to control M_n almost surely.

Step 1: Gaussian tail. For $x>0$,

$$\mathbb{P}\big(|X_k|>x\big) \;=\; \frac{2}{\sqrt{2\pi}}\int_x^\infty e^{-z^2/2}\,dz \;\leq\; c_1\,e^{-x^2/4}\int_x^\infty e^{-z^2/4}\,dz \;\leq\; c_2\,e^{-x^2/4}, \tag{1.17}$$

where the middle inequality splits $e^{-z^2/2}=e^{-z^2/4}e^{-z^2/4}$ and bounds the first factor by $e^{-x^2/4}$ on $\{z\geq x\}$.

Step 2: Borel–Cantelli. Let

$$A_k := \Big(|X_k|\geq 4\sqrt{\ln(k)}\Big).$$

By (1.17) applied at $x=4\sqrt{\ln(k)}$, it is readily checked that

$$\mathbb{P}(A_k)\;\leq\;\frac{c}{k^4}, \qquad \text{so that} \qquad \sum_{k\geq 1}\mathbb{P}(A_k)\;<\;\infty. \tag{1.18}$$

By the (first) Borel–Cantelli lemma, we thus get $\mathbb{P}(\limsup_k A_k)=0$.

Step 3: Conclusion. By Step 2, almost surely there exists $k_0=k_0(\omega)$ such that $|X_k(\omega)|\leq 4\sqrt{\ln(k)}$ for all $k\geq k_0$. Taking the supremum over $1\leq k\leq n$ and letting $n\to\infty$ yields (1.16). ∎

We are now ready to construct the Wiener process on $[0,1]$.

Proposition 1.3.11

Let $\{s_k\,;\,k\geq 0\}$ be the Schauder functions of Definition 1.3.8 and let $\{X_k\,;\,k\geq 0\}$ be an i.i.d. sequence of $\mathcal{N}(0,1)$ random variables on some probability space $(\Omega,\mathcal{F},\mathbb{P})$. Set

$$W_t \;:=\; \sum_{k\geq 0} X_k\,s_k(t), \qquad t\in[0,1]. \tag{1.19}$$

Then $W=\{W_t\,;\,t\in[0,1]\}$ is a Wiener process on $[0,1]$, in the sense of Definition 1.3.2 restricted to $[0,1]$.

Proof. Property (i) of Definition 1.3.2 is immediate since every s_k satisfies $s_k(0) = 0$. We prove the remaining three properties in three steps.

Step 1: Uniform convergence, and continuity of the paths. We show that the series (1.19) converges uniformly on $[0,1]$, almost surely; this simultaneously makes W_t well defined and gives property (iv) of Definition 1.3.2, since a uniform limit of continuous functions is continuous. It suffices to show that, almost surely, for every $\varepsilon > 0$ there is $n_0 = n_0(\omega)$ such that

$$\Big\| \sum_{k=2^m}^{2^n-1} X_k \, s_k \Big\|_\infty \le \varepsilon, \qquad \text{for all } n_0 \le m < n. \tag{1.20}$$

Fix $\eta \in (0, 1/2)$. By Lemma 1.3.10 applied to the sequence $\{X_k \,;\, k \ge 0\}$, almost surely there is $c = c(\omega)$ such that $|X_k| \le c\, k^\eta$ for every $k \ge 1$. On the other hand, for $2^p \le k < 2^{p+1}$ the functions s_k have pairwise disjoint support by (1.14), so (1.15) gives

$$\Big\| \sum_{k=2^p}^{2^{p+1}-1} s_k \Big\|_\infty \le \frac{1}{2^{p/2+1}}. \tag{1.21}$$

Combining the two bounds, for every $t \in [0,1]$,

$$\begin{aligned} \Big| \sum_{k=2^m}^{2^n} X_k \, s_k(t) \Big| &\le \sum_{p \ge m} \sum_{k=2^p}^{2^{p+1}-1} |X_k| \, s_k(t) \\ &\le \sum_{p \ge m} \Big(\sup_{2^p \le \ell < 2^{p+1}} |X_\ell| \Big) \Big| \sum_{k=2^p}^{2^{p+1}-1} s_k(t) \Big| \\ &\le c_1 \sum_{p \ge m} \frac{1}{2^{p(1/2-\eta)}} \le \frac{c_2}{2^{m(1/2-\eta)}}, \end{aligned} \tag{1.22}$$

using (1.21) and $\eta < 1/2$ for the last two inequalities. The right-hand side of (1.22) tends to 0 as $m \to \infty$ and does not depend on t, which proves (1.20) and hence the almost sure uniform convergence, on $[0,1]$, of the series (1.19) defining W. Being an almost sure uniform limit of continuous functions, W then has continuous paths, which is property (iv) of Definition 1.3.2.

Step 2: Law of the increments. We show that $\delta W_{rt} \sim \mathcal{N}(0, t-r)$ for $0 \le r < t \le 1$ (Notation 1.3.1), that is property (iii) of Definition 1.3.2. It suffices to show that, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E}\Big[e^{\imath \lambda \, \delta W_{rt}} \Big] = e^{-\frac{(t-r)\lambda^2}{2}}. \tag{1.23}$$

By (1.19), $\delta W_{rt} = \sum_{k \ge 0} X_k \Big(s_k(t) - s_k(r) \Big)$, so, invoking independence of the X_k's and dominated convergence (justified by Step 1) to exchange expectation and the infinite sum,

$$\begin{aligned} \mathbb{E}\Big[e^{\imath \lambda \, \delta W_{rt}} \Big] &= \prod_{k \ge 0} \mathbb{E}\Big[e^{\imath \lambda \, X_k (s_k(t) - s_k(r))} \Big] \\ &= \prod_{k \ge 0} e^{-\frac{\lambda^2}{2}(s_k(t)-s_k(r))^2} = e^{-\frac{\lambda^2}{2} \sum_{k \ge 0} (s_k(t)-s_k(r))^2}, \end{aligned} \tag{1.24}$$

the second equality being the characteristic function of the $\mathcal{N}(0,(s_k(t)-s_k(r))^2)$ law of $X_k(s_k(t)-s_k(r))$. It remains to evaluate the sum in the exponent of (1.24). By definition (1.13) of the Schauder functions as primitives of the Haar functions, $s_k(u) = \langle h_k, \mathbf{1}_{[0,u]}\rangle_{L^2([0,1])}$ for every $u \in [0,1]$, so Parseval's identity for the orthonormal basis $\{h_k\,;\,k\geq 0\}$ of Lemma 1.3.7 gives

$$\sum_{k\geq 0} s_k(r)\,s_k(t) \;=\; \sum_{k\geq 0}\langle h_k,\mathbf{1}_{[0,r]}\rangle\,\langle h_k,\mathbf{1}_{[0,t]}\rangle \;=\; \langle \mathbf{1}_{[0,r]},\mathbf{1}_{[0,t]}\rangle_{L^2([0,1])} \;=\; r, \tag{1.25}$$

since $r<t$. Combined with $\sum_{k\geq 0} s_k(u)^2 = u$ (the case $r=0$ of (1.25) together with $t=u$), this yields

$$\sum_{k\geq 0}\Big(s_k(t)-s_k(r)\Big)^2 \;=\; \sum_{k\geq 0} s_k(t)^2 - 2\sum_{k\geq 0} s_k(t)s_k(r) + \sum_{k\geq 0} s_k(r)^2 \;=\; t-2r+r \;=\; t-r. \tag{1.26}$$

Substituting (1.26) into (1.24) gives exactly (1.23), which proves Step 2.

Step 3: Independence of the increments. We finally prove property (ii) of Definition 1.3.2. We only detail the simplest case of two increments, $W_r \perp\!\!\!\perp \delta W_{rt}$ for $0\leq r<t\leq 1$; the general case of $n\geq 2$ increments over a partition $0=t_0<t_1<\cdots<t_n$ follows by the identical computation, with n Gaussian variables and n real parameters $\lambda_1,\dots,\lambda_n$ in place of λ_1,λ_2 below. For $\lambda_1,\lambda_2\in\mathbb{R}$, arguing exactly as in Step 2, we discover that

$$\begin{aligned}\mathbb{E}\Big[e^{\imath(\lambda_1 W_r+\lambda_2\,\delta W_{rt})}\Big] \;&=\; \prod_{k\geq 0}\mathbb{E}\Big[e^{\imath X_k[\lambda_1 s_k(r)+\lambda_2(s_k(t)-s_k(r))]}\Big]\\ &=\; e^{-\frac12\sum_{k\geq 0}[\lambda_1 s_k(r)+\lambda_2(s_k(t)-s_k(r))]^2}.\end{aligned} \tag{1.27}$$

Furthermore, expanding the square term by term for every $k\geq 0$, we get

$$\begin{aligned}\Big[\lambda_1 s_k(r)+\lambda_2\Big(s_k(t)-s_k(r)\Big)\Big]^2 \;=\; \lambda_1^2\,s_k(r)^2 \;+\; 2\lambda_1\lambda_2\,s_k(r)\Big(s_k(t)-s_k(r)\Big)\\ +\;\lambda_2^2\,\Big(s_k(t)-s_k(r)\Big)^2.\end{aligned} \tag{1.28}$$

Summing (1.28) over $k\geq 0$, the cross term vanishes:

$$\sum_{k\geq 0} s_k(r)\Big(s_k(t)-s_k(r)\Big) \;=\; \sum_{k\geq 0} s_k(r)s_k(t) \;-\; \sum_{k\geq 0} s_k(r)^2 \;=\; r-r \;=\; 0, \tag{1.29}$$

where we used (1.25) for the second identity. Summing (1.28) over $k\geq 0$ and substituting (1.29) for the cross term, the identity $\sum_{k\geq 0} s_k(r)^2 = r$ for the λ_1^2 term, and (1.26) for the λ_2^2 term then gives

$$\sum_{k\geq 0}\Big[\lambda_1 s_k(r)+\lambda_2\Big(s_k(t)-s_k(r)\Big)\Big]^2 \;=\; \lambda_1^2\sum_{k\geq 0} s_k(r)^2+\lambda_2^2\sum_{k\geq 0}\Big(s_k(t)-s_k(r)\Big)^2 \;=\; \lambda_1^2\,r+\lambda_2^2\,(t-r). \tag{1.30}$$

Substituting (1.30) into (1.27), we end up with

$$\mathbb{E}\Big[e^{\imath(\lambda_1 W_r+\lambda_2\,\delta W_{rt})}\Big] \;=\; e^{-\frac12[\lambda_1^2 r+\lambda_2^2(t-r)]} \;=\; \mathbb{E}\Big[e^{\imath\lambda_1 W_r}\Big]\,\mathbb{E}\Big[e^{\imath\lambda_2\,\delta W_{rt}}\Big], \tag{1.31}$$

Since the joint characteristic function of $(W_r,\delta W_{rt})$ factors as the product of the marginal characteristic functions, we have obtained that W_r and δW_{rt} are independent. This proves Step 3 and concludes the proof of the proposition. ∎

1.3.2 Extension to $\mathbb{R}_+$

Proposition 1.3.11 produces a Wiener process on the unit interval $[0,1]$. To extend the construction to all of $\mathbb{R}_+$, we glue together countably many independent copies of this construction, one per unit interval $[n, n+1]$.

Proposition 1.3.12

For $k \geq 1$, let $(\Omega_k, \mathcal{F}_k, \mathbb{P}_k)$ be a probability space carrying a Wiener process W^k on $[0,1]$, as in Proposition 1.3.11, and assume the W^k's are mutually independent. Let

$$\overline{\Omega} = \prod_{k\geq 1} \Omega_k, \qquad \overline{\mathcal{F}} = \bigotimes_{k\geq 1} \mathcal{F}_k, \qquad \overline{\mathbb{P}} = \bigotimes_{k\geq 1} \mathbb{P}_k. \tag{1.32}$$

Define $W_0 := 0$ and, recursively,

$$W_t := W_n + W^{n+1}_{t-n}, \qquad \text{for } t \in [n, n+1],\ n \geq 0. \tag{1.33}$$

Then $W = \{W_t\,;\, t \geq 0\}$ is a Wiener process on $\mathbb{R}_+$, in the sense of Definition 1.3.2, defined on $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$.

Partial proof. Continuity of the paths, property (iv) of Definition 1.3.2, and independence of increments, property (ii), follow by patching together the corresponding properties of each W^k, established in Steps 1 and 3 of the proof of Proposition 1.3.11: two increments lying in different unit intervals $[m, m+1]$ and $[n, n+1]$, $m \neq n$, are independent because they are built from independent copies W^{m+1} and W^{n+1}, while two increments inside the same unit interval inherit independence from Proposition 1.3.11 itself.

We detail instead property (iii), the law of the increments, which is the only point where the recursive definition (1.33) requires a short computation. Fix $0 \leq s < t$ and integers $m \leq n$ with $m \leq s < m+1 \leq n \leq t < n+1$ (the case $m = n$, that is s, t in the same unit interval, is already covered by Proposition 1.3.11). By (1.33),

$$\delta W_{st} = \Big(\sum_{k=1}^{n} W_1^k + W^{n+1}_{t-n}\Big) - \Big(\sum_{k=1}^{m} W_1^k + W^{m+1}_{s-m}\Big) = Z_1 + Z_2 + Z_3, \tag{1.34}$$

with

$$Z_1 = \sum_{k=m+2}^{n} W_1^k, \qquad Z_2 = W_1^{m+1} - W^{m+1}_{s-m}, \qquad Z_3 = W^{n+1}_{t-n}. \tag{1.35}$$

The three variables Z_1, Z_2, Z_3 are built from disjoint groups of the mutually independent processes W^k, hence they are independent, and each is centered Gaussian: Z_1 as a sum of independent $\mathcal{N}(0,1)$ increments $W_1^k = \delta W_{01}^k$, and Z_2, Z_3 as single increments of a Wiener process on $[0,1]$ (Proposition 1.3.11). Their variances are

$$\mathrm{Var}(Z_1) = n - (m+1), \qquad \mathrm{Var}(Z_2) = 1 - (s-m), \qquad \mathrm{Var}(Z_3) = t - n, \tag{1.36}$$

the first from the sum of $n-(m+1)$ unit variances in Z_1, and the other two from the law (1.7) of a single increment. Consequently $\delta W_{st} = Z_1 + Z_2 + Z_3 \sim \mathcal{N}(0, \sigma^2)$, with, by independence and (1.36),

$$\sigma^2 \;=\; \mathrm{Var}(Z_1)+\mathrm{Var}(Z_2)+\mathrm{Var}(Z_3) \;=\; \Big(n-(m+1)\Big)+\Big(1-(s-m)\Big)+(t-n) \;=\; t-s, \quad (1.37)$$

which is property (iii) of Definition 1.3.2. ■

1.3.3 The multivariate Wiener process

Combining Proposition 1.3.12 with independent copies in each coordinate produces, without further work, a multi-dimensional version of the Wiener process.

Definition 1.3.13

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{W_t\,;\, t \geq 0\}$ be an $\mathbb{R}^n$-valued stochastic process. We say that W is a (standard) *Wiener process on* $\mathbb{R}^n$ if

(i) $W_0 = 0$ almost surely;

(ii) for every $k \geq 1$ and $0 = t_0 < t_1 < \cdots < t_k$, the increments $\delta W_{t_0 t_1}, \ldots, \delta W_{t_{k-1} t_k}$ are independent;

(iii) for every $0 \leq s < t$, $\delta W_{st} \sim \mathcal{N}(0, (t-s)\,\mathrm{Id}_{\mathbb{R}^n})$;

(iv) W has continuous paths almost surely.

Remark 1.3.14. *Definition 1.3.13 reduces entirely to the scalar case: if $W^{(1)}, \ldots, W^{(n)}$ are independent real-valued Wiener processes, as constructed in Proposition 1.3.12, then $W := (W^{(1)}, \ldots, W^{(n)})$ is a Wiener process on $\mathbb{R}^n$ in the sense of Definition 1.3.13, and conversely every $\mathbb{R}^n$-valued Wiener process arises this way from its coordinate processes.*

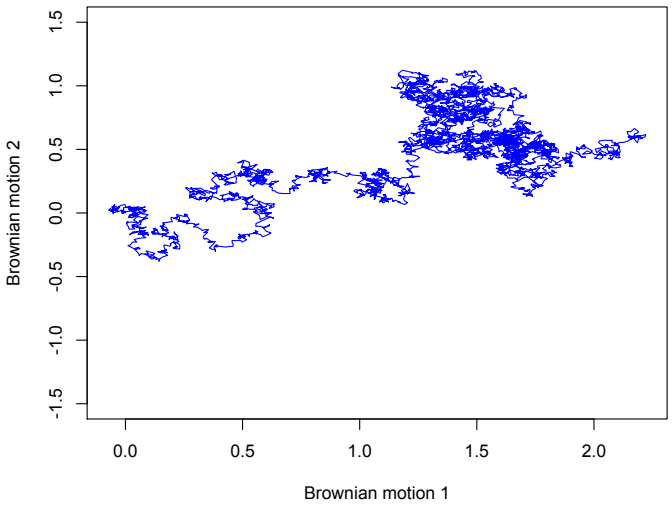


Figure 1.4 – A simulated path of a Wiener process on $\mathbb{R}^2$ (Definition 1.3.13).

We close the section by relativizing Definition 1.3.13 to an ambient filtration, which is the version of the Wiener process most directly compatible with the stochastic calculus developed later in the chapter: independence of increments, formulated with respect to the filtration rather than only with respect to the process itself, is what will let us build stochastic integrals against W.

Definition 1.3.15

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis (Definition 1.2.4) and let $\{W_t\,;\, t \geq 0\}$ be an $\mathbb{R}^n$-valued stochastic process on it. We say that W is a *Wiener process with respect to* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ if

(i) $W_0 = 0$ almost surely;

(ii) for every $0 \leq s < t$, $\delta W_{st} \perp\!\!\!\perp \mathcal{F}_s$;

(iii) for every $0 \leq s < t$, $\delta W_{st} \sim \mathcal{N}(0, (t-s)\,\mathrm{Id}_{\mathbb{R}^n})$;

(iv) W has continuous paths almost surely;

(v) W is $\mathcal{F}_t$-adapted (Definition 1.2.6).

Remark 1.3.16. *Condition (ii) of Definition 1.3.15 involves the independence $\delta W_{st} \perp\!\!\!\perp \mathcal{F}_s$ of a random vector from a σ-algebra, which extends the independence of two random variables recalled in Notation 1.1.3. Precisely, for a random vector X on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^n$ and a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, we write $X \perp\!\!\!\perp \mathcal{G}$ to mean that*

$$\mathbb{P}(X \in B,\, A) \,=\, \mathbb{P}(X \in B)\,\mathbb{P}(A), \qquad \text{for every Borel set } B \subset \mathbb{R}^n \text{ and every } A \in \mathcal{G}. \tag{1.38}$$

In other words, every event described by X is independent, in the sense of Notation 1.1.3, of every event in $\mathcal{G}$. Equivalently, denoting by $\sigma(X) \subset \mathcal{F}$ the σ-algebra generated by X (the smallest σ-algebra on Ω that makes X measurable), condition (1.38) says that the two σ-algebras $\sigma(X)$ and $\mathcal{G}$ are independent, that is

$$\mathbb{P}(A' \cap A) = \mathbb{P}(A')\,\mathbb{P}(A),$$

for every $A' \in \sigma(X)$ and $A \in \mathcal{G}$. In condition (ii) of Definition 1.3.15 this is applied with $X = \delta W_{st}$ and $\mathcal{G} = \mathcal{F}_s$: the increment δW_{st} is required to be independent of every event that is $\mathcal{F}_s$-measurable, that is of everything observable up to time s, and not merely of the past values of W itself.

Remark 1.3.17. *A Wiener process W in the sense of Definition 1.3.13 is automatically a Wiener process, in the sense of Definition 1.3.15, with respect to its own (completed) natural filtration $(\mathcal{F}_t^W)_{t\geq 0}$ (Remark 1.2.7): condition (ii) of Definition 1.3.15 strengthens condition (ii) of Definition 1.3.13 from independence of the increment δW_{st} from the past increments of W alone to independence from the full information $\mathcal{F}_s$ available at time s, and the two coincide precisely when $\mathcal{F}_s$ carries no more information than W itself does up to time s.*

We end this section with a word on the origin of the name Brownian motion. The Scottish botanist Robert Brown (1773–1858), an early and skilled user of the microscope who is also credited with the first detailed description of the cell nucleus, observed in 1827 the ceaseless, irregular motion of pollen grains suspended in water and, later, of much smaller inert mineral particles, ruling out a biological explanation for the phenomenon (Figure 1.5). It took nearly eighty years, and the independent work of Einstein and Bachelier at the turn of the twentieth century, for this physical observation to be matched with the mathematical object of Definition 1.3.2. We return to this history, and to the physical derivation of the model, in Section 1.5 later in this chapter.

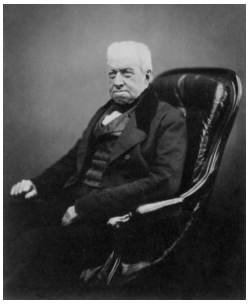


Figure 1.5 – Robert Brown (1773–1858).

1.4 First properties

Having constructed the Wiener process and fixed the vocabulary needed to state its defining properties rigorously, we now record a first batch of consequences of Definition 1.3.2: the process W is Gaussian, which yields a convenient alternative characterization of the Wiener process in terms of its mean and covariance functions, together with a scaling invariance. We close the section by describing the law of W directly as a probability measure on the space of continuous paths, the *Wiener measure*. This point of view will be useful when we discretize or perturb W in later chapters.

1.4.1 Gaussian property and scaling

We first show that the Wiener process is Gaussian (Proposition 1.4.2), and that this property in fact characterizes it among continuous processes with the right mean and covariance functions (Proposition 1.4.4). We then use this characterization to establish the scaling invariance of the Wiener process (Proposition 1.4.5), a property that will reappear throughout the notes.

Definition 1.4.1

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $I \subset \mathbb{R}_+$ be an interval, and let $X = \{X_t\,;\, t \in I\}$ be an $\mathbb{R}$-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ (Definition 1.2.1). We say that X is a *Gaussian process* if, for every $n \geq 1$ and every $t_1, \dots, t_n \in I$, the random vector $(X_{t_1}, \dots, X_{t_n})$ is a (possibly degenerate) Gaussian vector in $\mathbb{R}^n$. That is every linear combination

$$\sum_{i=1}^{n} a_i X_{t_i}, \quad \text{with} \quad a_1, \dots, a_n \in \mathbb{R},$$

is a real Gaussian random variable.

Proposition 1.4.2

Let W be a real-valued Wiener process (Definition 1.3.2). Then W is a Gaussian process (Definition 1.4.1), with mean function $\mathbb{E}[W_t] = 0$, $t \geq 0$, and covariance function (Notation 1.1.3)

$$\mathrm{Cov}(W_s, W_t) = s \wedge t, \qquad s, t \geq 0. \tag{1.39}$$

Proof. Step 1: The increments form a Gaussian vector. Fix $n \geq 1$, $0 = t_0 < t_1 < \cdots < t_n$, and set (see Notation 1.3.1)

$$Y_n := (\delta W_{t_0 t_1}, \dots, \delta W_{t_{n-1} t_n}).$$

By property (iii) of Definition 1.3.2, each coordinate of Y_n is a real Gaussian random variable, and by property (ii) (eq. (1.6)) the coordinates are independent. A linear combination $\sum_{i=1}^{n} a_i\, \delta W_{t_{i-1} t_i}$ of independent Gaussian random variables is itself Gaussian, so Y_n is a Gaussian vector.

Step 2: W is a Gaussian process. Still considering $0 = t_0 < t_1 < \cdots < t_n$, we set

$$X_n := (W_{t_1}, \dots, W_{t_n}),$$

the random vector whose Gaussianity is to be established. Since $W_0 = 0$ almost surely (property (i) of Definition 1.3.2), we have $W_{t_k} = \sum_{i=1}^{k} \delta W_{t_{i-1} t_i}$ so that

$$X_n = M\, Y_n, \tag{1.40}$$

where $M \in \mathbb{R}^{n\times n}$ is the deterministic lower triangular matrix with $M_{ki} = 1$ for $i \le k$ and $M_{ki} = 0$ otherwise. Every linear combination of the coordinates of X_n is, by (1.40), a linear combination of the coordinates of Y_n, hence Gaussian by Step 1. So X_n is a Gaussian vector. Due to the fact that $n \ge 1$ and $0 < t_1 < \cdots < t_n$ were arbitrarily chosen, we obtain that W is a Gaussian process.

Step 3: Mean and covariance. By property (iii) of Definition 1.3.2, we have $\delta W_{0t} = W_t \sim \mathcal{N}(0,t)$ for every $t \ge 0$. Thus $\mathbb{E}[W_t] = 0$ and $\mathrm{Var}(W_t) = t$. Let now $0 \le s \le t$. Applying property (ii) of Definition 1.3.2 to the partition $0 < s < t$ shows that $\delta W_{0s} = W_s$ and δW_{st} are independent, so

$$\begin{aligned}\mathrm{Cov}(W_s, W_t) &= \mathbb{E}\Big[W_s\Big(W_s + \delta W_{st}\Big)\Big] = \mathbb{E}\Big[W_s^2\Big] + \mathbb{E}[W_s]\,\mathbb{E}[\delta W_{st}] \\ &= s + 0 = s = s \wedge t, \end{aligned} \tag{1.41}$$

This proves (1.39). ■

Remark 1.4.3. *The law of a Gaussian process X (Definition 1.4.1) is entirely determined by its mean function $\mu_t := \mathbb{E}[X_t]$ and its covariance function $\rho(s,t) := \mathrm{Cov}(X_s, X_t)$: for every $n \ge 1$ and $t_1, \dots, t_n \in I$, the law of a Gaussian vector in $\mathbb{R}^n$ is determined by its mean vector and covariance matrix, and here the mean vector of $(X_{t_1}, \dots, X_{t_n})$ is $(\mu_{t_1}, \dots, \mu_{t_n})$ while its covariance matrix has entries $\rho(t_i, t_j)$, $1 \le i, j \le n$. By Proposition 1.4.2, the finite-dimensional distributions of a Wiener process are thus exactly those of the centered Gaussian process with covariance $\rho(s,t) = s \wedge t$.*

Proposition 1.4.4

Let W be a real-valued, continuous Gaussian process (Definitions 1.2.1 and 1.4.1) with $\mathbb{E}[W_t] = 0$ for every $t \ge 0$ and $\mathrm{Cov}(W_s, W_t) = s \wedge t$ for every $s, t \ge 0$. Then W is a Wiener process.

Proof. By Remark 1.4.3, the hypotheses on the mean and covariance functions of W determine its finite-dimensional distributions to be exactly those of a Wiener process (Proposition 1.4.2), which are properties (i)–(iii) of Definition 1.3.2. Specifically, we have $W_0 \sim \mathcal{N}(0,0)$ almost surely equal to 0 which gives (i). Next for $0 \le s < t$, the increment

$$\delta W_{st} = W_t - W_s$$

is a linear combination of the jointly Gaussian couple (W_s, W_t). Therefore δW_{st} is Gaussian, with mean 0 and variance

$$\mathrm{Var}(W_t) - 2\mathrm{Cov}(W_s, W_t) + \mathrm{Var}(W_s) = t - 2s + s = t - s,$$

which gives (iii). Independence of increments over an arbitrary partition, property (ii), follows since jointly Gaussian random variables are independent as soon as they are uncorrelated, and the computation of Step 3 in the proof of Proposition 1.4.2 shows that $\mathrm{Cov}(\delta W_{t_{i-1}t_i}, \delta W_{t_{j-1}t_j}) = 0$ for $i \ne j$ under $\rho(s,t) = s \wedge t$. Property (iv), continuity of the paths, is part of the standing hypothesis on W. Hence W satisfies (i)–(iv) of Definition 1.3.2. ■

Proposition 1.4.5: Brownian scaling

Let W be a real-valued Wiener process (Definition 1.3.2) and let $a > 0$. The rescaled process

$$W_t^a := a^{-1/2} W_{at}, \qquad t \geq 0, \tag{1.42}$$

is again a Wiener process.

Proof. We check the hypotheses of Proposition 1.4.4 for W^a.

Step 1: Continuity. The map $t \mapsto at$ is continuous, and W has continuous paths almost surely (property (iv) of Definition 1.3.2), so $t \mapsto W_{at}$, and hence $t \mapsto W_t^a = a^{-1/2} W_{at}$, has continuous paths almost surely.

Step 2: Gaussian property and mean. For $t_1, \ldots, t_n \geq 0$, the random vector

$$\left(W_{t_1}^a, \ldots, W_{t_n}^a\right) = a^{-1/2} \left(W_{at_1}, \ldots, W_{at_n}\right) \tag{1.43}$$

is a scalar multiple of a Gaussian vector (see Proposition 1.4.2), hence Gaussian. Thus W^a is a Gaussian process with $\mathbb{E}[W_t^a] = a^{-1/2}\mathbb{E}[W_{at}] = 0$.

Step 3: Covariance. By (1.39), it is readily checked that

$$\operatorname{Cov}(W_s^a, W_t^a) = a^{-1} \operatorname{Cov}(W_{as}, W_{at}) = a^{-1} (as \wedge at) = a^{-1} a (s \wedge t) = s \wedge t. \tag{1.44}$$

By Steps 1–3, W^a satisfies the hypotheses of Proposition 1.4.4, so W^a is a Wiener process. ∎

Having described the Wiener process through its finite-dimensional distributions, we now turn to a global description of its law as a single probability measure on a space of continuous paths.

1.4.2 The canonical space and Wiener measure

We now describe the law of an $\mathbb{R}^n$-valued Wiener process directly as a probability measure on a space of continuous functions, rather than only through its finite-dimensional distributions: this is the point of view underlying, for instance, the invariance principle mentioned in Remark 1.3.5.

Definition 1.4.6

Let $n \geq 1$. The *canonical space* is the space

$$E := \mathcal{C}(\mathbb{R}_+; \mathbb{R}^n) \tag{1.45}$$

of continuous functions $f : \mathbb{R}_+ \to \mathbb{R}^n$, equipped with the distance

$$d(f,g) := \sum_{k\geq 1} \frac{1}{2^k} \frac{\|f-g\|_{\infty,k}}{1+\|f-g\|_{\infty,k}}, \qquad \|f-g\|_{\infty,k} := \sup_{t\in[0,k]} |f_t - g_t|. \tag{1.46}$$

Before using (E,d) to carry probability measures, we check that it is a well-behaved metric space, in the sense of the following proposition.

Proposition 1.4.7

The canonical space (E,d) of Definition 1.4.6 is a separable, complete metric space, and a sequence $(f_m)_m$ converges to f for d if and only if $f_m \to f$ uniformly on every compact subset of $\mathbb{R}_+$.

Proof. Step 1: d is a distance inducing the topology of local uniform convergence. The function $\varphi(x) := x/(1+x)$, $x \geq 0$, is increasing and subadditive ($\varphi(x+y) \leq \varphi(x)+\varphi(y)$), so each term $2^{-k}\varphi(\|f-g\|_{\infty,k})$ in (1.46) satisfies the triangle inequality. Hence the convergent series d also verifies the triangle inequality. Together with the easily checked property

$$d(f,g) = 0 \iff f = g,$$

this shows d is a distance on E. Since $\|f-g\|_{\infty,k}$ is nondecreasing in k and $\varphi \leq 1$, a sequence $(f_m)_m$ satisfies $d(f_m,f) \to 0$ if and only if $\|f_m - f\|_{\infty,k} \to 0$ for every fixed $k \geq 1$, that is if and only if $f_m \to f$ uniformly on every compact subset of $\mathbb{R}_+$.

Step 2: Completeness. Let $(f_m)_m$ be d-Cauchy. By Step 1, $(f_m)_m$ is uniformly Cauchy on $[0,k]$ for every $k \geq 1$, hence converges uniformly there to some continuous $f^{(k)} : [0,k] \to \mathbb{R}^n$; the restriction of $f^{(k+1)}$ to $[0,k]$ equals $f^{(k)}$ by uniqueness of uniform limits, so the $f^{(k)}$ glue into a single $f \in E$, and $f_m \to f$ for d by Step 1.

Step 3: Separability. Let $D \subset E$ be the countable set of continuous, piecewise affine functions $\mathbb{R}_+ \to \mathbb{R}^n$ with finitely many breakpoints, all with rational time and value coordinates. Given $f \in E$, $k \geq 1$, and $\varepsilon > 0$, uniform continuity of f on $[0,k]$ produces a piecewise affine $g : [0,k] \to \mathbb{R}^n$, with finitely many breakpoints at rational times and values, such that $\|f-g\|_{\infty,k} < \varepsilon$; extending g to $\mathbb{R}_+$ by setting $g \equiv g(k)$ on $[k,\infty)$ keeps it continuous, piecewise affine, and with rational coordinates throughout, so this extension lies in D. Therefore D is dense in (E,d) by Step 1. ∎

To describe the law of W as a measure on E, we still need a suitable σ-algebra, tailored to the finite-dimensional distributions of the process. It is generated by the following cylinder sets.

Definition 1.4.8

Let $m \geq 1$, let $0 \leq t_1 < \cdots < t_m$, and let $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}^n)$ be Borel sets. The *cylinder set*

$$R_{t_1,\ldots,t_m}(A_1, \ldots, A_m) := \left\{ f \in E \,;\, f_{t_1} \in A_1, \ \ldots, \ f_{t_m} \in A_m \right\} \subset E \tag{1.47}$$

is the set of paths passing through $A_1, \ldots, A_m$ at times $t_1, \ldots, t_m$. We denote by $\mathcal{A}$ the σ-algebra on E generated by all cylinder sets.

This σ-algebra $\mathcal{A}$ turns out to coincide with the Borel σ-algebra of (E, d), so that constructing a measure on $(E, \mathcal{A})$ will be the same as constructing a Borel measure on E.

Proposition 1.4.9

Let $\mathcal{A}$ be the cylinder σ-algebra on E (Definition 1.4.8). Then $\mathcal{A} = \mathcal{B}(E)$, the Borel σ-algebra of (E, d) (Definition 1.4.6).

Proof. Step 1: $\mathcal{A} \subset \mathcal{B}(E)$. For a fixed $t \geq 0$, the evaluation map $e_t : E \to \mathbb{R}^n$, $e_t(f) := f_t$, is continuous for d by Proposition 1.4.7 (convergence for d implies pointwise, in fact locally uniform, convergence). Hence e_t is $\mathcal{B}(E)/\mathcal{B}(\mathbb{R}^n)$-measurable. Furthermore, a generic cylinder set can be written as

$$R_{t_1,\ldots,t_m}(A_1, \ldots, A_m) = \bigcap_{i=1}^{m} e_{t_i}^{-1}(A_i).$$

Therefore it is a finite intersection of Borel sets of E, hence itself Borel. We thus get $\mathcal{A} \subset \mathcal{B}(E)$.

Step 2: $\mathcal{B}(E) \subset \mathcal{A}$. Fix $f \in E$ and $\varepsilon > 0$. Since $t \mapsto f_t - g_t$ is continuous for every $g \in E$, the countable dense subset $\mathbb{Q}_+ \subset \mathbb{R}_+$ (Remark 1.2.3(iii)) satisfies

$$\|f - g\|_{\infty,k} = \sup_{q \in \mathbb{Q}_+ \cap [0,k]} |f_q - g_q|$$

for every $k \geq 1$. This right-hand side is, as a countable supremum of the maps $g \mapsto |f_q - g_q|$, an $\mathcal{A}$-measurable function of g. Consequently

$$g \longmapsto d(f, g) = \sum_{k \geq 1} 2^{-k} \varphi(\|f - g\|_{\infty,k})$$

is $\mathcal{A}$-measurable, so the open ball $B(f, \varepsilon) = \{g \in E \,;\, d(f, g) < \varepsilon\}$ belongs to $\mathcal{A}$. Since (E, d) is separable (Proposition 1.4.7), $\mathcal{B}(E)$ is generated by the countable family of balls centered at points of a countable dense subset of E with rational radii, all of which lie in $\mathcal{A}$ by the above. We have thus obtained $\mathcal{B}(E) \subset \mathcal{A}$. ∎

With $(E, \mathcal{A}) = (E, \mathcal{B}(E))$ in hand, we can now transport the law of W onto the canonical space, defining the Wiener measure.

Proposition 1.4.10

Let $W = \{W_t\,;\, t \geq 0\}$ be an $\mathbb{R}^n$-valued Wiener process on $(\Omega, \mathcal{F}, \mathbb{P})$ (Definition 1.3.13), so that almost every path of W belongs to the canonical space E (Definition 1.4.6), and let

$$T\,:\,(\Omega,\mathcal{F}) \longrightarrow (E,\mathcal{A}), \qquad T(\omega)\,:=\,\{W_t(\omega)\,;\,t\geq 0\}, \tag{1.48}$$

with $\mathcal{A} = \mathcal{B}(E)$ as in Definition 1.4.8 and Proposition 1.4.9. Then the following holds true:

(i) The map T is measurable.

(ii) The pushforward measure $\mathbb{P}_0 := \mathbb{P} \circ T^{-1}$ on $(E, \mathcal{A})$ is a probability measure, called the *Wiener measure*.

(iii) Under $\mathbb{P}_0$, the canonical process $\{\omega_t\,;\, t \geq 0\}$ defined by $\omega_t(f) := f_t$ for $f \in E$ is an $\mathbb{R}^n$-valued Wiener process, with the same law as W.

Proof. Step 1: Measurability of T. For a cylinder set $R := R_{t_1,\dots,t_m}(A_1, \dots, A_m)$ (see Definition 1.4.8),

$$T^{-1}(R) \;=\; \Big\{\omega \in \Omega\,;\, W_{t_1}(\omega) \in A_1, \dots, W_{t_m}(\omega) \in A_m\Big\} \;\in\; \mathcal{F}, \tag{1.49}$$

since each W_{t_i} is $\mathcal{F}$-measurable (Definition 1.2.1). To extend this from cylinder sets to all of $\mathcal{A}$, introduce the collection

$$\mathcal{G} \;:=\; \Big\{B \subset E\,;\, T^{-1}(B) \in \mathcal{F}\Big\}.$$

We check that $\mathcal{G}$ satisfies the three axioms of Definition 1.1.1.

(i) $T^{-1}(E) = \Omega \in \mathcal{F}$, so $E \in \mathcal{G}$.

(ii) If $B \in \mathcal{G}$, then $T^{-1}(E \setminus B) = \Omega \setminus T^{-1}(B) \in \mathcal{F}$, since $\mathcal{F}$ is a σ-algebra. Therefore $E \setminus B \in \mathcal{G}$.

(iii) If $(B_k)_{k\geq 1} \subset \mathcal{G}$, then $T^{-1}\Big(\bigcup_k B_k\Big) = \bigcup_k T^{-1}(B_k) \in \mathcal{F}$, again since $\mathcal{F}$ is a σ-algebra. Hence $\bigcup_k B_k \in \mathcal{G}$.

Thus $\mathcal{G}$ is indeed a σ-algebra on E, and by (1.49) it contains every cylinder set. Since $\mathcal{A}$ (Definition 1.4.8) is by definition the *smallest* σ-algebra on E containing all cylinder sets, it follows that $\mathcal{A} \subset \mathcal{G}$, that is $T^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{A}$. This is precisely the statement that $T : (\Omega, \mathcal{F}) \to (E, \mathcal{A})$ is measurable.

Step 2: $\mathbb{P}_0$ is a probability measure. We check the two axioms of Definition 1.1.2 for $\mathbb{P}_0$ on $(E, \mathcal{A})$. First, $\mathbb{P}_0(E) = \mathbb{P}(T^{-1}(E)) = \mathbb{P}(\Omega) = 1$, which is axiom (i). Next, let $(B_k)_{k\geq 1} \subset \mathcal{A}$ be pairwise disjoint. Since $T^{-1}\Big(\bigcup_k B_k\Big) = \bigcup_k T^{-1}(B_k)$, and the sets $T^{-1}(B_k) \in \mathcal{F}$ are

themselves pairwise disjoint (as preimages of disjoint sets), σ-additivity of $\mathbb{P}$ (Definition 1.1.2) gives

$$\mathbb{P}_0\Big(\bigcup_{k\geq 1} B_k\Big) = \mathbb{P}\Big(\bigcup_{k\geq 1} T^{-1}(B_k)\Big) = \sum_{k\geq 1} \mathbb{P}\big(T^{-1}(B_k)\big) = \sum_{k\geq 1} \mathbb{P}_0(B_k),$$

which is axiom (ii). Hence $\mathbb{P}_0$ is a probability measure on $(E, \mathcal{A})$.

Step 3: Law of the canonical process. For a cylinder set R as above, we have

$$\begin{aligned}\mathbb{P}_0\big(\omega_{t_1} \in A_1, \ldots, \omega_{t_m} \in A_m\big) &= \mathbb{P}_0(R) = \mathbb{P}\big(T^{-1}(R)\big) \\ &= \mathbb{P}\big(W_{t_1} \in A_1, \ldots, W_{t_m} \in A_m\big), \end{aligned} \tag{1.50}$$

using (1.49) for the last equality. Hence, under $\mathbb{P}_0$, the canonical process has the same finite-dimensional distributions as W, that is properties (i)–(iii) of Definition 1.3.13 are satisfied. Property (iv), continuity of the paths, holds automatically since every point of E is, by Definition 1.4.6, a continuous function. So the canonical process is an $\mathbb{R}^n$-valued Wiener process under $\mathbb{P}_0$, with the same law as W. ■

We close this section with a word on the early physical history of Brownian motion. In 1905, while employed as a technical expert at the Swiss patent office in Bern, Albert Einstein published, among four papers that would reshape physics that year, a derivation of the erratic motion of pollen grains (Figure 1.5) from the incessant collisions of much smaller, invisible water molecules. Beyond explaining Robert Brown's observation, this argument provided some of the first quantitative evidence for the existence of atoms, and derived, from physical principles rather than from Definition 1.3.2, the transition density that governs the process: we recover this transition density mathematically in Section 1.6, when we study the Markov property of W.

1.5 Martingale property

We now turn to the property of Brownian motion that structures most of the stochastic calculus developed in this course: relative to an admissible filtration, W is a martingale. We use this fact to build the elementary theory of stopping times, culminating in the optional sampling theorem, and illustrate it with an explicit computation of the exit time of W from an interval.

1.5.1 Conditional expectation

The martingale property is phrased in terms of conditional expectation, whose notation we first recall.

Notation 1.5.1

For a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ and an integrable random variable Y on $(\Omega, \mathcal{F}, \mathbb{P})$, we write $\mathbb{E}[Y \,|\, \mathcal{G}]$ for the conditional expectation of Y given $\mathcal{G}$, that is the (almost surely unique) $\mathcal{G}$-measurable, integrable random variable satisfying

$$\mathbb{E}\big[Y\,\mathbf{1}_A\big] \;=\; \mathbb{E}\big[\mathbb{E}[Y \,|\, \mathcal{G}]\,\mathbf{1}_A\big], \qquad \text{for every } A \in \mathcal{G}. \tag{1.51}$$

We use freely two elementary properties of conditional expectation not restated below: linearity, and pulling out $\mathcal{G}$-measurable factors.

Beyond these two properties, we record here, for later reference, three further properties of conditional expectation that we invoke repeatedly in the rest of the course. All three are standard, so we state them without proof and refer to Durrett [2] for detailed arguments.

Proposition 1.5.2: Tower property, unconditional form

Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra and let Y be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\mathbb{E}\big[\mathbb{E}[Y \,|\, \mathcal{G}]\big] \;=\; \mathbb{E}[Y]. \tag{1.52}$$

Proposition 1.5.3: Freezing lemma

Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra, let X be a $\mathcal{G}$-measurable random vector with values in $\mathbb{R}^m$, and let Y be a random vector with values in $\mathbb{R}^k$ such that $Y \perp\!\!\!\perp \mathcal{G}$ (Remark 1.3.16). Then, for every Borel function $\varphi : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ such that $\varphi(X, Y)$ is integrable,

$$\mathbb{E}\big[\varphi(X,Y)\,\big|\,\mathcal{G}\big] \;=\; \mathbb{E}\big[\varphi(x,Y)\big]\Big|_{x=X}, \qquad \text{almost surely.} \tag{1.53}$$

Remark 1.5.4. *Equation (1.53) recovers, as a particular case, an elementary fact that we use repeatedly throughout the course: if $Y \perp\!\!\!\perp \mathcal{G}$ (Remark 1.3.16), then*

$$\mathbb{E}[Y \,|\, \mathcal{G}] \;=\; \mathbb{E}[Y], \qquad \textit{almost surely.} \tag{1.54}$$

The last property we record extends the tower property of Proposition 1.5.2 to two nested σ-algebras, rather than a σ-algebra and the trivial one.

Proposition 1.5.5: Tower property, nested σ-algebras

Let $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ be sub-σ-algebras and let Y be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Then

$$\mathbb{E}\Big[\mathbb{E}[Y \,|\, \mathcal{G}_1]\,\Big|\,\mathcal{G}_2\Big] = \mathbb{E}\Big[\mathbb{E}[Y \,|\, \mathcal{G}_2]\,\Big|\,\mathcal{G}_1\Big] = \mathbb{E}[Y \,|\, \mathcal{G}_1], \qquad \text{almost surely.} \tag{1.55}$$

1.5.2 Martingale property

With conditional expectation now in hand, we turn to defining the martingale property itself.

Definition 1.5.6

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis (Definition 1.2.4) and let $X = \{X_t\,;\, t \geq 0\}$ be an $\mathbb{R}^n$-valued stochastic process on it. We say that X is an *$\mathcal{F}_t$-martingale* if

(i) $X_t \in L^1(\Omega)$ for every $t \geq 0$;

(ii) X is $\mathcal{F}_t$-adapted (Definition 1.2.6);

(iii) for every $0 \leq s < t$, $\mathbb{E}[\delta X_{st} \,|\, \mathcal{F}_s] = 0$ almost surely (Notations 1.3.1 and 1.5.1).

The paradigmatic example, and the one that concerns us here, is the Wiener process itself.

Proposition 1.5.7

Let W be a Wiener process with respect to $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ (Definition 1.3.15). Then W is an $\mathcal{F}_t$-martingale (Definition 1.5.6).

Proof. Step 1: Integrability. From Definition 1.3.15, we have $W_t \sim \mathcal{N}(0, t\,\mathrm{Id}_{\mathbb{R}^n})$ for every $t \geq 0$. In particular, we get $W_t \in L^1(\Omega)$.

Step 2: Adaptedness. This is property (v) of Definition 1.3.15.

Step 3: Vanishing conditional increments. Let $0 \leq s < t$. By property (ii) of Definition 1.3.15, $\delta W_{st} \perp\!\!\!\perp \mathcal{F}_s$, so (1.54) gives

$$\mathbb{E}[\delta W_{st} \,|\, \mathcal{F}_s] = \mathbb{E}[\delta W_{st}] = 0, \tag{1.56}$$

the last equality by property (iii) of Definition 1.3.15. By Steps 1–3, W satisfies (i)–(iii) of Definition 1.5.6, so W is an $\mathcal{F}_t$-martingale. ∎

Having identified W as a martingale, we now introduce the random times at which we will be allowed to evaluate it without destroying this property: stopping times.

Definition 1.5.8

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis (Definition 1.2.4) and let $S : \Omega \to [0, \infty]$ be a random variable. We say that S is a *stopping time* if

$$(S \leq t) \in \mathcal{F}_t, \qquad \text{for every } t \geq 0. \tag{1.57}$$

Remark 1.5.9. *Condition (1.57) formalizes the idea that deciding whether S has occurred by time t only requires the information available at time t: an observer who knows $\mathcal{F}_t$, that is everything about the evolution of the system up to and including time t, can always tell whether $S \leq t$ or $S > t$, without needing to look into the future. This is the property one expects, for instance, of an instant at which a player decides to stop playing a game based only on what has happened so far. This is in contrast to an instant defined by information about the future (such as the time of the last visit of W to 0 before a fixed horizon, which is not a stopping time).*

The basic examples of stopping times, to which we will repeatedly return, are hitting times of closed or open sets.

Proposition 1.5.10

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis, let $X = \{X_t\,;\, t \geq 0\}$ be a continuous, $\mathcal{F}_t$-adapted process with values in $\mathbb{R}^d$ (Definitions 1.2.1 and 1.2.6), let $F \subset \mathbb{R}^d$ be closed, and let $G \subset \mathbb{R}^d$ be open. Set

$$T_F := \inf\{t \geq 0\,;\, X_t \in F\}, \qquad T_G := \inf\{t \geq 0\,;\, X_t \in G\} \qquad (\inf \varnothing := +\infty). \tag{1.58}$$

Then T_F is a stopping time (Definition 1.5.8), and T_G is a stopping time when $X = W$ is a Wiener process with respect to $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ (Definition 1.3.15).

Proof. Step 1: T_F is a stopping time. Fix $t \geq 0$. Since F is closed, $x \in F$ if and only if $\operatorname{dist}(x, F) = 0$, so

$$(T_F \leq t) = \Big(\inf_{s\in[0,t]} \operatorname{dist}(X_s, F) = 0\Big). \tag{1.59}$$

The map $s \mapsto \operatorname{dist}(X_s, F)$ is continuous on $[0, t]$ (a composition of the continuous path $s \mapsto X_s$ with the continuous map $x \mapsto \operatorname{dist}(x, F)$), so its infimum over the compact set $[0, t]$ coincides with its infimum over the countable dense subset $\mathbb{Q} \cap [0, t]$. Hence

$$(T_F \leq t) = \bigcap_{n\geq 1} \bigcup_{s\in\mathbb{Q}\cap[0,t]} \Big(\operatorname{dist}(X_s, F) < \tfrac{1}{n}\Big). \tag{1.60}$$

For fixed $n \geq 1$ and $s \in \mathbb{Q} \cap [0, t]$, the event $(\operatorname{dist}(X_s, F) < 1/n)$ belongs to $\mathcal{F}_s \subset \mathcal{F}_t$, since X is $\mathcal{F}_t$-adapted and $x \mapsto \operatorname{dist}(x, F)$ is Borel. The right-hand side of (1.60) is thus

a countable intersection of countable unions of events in $\mathcal{F}_t$, hence itself in $\mathcal{F}_t$. By (1.59)–(1.60), $(T_F \le t) \in \mathcal{F}_t$, so T_F is a stopping time.

Step 2: reduction for T_G. Assume now $X = W$ and fix $t > 0$. We first show

$$(T_G < t) \;=\; \bigcup_{s \in \mathbb{Q} \cap [0,t)} (W_s \in G). \tag{1.61}$$

The inclusion "$\supset$" is immediate from the definition of T_G in (1.58). Conversely, if $T_G < t$, there is $s < t$ with $W_s \in G$; write $x := W_s$ and let $\varepsilon > 0$ be such that the open ball $B(x, \varepsilon) \subset G$. By continuity of the path of W at s, there is $\delta > 0$ such that

$$W_r \;\in\; B(x, \varepsilon), \qquad \text{for every } r \in (s - \delta, s + \delta). \tag{1.62}$$

Figure 1.6 illustrates relation (1.62): once the path reaches $x = W_s \in G$, it stays trapped in the ball $B(x, \varepsilon) \subset G$ throughout the whole time window $(s - \delta, s + \delta)$.

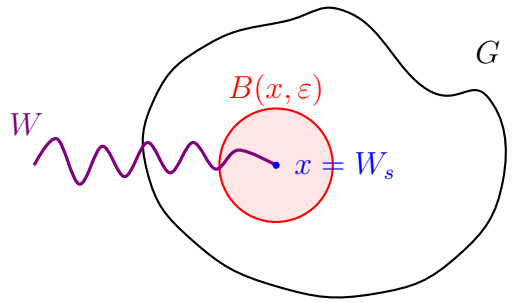


Figure 1.6 – Continuity of W at s. Since $x = W_s \in G$ and $B(x, \varepsilon) \subset G$, there is $\delta > 0$ such that $W_r \in B(x, \varepsilon)$ for every $r \in (s - \delta, s + \delta)$, as in (1.62).

In particular there exists $q \in \mathbb{Q} \cap (s - \delta, s] \subset \mathbb{Q} \cap [0, t)$ with $W_q \in B(x, \varepsilon) \subset G$, which proves the inclusion "$\subset$" and hence (1.61). Since $\mathbb{Q} \cap [0, t)$ is countable and each $(W_s \in G) \in \mathcal{F}_s \subset \mathcal{F}_t$ by adaptedness, relation (1.61) shows $(T_G < t) \in \mathcal{F}_t$.

Step 3: from optional to stopping time. We say, more generally, that a random variable $T : \Omega \to [0, \infty]$ is an *optional time* if

$$(T < t) \;\in\; \mathcal{F}_t, \qquad \text{for every } t > 0. \tag{1.63}$$

Step 2 shows that T_G is optional. Every stopping time is optional (since $(T < t) = \bigcup_{n \ge 1} (T \le t - 1/n) \in \mathcal{F}_t$). Furthermore, an optional time T always satisfies the weaker measurability

$$(T \le t) \;\in\; \mathcal{F}_{t+} \;:=\; \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}, \tag{1.64}$$

since $(T \le t) = \bigcap_{n \ge 1} (T < t + 1/n)$. The two notions coincide, optional time = stopping time, whenever the filtration satisfies $\mathcal{F}_{t+} = \mathcal{F}_t$ for every $t \ge 0$. This right-continuity property that will be established for the completed natural filtration of W in Section 1.6, as a consequence of the Markov property. Granting this, T_G is a stopping time. ∎

Stopping times also enjoy simple stability properties, which we record for later use.

Proposition 1.5.11

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis and let S, T be stopping times (Definition 1.5.8). Then the following holds true:

(i) $S \wedge T$ and $S \vee T$ (Notation 1.1.3) are stopping times.

(ii) Every deterministic time $T \equiv c$, $c \geq 0$, is a stopping time.

Proof. For every $t \geq 0$, $(S \wedge T \leq t) = (S \leq t) \cup (T \leq t) \in \mathcal{F}_t$ and $(S \vee T \leq t) = (S \leq t) \cap (T \leq t) \in \mathcal{F}_t$, since $\mathcal{F}_t$ is a σ-algebra and both $(S \leq t)$ and $(T \leq t)$ belong to it. This proves (i). Claim (ii) is immediate, since a deterministic (hence constant) random variable is measurable with respect to any σ-algebra. ■

To state the optional sampling theorem below, we also need to make sense of "the information available at a stopping time," extending the deterministic-time σ-algebras $\mathcal{F}_t$.

Definition 1.5.12

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis and let S be a stopping time (Definition 1.5.8). The *information at time* S is the σ-algebra

$$\mathcal{F}_S := \Big\{ A \in \mathcal{F} \,;\, A \cap (S \leq t) \in \mathcal{F}_t \text{ for every } t \geq 0 \Big\}. \tag{1.65}$$

Remark 1.5.13. *One checks readily that $\mathcal{F}_S$ of Definition 1.5.12 is indeed a sub-σ-algebra of $\mathcal{F}$, and, when $S \equiv t_0$ is the deterministic stopping time of Proposition 1.5.11(ii), that $\mathcal{F}_S = \mathcal{F}_{t_0}$. Hence Definition 1.5.12 is exactly the extension, to a stopping time horizon, of the interpretation of $\mathcal{F}_t$ recalled in Definition 1.2.4. Namely $\mathcal{F}_S$ represents the information available to an observer up to the (random) time S.*

We can now state the main consequence of the martingale property that we will use throughout the course: the optional sampling theorem, which extends the identity $\mathbb{E}[X_t] = \mathbb{E}[X_0]$, valid at deterministic times, to a pair of stopping times.

Theorem 1.5.14: Optional sampling

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis, let X be a continuous $\mathcal{F}_t$-martingale (Definitions 1.2.1 and 1.5.6), and let $S \leq T$ be two stopping times (Definition 1.5.8) such that $\{X_{t\wedge T}\,;\, t \geq 0\}$ is a uniformly integrable martingale. Then

$$\mathbb{E}[X_T \,|\, \mathcal{F}_S] \;=\; X_S, \qquad \text{almost surely.} \tag{1.66}$$

In particular, taking expectations in (1.66) and applying the tower property (Proposition 1.5.2, with $\mathcal{G} := \mathcal{F}_S$) to $\mathbb{E}[X_T] = \mathbb{E}[\mathbb{E}[X_T \,|\, \mathcal{F}_S]]$, we have

$$\mathbb{E}[X_T] \;=\; \mathbb{E}[X_S] \;=\; \mathbb{E}[X_0]. \tag{1.67}$$

Sketch of proof. The result is first established for stopping times taking finitely many values, by reduction to the (elementary) discrete-time optional sampling theorem applied to the discrete-time martingale

$$Y_m := X_{t_m \wedge T}, \quad \text{with} \quad t_m := m/2^n,$$

along a dyadic grid fine enough to resolve S and T. The general case follows by approximating S and T from above by such dyadic stopping times and passing to the limit, using the uniform integrability of $\{X_{t\wedge T}\,;\, t \geq 0\}$ and continuity of X to justify the limiting procedure. We omit the details, which are elementary but lengthy, and refer to Karatzas–Shreve [6, Section 1.3.B]. ∎

Remark 1.5.15. *Checking the uniform integrability hypothesis of Theorem 1.5.14 directly can be delicate in general, but is automatic in either of the following situations, writing $Y_t := X_{t\wedge T}$:*

(i) *Y is bounded, that is $|Y_t| \leq M$ for every $t \geq 0$, for some deterministic constant M independent of t;*

(ii) *Y is bounded in $L^p(\Omega)$ for some $p > 1$, that is $\sup_{t\geq 0} \mathbb{E}\big[|Y_t|^p\big] \leq M$ for some deterministic constant M.*

Boundedness in $L^2(\Omega)$, namely the case $p = 2$, is the version we use below.

We illustrate the optional sampling theorem with a first, explicit computation, of a kind that will recur throughout the course: the exit time of a Wiener process from a bounded interval.

Proposition 1.5.16: Exit time from an interval

Let B be a real-valued Wiener process with respect to $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ (Definition 1.3.15). Let also $a, b > 0$, and set

$$T_a := \inf\{t \geq 0\,;\, B_t = -a\}, \qquad T_b := \inf\{t \geq 0\,;\, B_t = b\}, \qquad T := T_a \wedge T_b. \tag{1.68}$$

Then

$$\mathbb{P}(T_a < T_b) = \frac{b}{a+b}, \qquad \text{and} \qquad \mathbb{E}[T] = ab. \tag{1.69}$$

Proof. We admit that $T < \infty$ almost surely, a consequence of the recurrence of B that will be established, via the Markov property, in Section 1.6. Since $\{-a\}$ and $\{b\}$ are closed subsets of $\mathbb{R}$, T_a, T_b, and $T = T_a \wedge T_b$ are stopping times by Propositions 1.5.10 and 1.5.11.

Step 1: B and $t \mapsto B_t^2 - t$ are martingales. By Proposition 1.5.7, B is an $\mathcal{F}_t$-martingale. We claim that $M_t := B_t^2 - t$, $t \geq 0$, is also an $\mathcal{F}_t$-martingale. Indeed $M_t \in L^1(\Omega)$ since $\mathbb{E}[B_t^2] = t < \infty$ (Proposition 1.4.2), M is $\mathcal{F}_t$-adapted since B is, and, for $0 \leq s < t$, writing $B_t = B_s + \delta B_{st}$,

$$\delta M_{st} = B_t^2 - B_s^2 - (t-s) = \big(\delta B_{st}\big)^2 + 2B_s\,\delta B_{st} - (t-s). \tag{1.70}$$

Since $\delta B_{st} \perp\!\!\!\perp \mathcal{F}_s$ with $\delta B_{st} \sim \mathcal{N}(0, t-s)$ (see Definition 1.3.15), relation (1.54) gives

$$\mathbb{E}[(\delta B_{st})^2 \,|\, \mathcal{F}_s] = \mathbb{E}[(\delta B_{st})^2] = t - s,$$

while pulling out the $\mathcal{F}_s$-measurable factor B_s and using (1.56) gives

$$\mathbb{E}[2B_s\,\delta B_{st} \,|\, \mathcal{F}_s] = 2B_s\,\mathbb{E}[\delta B_{st} \,|\, \mathcal{F}_s] = 0.$$

Taking $\mathbb{E}[\,\cdot\,|\,\mathcal{F}_s]$ in (1.70) therefore yields

$$\mathbb{E}[\delta M_{st} \,|\, \mathcal{F}_s] = (t-s) + 0 - (t-s) = 0.$$

We have thus obtained that M is an $\mathcal{F}_t$-martingale.

Step 2: computation of $\mathbb{P}(T_a < T_b)$. Fix $\tau > 0$ and consider the bounded stopping time $T \wedge \tau$. By definition of T, the path B stays in $[-a, b]$ up to time T, and is frozen thereafter, so

$$\big|B_{t\wedge(T\wedge\tau)}\big| \leq a \vee b, \qquad \text{for every } t \geq 0.$$

In particular $\{B_{t\wedge(T\wedge\tau)}\,;\, t \geq 0\}$ is uniformly integrable (Remark 1.5.15(i)). We may therefore apply Theorem 1.5.14 to the martingale B, with stopping times $S = 0 \leq T \wedge \tau$. This gives

$$\mathbb{E}[B_{T\wedge\tau}] = \mathbb{E}[B_0] = 0.$$

We now let $\tau \to \infty$. Since $T < \infty$ almost surely, $T \wedge \tau \to T$ almost surely, and continuity of B gives $B_{T\wedge\tau} \to B_T$ almost surely. As $|B_{T\wedge\tau}| \le a \vee b$ for every τ, dominated convergence yields

$$\mathbb{E}[B_T] \;=\; 0.$$

It remains to identify B_T. On $(T_a < T_b)$ we have $B_T = -a$, and on $(T_b < T_a)$ we have $B_T = b$. Moreover $\mathbb{P}(T_a = T_b) = 0$, since B cannot be simultaneously equal to $-a$ and b. Writing $p := \mathbb{P}(T_a < T_b)$, the identity $\mathbb{E}[B_T] = 0$ therefore reads

$$0 \;=\; -a\,p \;+\; b\,(1-p). \tag{1.71}$$

Solving (1.71) gives $p = b/(a+b)$, the first identity in (1.69).

Step 3: computation of $\mathbb{E}[T]$. We reuse the stopping times $S = 0 \le T \wedge \tau$ of Step 2, now applying Theorem 1.5.14 to the martingale $M_t = B_t^2 - t$ of Step 1. Uniform integrability of $\{M_{t\wedge(T\wedge\tau)}\,;\, t \ge 0\}$ follows as in Step 2, since $B_{t\wedge(T\wedge\tau)}$ is bounded by $a \vee b$ and $t \wedge (T \wedge \tau)$ is bounded by τ. Optional sampling therefore gives $\mathbb{E}[B^2_{T\wedge\tau}] - \mathbb{E}[T \wedge \tau] = \mathbb{E}[M_0] = 0$, that is

$$\mathbb{E}\Big[B^2_{T\wedge\tau}\Big] \;=\; \mathbb{E}[T \wedge \tau]. \tag{1.72}$$

We now let $\tau \to \infty$ in (1.72), and treat each side in turn. On the left, $B^2_{T\wedge\tau} \to B^2_T$ almost surely, and $|B^2_{T\wedge\tau}| \le (a \vee b)^2$ for every τ, so dominated convergence gives

$$\mathbb{E}\Big[B^2_{T\wedge\tau}\Big] \;\longrightarrow\; \mathbb{E}\Big[B^2_T\Big].$$

On the right, $T \wedge \tau$ increases to T almost surely, so monotone convergence gives

$$\mathbb{E}[T \wedge \tau] \;\longrightarrow\; \mathbb{E}[T].$$

Passing to the limit in (1.72) thus yields $\mathbb{E}[B_T^2] = \mathbb{E}[T]$.

Step 4: conclusion. Using $B_T \in \{-a, b\}$ and Step 2,

$$\begin{aligned}\mathbb{E}[T] \;&=\; \mathbb{E}\Big[B_T^2\Big] \;=\; a^2\,\mathbb{P}(T_a < T_b) \;+\; b^2\,\mathbb{P}(T_b < T_a) \;=\; a^2\,\frac{b}{a+b} \;+\; b^2\,\frac{a}{a+b} \\ &=\; \frac{ab(a+b)}{a+b} \;=\; ab,\end{aligned} \tag{1.73}$$

which is the second identity in (1.69). ∎

We close this section with a word on the mathematical history of Brownian motion. Louis Bachelier, born in 1870 to a family of wine merchants, defended in 1900 a doctoral thesis, *Théorie de la spéculation*, that gave the first mathematical description of Brownian motion, five years before Einstein, in order to model the fluctuations of stock prices on the Paris exchange. Working outside the mainstream concerns of the academic establishment of his time, Bachelier's thesis went largely unrecognized during his lifetime, and it was only decades later, with the

Figure 1.7 – Louis Bachelier (1870–1946).

development of both rigorous stochastic calculus and mathematical finance, that the depth of his intuition was fully appreciated. The mathematical object he described by intuitive, non-rigorous means is precisely the Wiener process of Definition 1.3.2, whose rigorous construction, in Section 1.3, is due to Wiener. We continue to develop its properties in the next two sections: first its Markov property, which recovers the transition density underlying Einstein's and Bachelier's derivations, and then the pathwise regularity of its trajectories, which will motivate the construction of the stochastic integral in the next chapter.

1.6 Markov property

Having lifted the law of W to the canonical space $(E, \mathcal{A})$ in Proposition 1.4.10, we now use this construction to express the defining feature of Brownian motion as a Markov process. Recall that the Markov property has to be interpreted as *knowledge of the present position of the path is enough to describe its future evolution.* In this section we first index the Wiener measure by its starting point, then formalize *the future after time s* through a shift operator on E. Combining the two lets us state and prove the Markov property (Theorem 1.6.10), from which we derive the right-continuity of the Brownian filtration promised in the proof of Proposition 1.5.10 (Corollary 1.6.15), its extension to stopping times (Theorem 1.6.16), and the reflection principle (Theorem 1.6.17), one of the most striking explicit computations available for Brownian motion.

The Wiener measure $\mathbb{P}_0$ of Proposition 1.4.10 describes a path started at the origin. To let the starting point vary, we transport $\mathbb{P}_0$ by translation.

Definition 1.6.1

Let $x \in \mathbb{R}^n$. The *translation by x* is the map

$$\tau_x \,:\, E \longrightarrow E, \qquad (\tau_x f)_t \,:=\, x + f_t, \qquad t \geq 0. \tag{1.74}$$

Since τ_x is continuous for d (Definition 1.4.6), with continuous inverse τ_{-x}, it is a homeomorphism of E; in particular τ_x is $\mathcal{A}/\mathcal{A}$-measurable. We define the *Wiener measure started at x* as the pushforward

$$\mathbb{P}_x \,:=\, \mathbb{P}_0 \circ \tau_x^{-1} \tag{1.75}$$

of the Wiener measure $\mathbb{P}_0$ (Proposition 1.4.10) under τ_x.

Proposition 1.6.2

Let $x \in \mathbb{R}^n$ and let $\mathbb{P}_x$ be as in Definition 1.6.1. Then:

(i) $\mathbb{P}_x$ is a probability measure on $(E, \mathcal{A})$.

(ii) $\omega_0 = x$, $\mathbb{P}_x$-almost surely.

(iii) $\{\omega_t - x\,;\, t \geq 0\}$ is an $\mathbb{R}^n$-valued Wiener process under $\mathbb{P}_x$, in the sense of Definition 1.3.13.

Proof. Step 1: $\mathbb{P}_x$ is a probability measure. Since τ_x is $\mathcal{A}/\mathcal{A}$-measurable (Definition 1.6.1) and $\mathbb{P}_0$ is a probability measure on $(E, \mathcal{A})$ (Proposition 1.4.10), checking the two axioms of Definition 1.1.2 for $\mathbb{P}_x$ proceeds exactly as in Step 2 of the proof of Proposition 1.4.10, with T replaced by τ_x and $\mathbb{P}$ by $\mathbb{P}_0$. Hence $\mathbb{P}_x = \mathbb{P}_0 \circ \tau_x^{-1}$ is a probability measure on $(E, \mathcal{A})$.

Step 2: law of the canonical process. Let $m \geq 1$, let $0 \leq t_1 < \cdots < t_m$, let $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}^n)$, and let $R := R_{t_1,\ldots,t_m}(A_1, \ldots, A_m)$ (Definition 1.4.8). By definition of τ_x, we have

$$\tau_x^{-1}(R) \;=\; R_{t_1,\ldots,t_m}(A_1 - x, \ldots, A_m - x), \qquad A_i - x := \{y - x\,;\, y \in A_i\}. \tag{1.76}$$

Using (1.76) and (1.50) for the Wiener measure $\mathbb{P}_0$, we thus get

$$\begin{aligned} \mathbb{P}_x\Big(\omega_{t_1} \in A_1, \ldots, \omega_{t_m} \in A_m\Big) \;&=\; \mathbb{P}_x(R) \;=\; \mathbb{P}_0\Big(\tau_x^{-1}(R)\Big) \\ &=\; \mathbb{P}_0\Big(\omega_{t_1} \in A_1 - x, \ldots, \omega_{t_m} \in A_m - x\Big) \\ &=\; \mathbb{P}_0\Big(x + \omega_{t_1} \in A_1, \ldots, x + \omega_{t_m} \in A_m\Big). \end{aligned} \tag{1.77}$$

Since m, the times t_i, and the sets A_i were arbitrary, relation (1.77) shows that the random vector $(\omega_{t_1}, \ldots, \omega_{t_m})$ has, under $\mathbb{P}_x$, the same law as $(x + \omega_{t_1}, \ldots, x + \omega_{t_m})$ under $\mathbb{P}_0$. Taking $m = 1$, $t_1 = 0$, $A_1 = \{0\}$ gives $\mathbb{P}_x(\omega_0 = x) = \mathbb{P}_0(\omega_0 = 0) = 1$. More generally, properties (ii)–(iii) of Definition 1.3.13 for $\omega - x$ under $\mathbb{P}_x$ follow from (1.77) applied to the increments of ω, which coincide with those of $\omega - x$. Finally, property (iv), continuity of the paths, is automatic since every point of E is, by Definition 1.4.6, a continuous function. Hence $\omega - x$ is an $\mathbb{R}^n$-valued Wiener process under $\mathbb{P}_x$. ∎

With a Wiener measure available at every starting point, we fix, for the remainder of this section, the following working conventions.

Notation 1.6.3

We work on the canonical space $(E, \mathcal{A})$ (Definitions 1.4.6 and 1.4.8), and we fix the following pieces of notation.

(i) $W := \{\omega_t \,;\, t \geq 0\}$ denotes the canonical process (Proposition 1.4.10).

(ii) $\{\mathbb{P}_x \,;\, x \in \mathbb{R}^n\}$ denotes the family of Wiener measures (Definition 1.6.1), and $\mathbb{E}_x$ denotes the expectation under $\mathbb{P}_x$ (Notation 1.1.3).

(iii) $(\mathcal{F}_t)_{t\geq 0}$ denotes the natural filtration $\mathcal{F}_t := \sigma\{W_s \,;\, s \leq t\}$ of W, completed with respect to the $\mathbb{P}_x$-negligible sets, for a fixed but arbitrary $x \in \mathbb{R}^n$.

Under this notation, the law of the increments of W does not actually depend on the starting point x, only their base point does.

Remark 1.6.4. *By Proposition 1.6.2, $W_0 = x$ $\mathbb{P}_x$-almost surely, and $W - x$ is a Wiener process under $\mathbb{P}_x$. Hence the increments δW_{st}, $0 \leq s < t$, have the same law $\mathcal{N}(0, (t-s)\mathrm{Id}_{\mathbb{R}^n})$ and the same independence properties under every $\mathbb{P}_x$. We use this repeatedly below without further comment.*

Next in order to formalize *the future after time s*, we introduce a shift operator on the canonical space.

Definition 1.6.5

Let $t \geq 0$. The *shift by t* is the map

$$\theta_t \,:\, E \longrightarrow E, \quad \text{such that} \quad (\theta_t f)_s \,:=\, f_{t+s}, \qquad s \geq 0. \tag{1.78}$$

As for τ_x in Definition 1.6.1, θ_t is continuous for d, hence $\mathcal{A}/\mathcal{A}$-measurable by Proposition 1.4.9; in particular $Y \circ \theta_t$ is measurable whenever $Y : E \to \mathbb{R}$ is.

Figure 1.8 illustrates the shift operator θ_s. The red curve retraces the black path of W for $t \geq s$, and is exactly the path $u \mapsto (\theta_s W)_u = W_{s+u}$ once time is re-based at s.

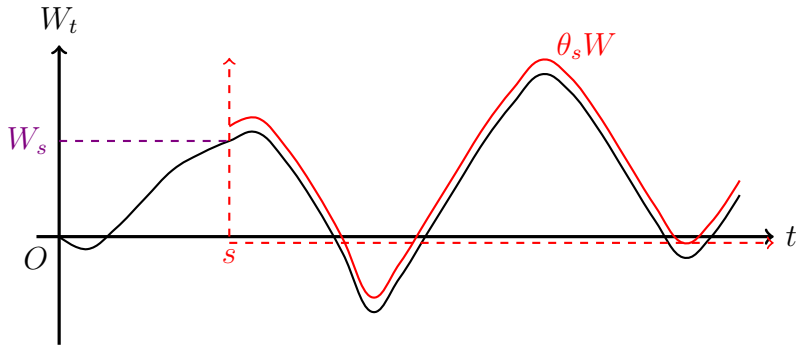


Figure 1.8 – The shift operator θ_s of Definition 1.6.5. The path $t \mapsto W_t$ is shown in black. Its restriction to $t \geq s$, retraced in red and labeled $W \circ \theta_s$, is exactly the path $u \mapsto (\theta_s W)_u = W_{s+u}$ of (1.78), read off against the shifted clock $u = t - s \geq 0$ based at s (dashed red). In particular $(\theta_s W)_0 = W_s$, read off on the vertical axis (dashed violet).

Remark 1.6.6. *For $Y : E \to \mathbb{R}$ measurable and $s \geq 0$, the random variable $Y \circ \theta_s$ depends only on the future of the path after time s. For instance, if*

$$Y(f) = g(f_{t_1}, \ldots, f_{t_k}), \quad 0 \leq t_1 < \cdots < t_k, \quad \text{with} \quad g : (\mathbb{R}^n)^k \to \mathbb{R} \text{ Borel}, \tag{1.79}$$

then, by definition of θ_s,

$$Y \circ \theta_s = g(W_{s+t_1}, \ldots, W_{s+t_k}). \tag{1.80}$$

In particular, $Y \circ \theta_s$ is measurable with respect to $\sigma\{W_{s+u} ; u \geq 0\}$, the σ-algebra generated by W after time s. Note that this σ-algebra already contains W_s, the case $u = 0$.

The Markov property below is most conveniently stated in terms of the transition operator of W, which we introduce now.

Definition 1.6.7

For $t > 0$ and $f : \mathbb{R}^n \to \mathbb{R}$ bounded and Borel, set

$$p_t f(x) := \mathbb{E}_x\big[f(W_t)\big] = \int_{\mathbb{R}^n} f(y) \frac{\exp\Big(-\frac{|y-x|^2}{2t}\Big)}{(2\pi t)^{n/2}}\, dy, \qquad x \in \mathbb{R}^n, \tag{1.81}$$

the second equality by Remark 1.6.4, and set $p_0 f := f$.

Extending an identity from simple, cylindrical functionals of W (Remark 1.6.6) to every bounded measurable functional is a recurring step below; we isolate the measure-theoretic tool once and for all.

Lemma 1.6.8: Functional monotone class theorem

Let $\mathcal{M}$ be a family of bounded functions $E \to \mathbb{R}$ such that:

- $\mathcal{M}$ is closed under multiplication.
- $\mathcal{M}$ generates $\mathcal{A}$, that is, $\mathcal{A}$ is the smallest σ-algebra on E making every $Y \in \mathcal{M}$ measurable.

Let $\mathcal{H}$ be a vector space of bounded, $\mathcal{A}$-measurable functions $E \to \mathbb{R}$ such that:

- $\mathcal{H}$ contains the constants and $\mathcal{M}$.
- $\mathcal{H}$ is closed under bounded, increasing pointwise limits.

Then $\mathcal{H}$ contains every bounded, $\mathcal{A}$-measurable function.

Lemma 1.6.8 is a standard consequence of Dynkin's π-λ theorem, applied to the family of sets $\{Y \le c\,;\, Y \in \mathcal{M},\, c \in \mathbb{R}\}$. We omit the (routine) proof and refer to Durrett [2] or Karatzas–Shreve [6]. We now exhibit a family $\mathcal{M}$ to which Lemma 1.6.8 applies, and that we use throughout the rest of this section.

Remark 1.6.9. *Let $\mathcal{M}$ be the family of bounded, continuous cylindrical functions*

$$Y(f) \;=\; g(f_{t_1}, \dots, f_{t_k}), \qquad k \ge 1,\; 0 \le t_1 < \cdots < t_k,\; g \in \mathcal{C}_b\Big((\mathbb{R}^n)^k\Big). \tag{1.82}$$

This family is closed under multiplication. It also generates $\mathcal{A}$. Indeed, the evaluation maps e_t already generate $\mathcal{A}$ (Definition 1.4.8), and every set $(e_t \in A)$, A Borel, is recovered from $\mathcal{M}$ by approximating $\mathbf{1}_A$ by bounded continuous functions. Hence $\mathcal{M}$ satisfies the assumptions of Lemma 1.6.8.

We can now state the Markov property.

Theorem 1.6.10: Markov property

Let $s \ge 0$ and let $Y : E \to \mathbb{R}$ be bounded and $\mathcal{A}$-measurable. Then, for every $x \in \mathbb{R}^n$,

$$\mathbb{E}_x\Big[Y \circ \theta_s \,\Big|\, \mathcal{F}_s\Big] \;=\; \mathbb{E}_{W_s}[Y], \qquad \mathbb{P}_x\text{-almost surely}, \tag{1.83}$$

where $\mathbb{E}_{W_s}[Y]$ denotes the (bounded, Borel) function $y \mapsto \mathbb{E}_y[Y]$ evaluated at $y = W_s$.

Remark 1.6.11. *Relation (1.83) says that the conditional expectation of any bounded, measurable functional of the future of W after time s, given the whole history $\mathcal{F}_s$ up to s, depends on that history only through the current position W_s: the future of W can be predicted, in law, from W_s alone, exactly as if the process were restarted from W_s at time s.*

Proof. Step 1: a single future time. Assume first $Y(f) = g(f_t)$ for some bounded Borel $g : \mathbb{R}^n \to \mathbb{R}$ and $t \geq 0$, so that

$$Y \circ \theta_s = g(W_{s+t}),$$

by Remark 1.6.6. Fix $x \in \mathbb{R}^n$. The random vector W_s is $\mathcal{F}_s$-measurable, and $\delta W_{s,s+t}$ is independent of $\mathcal{F}_s$ under $\mathbb{P}_x$, with law $\mathcal{N}(0, t\,\mathrm{Id}_{\mathbb{R}^n})$ by Remark 1.6.4. The freezing lemma (Proposition 1.5.3, with $X := W_s$, $Y := \delta W_{s,s+t}$, and $\varphi(x, y) := g(x + y)$) therefore gives

$$\mathbb{E}_x\Big[g(W_{s+t})\,\Big|\,\mathcal{F}_s\Big] \;=\; \mathbb{E}_x\Big[g(W_s + \delta W_{s,s+t})\,\Big|\,\mathcal{F}_s\Big] \;=\; p_t g(W_s), \tag{1.84}$$

using Definition 1.6.7 for the last equality. On the other hand, $\mathbb{E}_y[Y] = \mathbb{E}_y[g(W_t)] = p_t g(y)$ for every $y \in \mathbb{R}^n$, again by Definition 1.6.7. Combining with (1.84) gives (1.83) for this choice of Y.

Step 2: finitely many future times. Assume now $Y(f) = g(f_{t_1}, \ldots, f_{t_k})$ as in (1.79), so that

$$Y \circ \theta_s = g(W_{s+t_1}, \ldots, W_{s+t_k}).$$

Write $u_1 := t_1$ and $u_i := t_i - t_{i-1}$ for $2 \leq i \leq k$. By Notation 1.6.3, the increments $\delta W_{s,s+t_1}, \delta W_{s+t_1,s+t_2}, \ldots, \delta W_{s+t_{k-1},s+t_k}$ are independent of $\mathcal{F}_s$ under $\mathbb{P}_x$, and independent of one another, with respective laws $\mathcal{N}(0, u_1\mathrm{Id}_{\mathbb{R}^n}), \ldots, \mathcal{N}(0, u_k\mathrm{Id}_{\mathbb{R}^n})$. Writing $W_{s+t_i} = W_s + \sum_{j \leq i} \delta W_{s+t_{j-1},s+t_j}$ (with $t_0 := 0$) and repeating the computation of Step 1 successively in each increment gives

$$\mathbb{E}_x\Big[Y \circ \theta_s\,\Big|\,\mathcal{F}_s\Big] \;=\; q_{t_1,\ldots,t_k} g(W_s), \tag{1.85}$$

where

$$q_{t_1,\ldots,t_k} g(y) \;:=\; \mathbb{E}\Big[g(y + Z_1,\; y + Z_1 + Z_2,\; \ldots,\; y + Z_1 + \cdots + Z_k)\Big], \tag{1.86}$$

for $Z_1, \ldots, Z_k$ independent, $Z_i \sim \mathcal{N}(0, u_i\mathrm{Id}_{\mathbb{R}^n})$. Furthermore, under $\mathbb{P}_y$ the vector

$$(W_{t_1}, \ldots, W_{t_k})$$

has exactly the law of $(y + Z_1, \ldots, y + Z_1 + \cdots + Z_k)$ for the same Z_i's. Therefore $\mathbb{E}_y[Y] = q_{t_1,\ldots,t_k} g(y)$. Combining with (1.85) gives (1.83) for this Y.

Step 3: general Y. Fix $s \geq 0$ and $x \in \mathbb{R}^n$, and let

$$\mathcal{H} \;:=\; \Big\{Y : E \to \mathbb{R} \text{ bounded and } \mathcal{A}\text{-measurable}\,;\, (1.83) \text{ holds for } Y\Big\}.$$

By linearity of conditional expectation, $\mathcal{H}$ is a vector space containing the constants, and by (conditional) monotone convergence, $\mathcal{H}$ is closed under bounded, increasing pointwise limits. We now check that $\mathcal{H}$ contains the family $\mathcal{M}$ of Remark 1.6.9. Namely the fact that a generic $Y \in \mathcal{M}$ belongs to $\mathcal{H}$ requires two things. First, that (1.83) holds for Y, which is exactly the content of Steps 1–2. Second, that $y \mapsto \mathbb{E}_y[Y]$ is Borel measurable, as needed for $\mathbb{E}_{W_s}[Y]$ to make sense in the statement of the theorem. To this aim, write

$$Y(f) = g(f_{t_1}, \ldots, f_{t_k}),$$

with $g \in \mathcal{C}_b((\mathbb{R}^n)^k)$, as in Remark 1.6.9. By Step 2, $\mathbb{E}_y[Y] = q_{t_1,\dots,t_k}g(y)$, and we check that this function of y is continuous (which is stronger than Borel measurable). To this aim, let $y_{k'} \to y$ in $\mathbb{R}^n$ as $k' \to \infty$. Since g is continuous,

$$g\Big(y_{k'} + Z_1,\, y_{k'} + Z_1 + Z_2,\, \dots\Big) \longrightarrow g\Big(y + Z_1,\, y + Z_1 + Z_2,\, \dots\Big) \tag{1.87}$$

almost surely as $k' \to \infty$. Since g is also bounded, dominated convergence applied to (1.86) gives

$$q_{t_1,\dots,t_k}g(y_{k'}) \longrightarrow q_{t_1,\dots,t_k}g(y), \qquad k' \to \infty. \tag{1.88}$$

Hence $y \mapsto q_{t_1,\dots,t_k}g(y)$ is continuous, and $\mathcal{H}$ contains $\mathcal{M}$. Lemma 1.6.8 therefore shows that $\mathcal{H}$ contains every bounded, $\mathcal{A}$-measurable function, that is (1.83) holds for every such Y. ∎

As a first application of the Markov property, we record the semigroup structure of the transition operator p_t, and its link to the heat equation.

Proposition 1.6.12: Heat semigroup

The family $\{p_t\,;\, t \geq 0\}$ (Definition 1.6.7) is a semigroup of operators on the space of bounded Borel functions $\mathbb{R}^n \to \mathbb{R}$, that is

$$p_{t+u} \;=\; p_t \circ p_u, \qquad \text{for every } t, u \geq 0. \tag{1.89}$$

Proof. Fix $t, u \geq 0$, $x \in \mathbb{R}^n$, and a bounded Borel $f : \mathbb{R}^n \to \mathbb{R}$. Applying the Markov property (Theorem 1.6.10) at $s = t$ to $Y := f(W_u)$ (so that $Y \circ \theta_t = f(W_{t+u})$) and taking $\mathbb{E}_x[\,\cdot\,]$ of both sides, the tower property (Proposition 1.5.2, with $\mathcal{G} := \mathcal{F}_t$) gives

$$\begin{aligned} p_{t+u}f(x) \;=\; \mathbb{E}_x\Big[f(W_{t+u})\Big] \;=\; \mathbb{E}_x\Big[Y \circ \theta_t\Big] \;=\; \mathbb{E}_x\Big[\mathbb{E}_{W_t}[Y]\Big] \\ = \mathbb{E}_x\Big[p_u f(W_t)\Big] \;=\; p_t(p_u f)(x), \end{aligned} \tag{1.90}$$

the last equality again by Definition 1.6.7. This is (1.89). ∎

Remark 1.6.13 (Generator and Feynman–Kac formula)**.** *Differentiating the Gaussian integral in* (1.81) *under the integral sign shows that, for bounded continuous f, the function $u(t,x) := p_t f(x)$ solves the heat equation*

$$\partial_t u(t,x) \;=\; \tfrac{1}{2}\Delta u(t,x), \qquad u(0,x) = f(x). \tag{1.91}$$

In other words, the generator of the semigroup $\{p_t\}$ of Proposition 1.6.12, in the sense that $\partial_t p_t f = \mathcal{L} p_t f$, is half the Laplace operator, $\mathcal{L} = \frac{1}{2}\Delta$. The representation (1.91) *of the solution of the heat equation as $u(t,x) = \mathbb{E}_x[f(W_t)]$ is the simplest instance of the* Feynman–Kac formula,*. This probabilistic representation of solutions of parabolic partial differential equations that will reappear, in more elaborate forms, later in the course.*

We now turn to the promise made in the proof of Proposition 1.5.10: the completed natural filtration of W is right-continuous. This follows from a zero-one law of Blumenthal, itself a direct consequence of the Markov property. It expresses the intuitive fact that no information about the immediate future of W, beyond the starting point x, is available at time 0.

Proposition 1.6.14: Blumenthal's zero-one law

Let $x \in \mathbb{R}^n$ and let $\mathcal{F}_{0+} := \bigcap_{\varepsilon>0} \mathcal{F}_\varepsilon$ (see (1.64) with $t = 0$). Then every $A \in \mathcal{F}_{0+}$ satisfies $\mathbb{P}_x(A) \in \{0, 1\}$.

Proof. Step 1: $\mathcal{F}_{0+}$ is independent of $\mathcal{F}$. Fix $A \in \mathcal{F}_{0+}$ and let $Y \in \mathcal{M}$ be a bounded, continuous cylindrical function as in (1.82). For $\varepsilon > 0$, since $A \in \mathcal{F}_{0+} \subset \mathcal{F}_\varepsilon$, the Markov property (Theorem 1.6.10) at $s = \varepsilon$ gives

$$\mathbb{E}_x\Big[Y \circ \theta_\varepsilon \, \mathbf{1}_A\Big] \;=\; \mathbb{E}_x\Big[\mathbb{E}_x[Y \circ \theta_\varepsilon \,|\, \mathcal{F}_\varepsilon] \, \mathbf{1}_A\Big] \;=\; \mathbb{E}_x\Big[\mathbb{E}_{W_\varepsilon}[Y] \, \mathbf{1}_A\Big]. \tag{1.92}$$

As $\varepsilon \to 0^+$, $W_\varepsilon \to W_0 = x$ ($\mathbb{P}_x$-almost surely), by continuity of the paths. Moreover, since Y is continuous and cylindrical, the function $y \mapsto \mathbb{E}_y[Y]$ is continuous (by the argument of Step 3 in the proof of Theorem 1.6.10). It is also readily checked that $Y \circ \theta_\varepsilon \to Y$, $\mathbb{P}_x$-almost surely. Both sides of (1.92) are bounded by $\sup|Y|$, so dominated convergence gives, letting $\varepsilon \to 0^+$,

$$\mathbb{E}_x\Big[Y \, \mathbf{1}_A\Big] \;=\; \mathbb{E}_x[Y] \, \mathbb{P}_x(A). \tag{1.93}$$

Relation (1.93) holds for every $Y \in \mathcal{M}$; by the same monotone class argument as in Step 3 of the proof of Theorem 1.6.10 (applied to the vector space of bounded, $\mathcal{A}$-measurable Y), it extends to every bounded, $\mathcal{A}$-measurable Y. In other words, A is independent of $\mathcal{F} = \mathcal{A}$.

Step 2: conclusion. Since $A \in \mathcal{F}_{0+} \subset \mathcal{F}$, taking $Y = \mathbf{1}_A$ in (1.93) gives $\mathbb{P}_x(A) = \mathbb{P}_x(A)^2$, that is $\mathbb{P}_x(A) \in \{0, 1\}$. ∎

Corollary 1.6.15: Right-continuity of the Brownian filtration

For every $t \geq 0$, $\mathcal{F}_{t+} = \mathcal{F}_t$, where $(\mathcal{F}_t)_{t\geq 0}$ is the completed natural filtration of W (Notation 1.6.3) and $\mathcal{F}_{t+}$ is as in (1.64). Consequently, by Step 3 of the proof of Proposition 1.5.10, optional times and stopping times coincide for $(\mathcal{F}_t)_{t\geq 0}$.

Sketch of proof. The inclusion $\mathcal{F}_t \subset \mathcal{F}_{t+}$ is immediate from the definition of $\mathcal{F}_{t+}$. For the converse, fix $A \in \mathcal{F}_{t+}$. By the Markov property (Theorem 1.6.10) at $s = t$, the shifted process $W' := W \circ \theta_t$ decomposes as

$$W'_u \;=\; W_t + \widetilde{W}_u, \qquad u \geq 0, \tag{1.94}$$

where $\widetilde{W} := W' - W_t$ is a Wiener process started at 0, independent of $\mathcal{F}_t$. Since W_t is $\mathcal{F}_t$-measurable, W' itself is not independent of $\mathcal{F}_t$. What Theorem 1.6.10 actually gives

is that, conditionally on $\mathcal{F}_t$, W' is a Wiener process started at W_t. One checks that $A \in \mathcal{F}_t \vee \mathcal{G}_{0+}$, where

$$\mathcal{G}_{0+} := \bigcap_{\varepsilon>0} \sigma\{W'_u\,;\, u \le \varepsilon\} \tag{1.95}$$

is the germ σ-algebra of W' at 0. By (1.94) and since W_t is $\mathcal{F}_t$-measurable, $\mathcal{G}_{0+}$ coincides, modulo $\mathcal{F}_t$, with the germ σ-algebra of $\widetilde{W}$ at 0. By Blumenthal's zero-one law (Proposition 1.6.14), applied to $\widetilde{W}$, every event of $\mathcal{G}_{0+}$ has, conditionally on $\mathcal{F}_t$, probability 0 or 1. This forces A to coincide, up to a $\mathbb{P}_x$-negligible set, with an $\mathcal{F}_t$-measurable event, hence $A \in \mathcal{F}_t$, since $(\mathcal{F}_t)_{t\ge0}$ is complete. Observe that above we have omitted the routine but lengthy verification that $A \in \mathcal{F}_t \vee \mathcal{G}_{0+}$. We refer to Revuz–Yor [7] or Karatzas–Shreve [6] for the complete argument. ■

The (simple) Markov property evaluates the future after a deterministic time s. The reflection principle below, like many of the most useful computations involving Brownian motion, requires evaluating it instead at a stopping time. This extension is the strong Markov property, and it is precisely the right-continuity of Corollary 1.6.15 that lets $\mathcal{F}_S$ (Definition 1.5.12) play, for a stopping time S, the same role that $\mathcal{F}_s$ plays for a deterministic time.

Theorem 1.6.16: Strong Markov property

Let S be a stopping time (Definition 1.5.8) and let $Y : \mathbb{R}_+ \times E \to \mathbb{R}$ be bounded and measurable. Then for every $x \in \mathbb{R}^n$, we have

$$\mathbb{E}_x\Big[Y_S \circ \theta_S \,\Big|\, \mathcal{F}_S\Big]\, \mathbf{1}_{(S<\infty)} \;=\; \mathbb{E}_{W_S}[Y_S]\, \mathbf{1}_{(S<\infty)}, \qquad \mathbb{P}_x\text{-almost surely,} \tag{1.96}$$

where $\mathcal{F}_S$ is the information at time S (Definition 1.5.12). In particular, if $S < \infty$ $\mathbb{P}_x$-almost surely, we get

$$\mathbb{E}_x\Big[Y_S \circ \theta_S \,\Big|\, \mathcal{F}_S\Big] \;=\; \mathbb{E}_{W_S}[Y_S], \qquad \mathbb{P}_x\text{-almost surely.} \tag{1.97}$$

Sketch of proof. As in the proof of Theorem 1.5.14, one first treats stopping times S taking finitely many values $s_1 < \cdots < s_m$ (together, possibly, with the value $+\infty$), for which (1.96) follows by applying the Markov property (Theorem 1.6.10) on each event $(S = s_i) \in \mathcal{F}_{s_i} \subset \mathcal{F}_S$ separately. The general case follows by approximating S from above by a decreasing sequence of dyadic stopping times, and passing to the limit using the right-continuity of $(\mathcal{F}_t)_{t\ge0}$ (Corollary 1.6.15) together with continuity of the paths of W. We omit the details and refer to Karatzas–Shreve [6] or Revuz–Yor [7]. ■

As an application of the strong Markov property, we compute the law of the first hitting time of a level by a real-valued Wiener process: reflecting the path after it first reaches that level produces another Wiener process, from which the reflection principle below follows.

Theorem 1.6.17: Reflection principle

Let $n = 1$, let $a > 0$, and consider the following stopping time (see Proposition 1.5.10):

$$T_a := \inf\{t \geq 0\,;\, W_t = a\}, \tag{1.98}$$

that is the hitting time of a by W. Then for every $t \geq 0$, we have

$$\mathbb{P}_0(T_a < t) \;=\; 2\,\mathbb{P}_0(W_t > a). \tag{1.99}$$

Proof. Step 1: reduction. Since $\mathbb{P}_0(W_t = a) = 0$, splitting the event $(T_a < t)$ according to the value of W_t gives the classical decomposition

$$\mathbb{P}_0(T_a < t) \;=\; \mathbb{P}_0(T_a < t,\, W_t > a) \;+\; \mathbb{P}_0(T_a < t,\, W_t < a). \tag{1.100}$$

In addition, owing to the fact that $W_0 = 0 < a$ almost surely under $\mathbb{P}_0$, continuity of the paths and the intermediate value theorem show that $(W_t > a) \subset (T_a < t)$. In particular

$$\mathbb{P}_0(T_a < t,\, W_t > a) \;=\; \mathbb{P}_0(W_t > a). \tag{1.101}$$

It therefore suffices to show

$$\mathbb{P}_0(T_a < t,\, W_t > a) \;=\; \tfrac{1}{2}\,\mathbb{P}_0(T_a < t). \tag{1.102}$$

Indeed, multiplying (1.102) by 2 gives

$$\mathbb{P}_0(T_a < t) \;=\; 2\,\mathbb{P}_0(T_a < t,\, W_t > a).$$

Hence substituting the identity (1.101) for the term on the right then gives $\mathbb{P}_0(T_a < t) = 2\,\mathbb{P}_0(W_t > a)$, which is (1.99). Combined with the decomposition (1.100), this also shows that the two terms on its right-hand side coincide, both equal to $\frac{1}{2}\,\mathbb{P}_0(T_a < t)$. Otherwise stated, after W has crossed the level a, it is equally likely to end up above or below it at time t. This is the so-called *reflection principle*.

Step 2: a stopping time and a functional. Set (with $\inf \varnothing := \infty$)

$$S := \inf\{s < t\,;\, W_s = a\}. \tag{1.103}$$

By construction, $(S < \infty) = (S < t) = (T_a < t)$, and $S = T_a$ on this event. For $u < t$, we have

$$(S \leq u) = (T_a \leq u) \in \mathcal{F}_u,$$

since T_a is a stopping time. Next for $u \geq t$, we get

$$(S \leq u) = (T_a < t) = \bigcup_{k \geq 1} (T_a \leq t - 1/k) \in \mathcal{F}_t \subset \mathcal{F}_u.$$

Hence S is itself a stopping time. Define

$$Y_s(f) := \mathbf{1}_{(s<t)}\,\mathbf{1}_{(f_{t-s}>a)}, \qquad s \geq 0,\ f \in E, \tag{1.104}$$

which is a bounded and jointly measurable functional in (s, f), since $(s, f) \mapsto f_{t-s}$ is jointly continuous on $[0, t] \times E$. By definition of θ_S (Definition 1.6.5) and since $S < t$ exactly on $(S < \infty) = (T_a < t)$, we end up with

$$Y_S \circ \theta_S \;=\; \mathbf{1}_{(S<t)}\,\mathbf{1}_{(W_{S+(t-S)}>a)} \;=\; \mathbf{1}_{(T_a<t)}\,\mathbf{1}_{(W_t>a)}. \tag{1.105}$$

Step 3: applying the strong Markov property. By continuity of the paths, $W_S = a$ on $(S < \infty) = (T_a < t)$. The strong Markov property (Theorem 1.6.16), applied to S and Y above, gives

$$\mathbb{E}_0\Big[Y_S \circ \theta_S \,\Big|\, \mathcal{F}_S\Big]\,\mathbf{1}_{(S<\infty)} \;=\; \mathbb{E}_{W_S}[Y_S]\,\mathbf{1}_{(S<\infty)} \;=\; \varphi(W_S, S)\,\mathbf{1}_{(S<\infty)}, \tag{1.106}$$

where φ is defined by

$$\varphi(y, s) := \mathbb{E}_y\Big[\mathbf{1}_{(W_{t-s}>a)}\Big]\,\mathbf{1}_{(s<t)}.$$

Now on $(S < \infty)$, we have $W_S = a$. Moreover under $\mathbb{P}_a$, it holds that the conditional law of $W_{t-S} - a$ is identified as $\mathcal{N}(0, t - S)$, which is symmetric about 0. Hence

$$\varphi(W_S, S) \;=\; \mathbb{P}_a(W_{t-S} > a)\,\mathbf{1}_{(S<t)} \;=\; \tfrac{1}{2}\,\mathbf{1}_{(T_a<t)}. \tag{1.107}$$

Step 4: conclusion. Taking $\mathbb{E}_0[\,\cdot\,]$ in (1.106) and applying the tower property (Proposition 1.5.2, with $\mathcal{G} := \mathcal{F}_S$) to the left-hand side, then using (1.105) on the left and (1.107) on the right, we discover that

$$\mathbb{P}_0(T_a < t,\, W_t > a) \;=\; \mathbb{E}_0\Big[Y_S \circ \theta_S\,\mathbf{1}_{(S<\infty)}\Big] \;=\; \mathbb{E}_0\Big[\tfrac{1}{2}\,\mathbf{1}_{(T_a<t)}\Big] \;=\; \tfrac{1}{2}\,\mathbb{P}_0(T_a < t), \tag{1.108}$$

which is (1.102). By Step 1, this completes the proof. ∎

Having used the Markov property to compute explicit path functionals of W, such as its exit and hitting times, we turn, in the last section of this chapter, to the local regularity of the paths themselves: exactly how irregular a typical trajectory of W is, and in what sense.

1.7 Pathwise properties

We close the chapter by studying the paths of W directly, as functions of time: how regular they are (Theorem 1.7.4), and, in the other direction, exactly how irregular (Theorem 1.7.11 and Proposition 1.7.12). The infinite variation of the paths established along the way is the central obstruction that the stochastic integral, built in the next chapter, is designed to overcome. We first recall the relevant notion of regularity for a function of time.

Definition 1.7.1

Let $a < b$ and $\gamma \in (0, 1]$. For $f : [a, b] \to \mathbb{R}^n$, set

$$\|f\|_\gamma := \sup_{s,t\in[a,b],\, s\neq t} \frac{|\delta f_{st}|}{|t-s|^\gamma} \tag{1.109}$$

(see Notation 1.3.1 for δf). We say that f is *γ-Hölder* if $\|f\|_\gamma < \infty$, and write $\mathcal{C}^\gamma([a, b]; \mathbb{R}^n)$ for the space of such functions.

Remark 1.7.2. *$\|\cdot\|_\gamma$ is only a semi-norm, not a norm, on $\mathcal{C}^\gamma([a, b]; \mathbb{R}^n)$: it vanishes on every constant function, not only on $f \equiv 0$, so $\|f\|_\gamma = 0$ does not imply $f = 0$.*

The basic tool for producing Hölder-continuous versions of a process is Kolmogorov's continuity criterion, which we recall without proof.

Theorem 1.7.3: Kolmogorov's continuity criterion

Let $\tau > 0$ and let $X = \{X_t\,;\, t \in [0, \tau]\}$ be an $\mathbb{R}^n$-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ such that, for some $\alpha, \beta, c > 0$,

$$\mathbb{E}\Big[|\delta X_{st}|^\alpha\Big] \le c\,|t-s|^{1+\beta}, \qquad \text{for every } s, t \in [0, \tau]. \tag{1.110}$$

Then X admits a modification $\widehat{X}$ (Definition 1.2.2) such that, almost surely, $\widehat{X} \in \mathcal{C}^\gamma([0, \tau]; \mathbb{R}^n)$ for every $\gamma \in (0, \beta/\alpha)$.

We omit the proof, a chaining argument along dyadic partitions of $[0, \tau]$, and refer to Karatzas–Shreve [6] or Revuz–Yor [7].

Theorem 1.7.4: Hölder regularity of Brownian motion

Let W be an n-dimensional Wiener process, let $\tau > 0$ and $\gamma \in (0, 1/2)$. There exists a modification $\widehat{W}$ of W (restricted to $[0, \tau]$) such that, almost surely, $\widehat{W} \in \mathcal{C}^\gamma([0, \tau]; \mathbb{R}^n)$ (Definition 1.7.1).

Remark 1.7.5. *As is customary, and as we already did in Proposition 1.4.7 and thereafter, we denote $\widehat{W}$ again by W.*

Proof. Step 1: moments of the increments. Fix $m \ge 1$. Since $\delta W_{st} \sim \mathcal{N}(0, (t-s)\mathrm{Id}_{\mathbb{R}^n})$ (Definition 1.3.13), the moments of the Gaussian distribution give

$$\mathbb{E}\Big[|\delta W_{st}|^{2m}\Big] \;=\; c_m\,|t-s|^m \;=\; c_m\,|t-s|^{1+(m-1)}, \tag{1.111}$$

for a constant $c_m > 0$ depending only on m and n.

Step 2: application of Kolmogorov's criterion. Relation (1.111) is (1.110) with $\alpha = 2m$, $\beta = m - 1$, $c = c_m$. Theorem 1.7.3 therefore produces a modification of W on $[0, \tau]$, almost surely γ-Hölder for every

$$\gamma < \frac{m-1}{2m} = \frac{1}{2} - \frac{1}{2m}. \tag{1.112}$$

Letting $m \to \infty$ in (1.112), the right-hand side increases to $1/2$, so W admits an almost surely γ-Hölder modification on $[0, \tau]$ for every $\gamma < 1/2$. ∎

Theorem 1.7.4 shows that W admits an almost surely γ-Hölder modification for every $\gamma < 1/2$, but it says nothing about the borderline exponent $\gamma = 1/2$ itself. The next result closes this gap: it pins down the exact rate at which the increments of W vanish, up to a sharp logarithmic correction. We omit the proof, a refinement of the chaining argument underlying Theorem 1.7.3, and refer to Karatzas–Shreve [6] or Revuz–Yor [7].

Theorem 1.7.6: Lévy's modulus of continuity

Let W be an n-dimensional Wiener process and let $\tau > 0$. Almost surely,

$$\limsup_{\delta \to 0^+} \sup_{0 \le s < t \le \tau,\, |t-s| \le \delta} \frac{|\delta W_{st}|}{\left(2\delta \ln(1/\delta)\right)^{1/2}} = 1. \tag{1.113}$$

Remark 1.7.7. *Theorem 1.7.6 refines Theorem 1.7.4: it says that W has Hölder regularity exactly $1/2$ at every point, up to the logarithmic correction $(\ln(1/\delta))^{1/2}$, which is itself unavoidable.*

We now turn to the opposite question: exactly how irregular the paths of W are. This is best measured through the variation and quadratic variation of a function along a partition.

Definition 1.7.8

Let $a < b$. A *partition* of $[a, b]$ is a finite set $\pi = \{t_0, \dots, t_m\}$ with $a = t_0 < \cdots < t_m = b$; we write $|\pi| := \max_{0 \le k \le m-1}(t_{k+1} - t_k)$ for its mesh, and $\Pi_{a,b}$ for the set of all partitions of $[a, b]$.

Definition 1.7.9: Variation and quadratic variation

Let $a < b$ and $f : [a, b] \to \mathbb{R}$. Recalling Notation 1.3.1, the *variation* and *quadratic variation* of f on $[a, b]$ are repectively given by

$$\begin{aligned} V_{a,b}(f) &= \lim_{\pi \in \Pi_{a,b},\, |\pi| \to 0} \sum_{t_k, t_{k+1} \in \pi} |\delta f_{t_k t_{k+1}}|, \\ V^2_{a,b}(f) &= \lim_{\pi \in \Pi_{a,b},\, |\pi| \to 0} \sum_{t_k, t_{k+1} \in \pi} |\delta f_{t_k t_{k+1}}|^2, \end{aligned} \tag{1.114}$$

when these limits exist in $[0, \infty]$. If $V_{a,b}(f) < \infty$ (respectively $V^2_{a,b}(f) < \infty$), we say that f has *finite variation* (respectively *finite quadratic variation*) on $[a, b]$.

Remark 1.7.10. *For a deterministic function f, the limits in (1.114), when they exist, do not depend on the particular sequence of partitions used, in the usual sense of a limit along the directed set $(\Pi_{a,b}, \supset)$ refining partitions as $|\pi| \to 0$. For a stochastic process, as in Theorem 1.7.11 below, some care is needed: the limit defining $V^2_{a,b}(W)$ is understood in $L^2(\Omega)$, along* any *sequence of partitions with mesh going to 0. On the other hand the resulting almost sure statement is along a suitably chosen subsequence, as the proof makes precise.*

Theorem 1.7.11: Variations of Brownian motion

Let $n = 1$. Almost surely, for every $0 \le a < b < \infty$, it holds

$$V^2_{a,b}(W) = b - a \qquad \text{and} \qquad V_{a,b}(W) = \infty. \tag{1.115}$$

In other words, the paths of W have finite quadratic variation but infinite variation on every interval of $\mathbb{R}_+$.

Proof. Step 1: $L^2(\Omega)$-convergence of $V^2_{a,b}$. Fix $a < b$. For $\pi = \{t_0, \dots, t_m\} \in \Pi_{a,b}$, set

$$S_\pi := \sum_{k=0}^{m-1} |\delta W_{t_k t_{k+1}}|^2, \qquad X_k := |\delta W_{t_k t_{k+1}}|^2 - (t_{k+1} - t_k). \tag{1.116}$$

By independence and centering of the increments of W (Definition 1.3.13), the X_k's are centered and independent, and

$$S_\pi - (b - a) = \sum_{k=0}^{m-1} X_k. \tag{1.117}$$

Since $\delta W_{t_k t_{k+1}} / (t_{k+1} - t_k)^{1/2} \sim \mathcal{N}(0, 1)$, we have

$$\mathrm{Var}(X_k) = 2(t_{k+1} - t_k)^2. \tag{1.118}$$

Thus invoking independence and (1.117), we get

$$\mathbb{E}\Big[(S_\pi - (b-a))^2\Big] \;=\; \sum_{k=0}^{m-1} \mathrm{Var}(X_k) \;=\; 2\sum_{k=0}^{m-1} (t_{k+1}-t_k)^2 \;\le\; 2\,|\pi|\,(b-a). \tag{1.119}$$

Letting $|\pi| \to 0$ in (1.119) gives

$$L^2(\Omega)\text{-}\lim_{|\pi|\to 0} S_\pi \;=\; b-a. \tag{1.120}$$

Step 2: almost sure convergence along a subsequence, and $V_{a,b}(W) = \infty$. By (1.120), there is a sequence of partitions $(\pi_n)_n$, with $|\pi_n| \to 0$, along which $S_{\pi_n} \to b-a$ almost surely; this gives the first identity in (1.115), namely $V^2_{a,b}(W) = b-a$ almost surely, along $(\pi_n)_n$. We now argue by contradiction to show $V_{a,b}(W) = \infty$ almost surely. To this aim, let Ω_0 be the almost sure event on which $S_{\pi_n} \to b-a > 0$, fix $\omega \in \Omega_0$, and suppose $V_{a,b}(W(\omega)) < \infty$. Writing $\pi_n = \{t_0, \dots, t_{m_n}\}$, continuity of $W(\omega)$ on $[a,b]$ gives

$$\sup_{n\ge 1} \sum_{k=0}^{m_n-1} \Big|\delta W_{t_k t_{k+1}}(\omega)\Big| \;\le\; c(\omega) \;<\; \infty, \qquad \lim_{n\to\infty} \max_{0\le k\le m_n-1} \Big|\delta W_{t_k t_{k+1}}(\omega)\Big| \;=\; 0, \tag{1.121}$$

the first bound by the assumption $V_{a,b}(W(\omega)) < \infty$, and the second by uniform continuity of $W(\omega)$ on $[a,b]$ and $|\pi_n| \to 0$. Combining the two bounds in (1.121), we end up with

$$S_{\pi_n}(\omega) \;\le\; \Big(\max_{0\le k\le m_n-1} \Big|\delta W_{t_k t_{k+1}}(\omega)\Big|\Big) \sum_{k=0}^{m_n-1} \Big|\delta W_{t_k t_{k+1}}(\omega)\Big| \;\longrightarrow\; 0, \tag{1.122}$$

contradicting $S_{\pi_n}(\omega) \to b-a > 0$. Hence $V_{a,b}(W(\omega)) = \infty$ for every $\omega \in \Omega_0$, that is $V_{a,b}(W) = \infty$ almost surely, for this fixed pair (a,b).

Step 3: simultaneously for all (a,b). It remains to upgrade "almost surely, for each fixed (a,b)" to "almost surely, for every (a,b)". Since $V_{a,b}(f) = \infty$ for some f and $a < b$ forces $V_{a',b'}(f) = \infty$ for every $a' \le a < b \le b'$ (a partition of $[a',b']$ refined at a, b has variation at least that of its restriction to $[a,b]$), it suffices to treat $(a,b) \in \mathbb{Q}^2_+$, $a < b$. By Step 2, for each such rational pair there is $\Omega_{a,b}$ with $\mathbb{P}(\Omega_{a,b}) = 1$ and $V_{a,b}(W(\omega)) = \infty$ for every $\omega \in \Omega_{a,b}$. Set

$$\Omega_0 \;:=\; \bigcap_{(a,b)\in\mathbb{Q}^2_+,\, a<b} \Omega_{a,b}. \tag{1.123}$$

As a countable intersection of almost sure events, we still have $\mathbb{P}(\Omega_0) = 1$. Moreover, for every $\omega \in \Omega_0$ and every $(a,b) \in \mathbb{Q}^2_+$ with $a < b$, we get $V_{a,b}(W(\omega)) = \infty$. By the monotonicity remark above, this extends to every $0 \le a < b < \infty$, on the same event Ω_0. This proves the second identity in (1.115). ∎

Proposition 1.7.12: Irregularity of W above exponent $1/2$

Let $n = 1$, $\gamma > 1/2$, and $0 \le a < b$. Then, almost surely, $W \notin \mathcal{C}^\gamma([a,b];\mathbb{R})$ (Definition 1.7.1).

Proof. We argue by contradiction, on the same almost sure event Ω_0 and along the same partitions $(\pi_n)_n$ of $[a,b]$ as in the proof of Theorem 1.7.11, on which $S_{\pi_n} \to b-a > 0$. Fix $\omega \in \Omega_0$ and suppose $W(\omega) \in \mathcal{C}^\gamma([a,b];\mathbb{R})$, that is $|\delta W_{st}(\omega)| \le L(\omega)\,|t-s|^\gamma$ for every $s,t \in [a,b]$, for some (random) constant $L(\omega) < \infty$. Writing $\pi_n = \{t_0, \dots, t_{m_n}\}$,

$$S_{\pi_n}(\omega) \;\le\; L(\omega)^2 \sum_{k=0}^{m_n-1} |t_{k+1}-t_k|^{2\gamma} \;\le\; L(\omega)^2\,|\pi_n|^{2\gamma-1}\,(b-a) \;\longrightarrow\; 0, \tag{1.124}$$

since $2\gamma - 1 > 0$ and $|\pi_n| \to 0$. This contradicts $S_{\pi_n}(\omega) \to b-a > 0$, so $W(\omega) \notin \mathcal{C}^\gamma([a,b];\mathbb{R})$ for every $\omega \in \Omega_0$, that is almost surely. ∎

Proposition 1.7.12 rules out γ-Hölder regularity of W, for $\gamma > 1/2$, on a fixed interval. A classical strengthening, due to Paley, Wiener, and Zygmund, upgrades this to a statement at every point simultaneously, and in particular rules out differentiability.

Theorem 1.7.13: Nowhere differentiability

Let $n = 1$ and $\tau > 0$. Almost surely,

(i) for every $\gamma > 1/2$, the paths of W are not γ-Hölder continuous at any point $s \in [0,\tau]$, in the sense that

$$\limsup_{h \to 0} \frac{|\delta W_{s,s+h}|}{|h|^\gamma} \;=\; \infty, \qquad \text{for every } s \in [0,\tau];$$

(ii) in particular, W is nowhere differentiable on $[0,\tau]$.

Sketch of proof. Proposition 1.7.12 already shows that, for a *fixed* s, the paths of W are almost surely not γ-Hölder at s, for $\gamma > 1/2$; the content of the theorem is that the exceptional (negligible) event can be taken independent of s. This requires a more delicate, uniform version of the argument in the proof of Proposition 1.7.12, controlling the increments of W over all sufficiently small intervals simultaneously, rather than one (a,b) at a time. We omit the details, and refer to Karatzas–Shreve [6] or Revuz–Yor [7]. Part (ii) follows from part (i) applied with $\gamma = 1$: existence of W'_s at some s would give $|\delta W_{s,s+h}| \le c\,|h|$ for small h, contradicting part (i) with any $\gamma \in (1/2, 1)$. ∎

We close this chapter with a brief look into the next chapter. Theorem 1.7.11 showed that the paths of W have infinite variation on every interval: consequently, integrals

of the form $\int_0^t f_s \, dW_s$ cannot be defined path by path as Riemann–Stieltjes integrals, for a general integrand f, the way one integrates against a function of finite variation. Making sense of such integrals rigorously, and of the calculus that comes with them (Itô's formula), is the subject of the next chapter. This will rely on the finite quadratic variation of Theorem 1.7.11.

CHAPTER 2

Itô's formula

Chapter 1 constructed Brownian motion W and established its two most consequential pathwise features: W has finite quadratic variation but infinite variation on every interval (Theorem 1.7.11), and W is almost surely nowhere differentiable (Theorem 1.7.13). Both facts point to the same problem. Namely for a smooth (or even merely finite-variation) function f, one can define $\int_0^t u_s\, df_s$ path by path as a Riemann–Stieltjes integral. In addition, differential calculus with respect to f follows from the ordinary chain rule applied to $\dot{f}$. Neither route is available for W: the Riemann-Stieltjes construction diverges because W has infinite variation, and there is no $\dot{W}$ to differentiate against in the first place.

This chapter builds the calculus that replaces these classical tools. In the present section we construct the Itô integral $\int_0^t u_s\, dW_s$, for a suitable class of random, time-dependent integrands u, as an $L^2(\Omega)$-limit of elementary Riemann sums. The finite quadratic variation of W (Theorem 1.7.11) is exactly what makes this limit exist. Later sections turn this integral into a full differential calculus: Itô's formula, the chain rule adapted to functions of W, and its main consequences.

For notational simplicity we carry out the construction for a real-valued ($n = 1$) Wiener process W. The extension to an $\mathbb{R}^n$-valued $W = (W^{(1)}, \dots, W^{(n)})$ (Definition 1.3.13) is obtained by working coordinate by coordinate and is only a matter of extra indices.

2.1 Itô's integral

Throughout this section we fix $\tau > 0$ and a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\tau]}, \mathbb{P})$ (Definition 1.2.4) carrying a real-valued Wiener process W with respect to $(\mathcal{F}_t)_{t\in[0,\tau]}$, in the sense of Definition 1.3.15. We construct the integral $\int_0^t u_s\, dW_s$ in two steps: first define it explicitly on a class of elementary integrands, then pass to the limit.

2.1.1 Simple processes and their integral

We first define the integral for integrands that are piecewise constant in time and adapted to the filtration.

Definition 2.1.1: Simple process

A process $u = \{u_t\,;\, t \in [0,\tau]\}$ is *simple* if there exist a partition $\pi = \{t_0, \dots, t_N\} \in \Pi_{0,\tau}$ (Definition 1.7.8), a constant $c > 0$, and random variables $\xi_0, \dots, \xi_{N-1}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, with ξ_i $\mathcal{F}_{t_i}$-measurable and $|\xi_i| \le c$ almost surely for every $0 \le i \le N-1$, such that

$$u_t \;=\; \xi_0\, \mathbf{1}_{\{0\}}(t) \;+\; \sum_{i=0}^{N-1} \xi_i\, \mathbf{1}_{(t_i, t_{i+1}]}(t), \qquad t \in [0,\tau]. \tag{2.1}$$

We denote by $\mathcal{E}([0,\tau])$ the set of simple processes.

Remark 2.1.2. *$\mathcal{E}([0,\tau])$ is a vector space. If $u, v \in \mathcal{E}([0,\tau])$ are of the form (2.1) relative to partitions $\pi, \pi' \in \Pi_{0,\tau}$ respectively, both are also of this form relative to the common refinement $\pi \cup \pi'$. Hence we have that $\alpha u + \beta v \in \mathcal{E}([0,\tau])$ for every $\alpha, \beta \in \mathbb{R}$.*

We now define the integral of a simple process over a subinterval $[s,t] \subset [0,\tau]$, as the corresponding sum against the increments of W (Notation 1.3.1), taken along any partition refining both the partition underlying u and the endpoints s, t.

Definition 2.1.3: Integral of a simple process

Let $u \in \mathcal{E}([0,\tau])$ be given by (2.1) relative to $\pi = \{t_0, \dots, t_N\} \in \Pi_{0,\tau}$, and let $s, t \in [0,\tau]$ with $s \le t$. Let

$$\{r_0, \dots, r_K\} := \pi \cup \{s, t\},$$

listed in increasing order, and let $k_0 \le k_1$ be the indices with $r_{k_0} = s$ and $r_{k_1} = t$. Since u is constant, equal to some $\xi_{i(k)}$, on each subinterval $(r_k, r_{k+1}]$ of the refined partition, we define

$$\mathcal{J}_{s,t}(u\, dW) \;:=\; \sum_{k=k_0}^{k_1-1} \xi_{i(k)}\, \delta W_{r_k, r_{k+1}}, \tag{2.2}$$

with the convention $\mathcal{J}_{s,t}(u\, dW) := 0$ if $s = t$.

Remark 2.1.4. *Definition 2.1.3 exhibits $\mathcal{J}_{s,t}(u\, dW)$ as a genuine Riemann–Stieltjes sum of u against W, along the partition $\{r_{k_0}, \dots, r_{k_1}\}$ of $[s,t]$: no limiting procedure is needed at this stage. The value does not depend on the representation (2.1) chosen for u: refining the underlying partition π further only subdivides some of the terms in (2.2) into telescoping sub-sums $\sum \delta W_{r_k r_{k+1}} = \delta W$ of the undivided interval. This leaves the total unchanged. In particular, u and a second simple process v can always be represented on a common partition, for instance $\pi \cup \pi'$, which we use freely below.*

We collect the elementary algebraic and probabilistic properties of $\mathcal{J}_{s,t}(u\, dW)$ that will be preserved, in the next subsection, when passing to the limit.

Proposition 2.1.5

Let $u, v \in \mathcal{E}([0,\tau])$ and $\alpha, \beta \in \mathbb{R}$. Fix $s, t \in [0,\tau]$ with $s \le t$, and recall that $\mathcal{J}_{s,t}(u\,dW)$ denotes the integral of u against W on $[s,t]$, defined by (2.2). Then, almost surely,

(i) $\mathcal{J}_{t,t}(u\,dW) = 0$.

(ii) $\mathcal{J}_{s,t}\Big((\alpha u + \beta v)\,dW\Big) = \alpha\,\mathcal{J}_{s,t}(u\,dW) + \beta\,\mathcal{J}_{s,t}(v\,dW)$.

(iii) $\mathbb{E}\Big[\mathcal{J}_{s,t}(u\,dW)\,\Big|\,\mathcal{F}_s\Big] = 0$.

(iv) It holds

$$\mathbb{E}\Big[\mathcal{J}_{s,t}(u\,dW)^2\,\Big|\,\mathcal{F}_s\Big] \;=\; \int_s^t \mathbb{E}\Big[u_r^2\,\Big|\,\mathcal{F}_s\Big]\,dr. \tag{2.3}$$

(v) In particular,

$$\mathbb{E}\Big[\mathcal{J}_{s,t}(u\,dW)^2\Big] \;=\; \int_s^t \mathbb{E}\Big[u_r^2\Big]\,dr. \tag{2.4}$$

Proof. Step 1: properties (i) and (ii). Property (i) is the convention adopted in Definition 2.1.3 when $s = t$. For property (ii), represent u and v on the common refinement $\pi \cup \pi'$ of the partitions underlying them (Remark 2.1.4). On this refinement, $\alpha u + \beta v$ is simple with coefficients $\alpha\xi_i + \beta\xi'_i$, and formula (2.2) is linear in these coefficients, which gives property (ii).

Step 2: martingale property (iii). By Definition 2.1.3, one can write

$$\mathcal{J}_{s,t}(u\,dW) \;=\; \sum_{k=k_0}^{k_1-1} \eta_k\,\delta W_{r_k,r_{k+1}}, \qquad \eta_k := \xi_{i(k)}, \tag{2.5}$$

with $r_{k_0} = s$. Since η_{k_0} is $\mathcal{F}_s$-measurable and $\delta W_{s,r_{k_0+1}} \perp\!\!\!\perp \mathcal{F}_s$ (property (ii) of Definition 1.3.15), relation (1.54) gives

$$\mathbb{E}\Big[\eta_{k_0}\,\delta W_{s,r_{k_0+1}}\,\Big|\,\mathcal{F}_s\Big] \;=\; \eta_{k_0}\,\mathbb{E}\Big[\delta W_{s,r_{k_0+1}}\Big] \;=\; 0. \tag{2.6}$$

Now for $k_0 < k \le k_1 - 1$, we have $r_k > s$. Hence we get $\mathcal{F}_s \subset \mathcal{F}_{r_k}$, and the nested tower property (1.55) gives

$$\mathbb{E}\Big[\eta_k\,\delta W_{r_k,r_{k+1}}\,\Big|\,\mathcal{F}_s\Big] \;=\; \mathbb{E}\Big[\mathbb{E}\Big[\eta_k\,\delta W_{r_k,r_{k+1}}\,\Big|\,\mathcal{F}_{r_k}\Big]\,\Big|\,\mathcal{F}_s\Big] \;=\; 0, \tag{2.7}$$

the inner conditional expectation vanishing by the same computation as in (2.6). Thus summing (2.6) and the vanishing terms (2.7), over $k = k_0 + 1, \dots, k_1 - 1$, gives property (iii).

Step 3: isometries (iv) and (v). Squaring (2.5), we obtain

$$\mathcal{J}_{s,t}(u\,dW)^2 \;=\; \sum_{k=k_0}^{k_1-1} \eta_k^2\,\delta W_{r_k,r_{k+1}}^2 \;+\; 2 \sum_{k_0\le k<l\le k_1-1} \eta_k\,\delta W_{r_k,r_{k+1}}\,\eta_l\,\delta W_{r_l,r_{l+1}}. \tag{2.8}$$

For $k < l$ in (2.8), both $\eta_k\,\delta W_{r_k,r_{k+1}}$ and η_l are $\mathcal{F}_{r_l}$-measurable, and $\delta W_{r_l,r_{l+1}} \perp\!\!\!\perp \mathcal{F}_{r_l}$, so the same computation as in (2.7), with $\mathcal{F}_{r_l}$ in place of $\mathcal{F}_{r_k}$, gives

$$\mathbb{E}\Big[\eta_k\,\delta W_{r_k,r_{k+1}}\,\eta_l\,\delta W_{r_l,r_{l+1}}\,\Big|\,\mathcal{F}_s\Big] \;=\; 0, \tag{2.9}$$

so the cross terms in (2.8) drop out after conditioning on $\mathcal{F}_s$. For the diagonal terms, η_k^2 is $\mathcal{F}_{r_k}$-measurable and $\delta W_{r_k,r_{k+1}} \perp\!\!\!\perp \mathcal{F}_{r_k}$ with $\delta W_{r_k,r_{k+1}} \sim \mathcal{N}(0, r_{k+1}-r_k)$. Hence repeating the nested tower argument of (2.7), we discover that

$$\mathbb{E}\Big[\eta_k^2\,\delta W_{r_k,r_{k+1}}^2\,\Big|\,\mathcal{F}_s\Big] \;=\; \mathbb{E}\Big[\eta_k^2\,\Big|\,\mathcal{F}_s\Big]\,(r_{k+1}-r_k) \;=\; \int_{r_k}^{r_{k+1}} \mathbb{E}\Big[u_r^2\,\Big|\,\mathcal{F}_s\Big]\,dr, \tag{2.10}$$

where the last identity stems from the fact that $u_r = \eta_k$ for $r \in (r_k, r_{k+1}]$. Summing (2.10) over $k = k_0, \dots, k_1 - 1$ gives (2.3). Furthermore, taking the expectation of (2.3) and invoking the tower property (1.52) we conclude that (2.4) holds true. This ends the proof. ■

2.1.2 The space L^2_a and density of simple processes

Simple processes are too restrictive a class of integrands for later applications, where typically $u_t = f(W_t)$ for a smooth function f. We now identify the class of integrands to which the integral of Definition 2.1.3 extends, and show that every such integrand is an L^2-limit of simple processes.

Definition 2.1.6: Progressive measurability

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\tau]}, \mathbb{P})$ be a stochastic basis (Definition 1.2.4) and let $u = \{u_t\,;\, t \in [0,\tau]\}$ be a real-valued stochastic process on it (Definition 1.2.1). We say that u is *progressively measurable* if, for every $t \in [0,\tau]$, the map

$$(r,\omega) \longmapsto u_r(\omega), \qquad [0,t]\times\Omega \longrightarrow \mathbb{R}, \tag{2.11}$$

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

Remark 2.1.7. *Progressive measurability is a mild strengthening of adaptedness (Definition 1.2.6): restricting (2.11) to the slice $\{t\}\times\Omega$ shows that a progressively measurable process is in particular $\mathcal{F}_t$-adapted, and the extra content of (2.11) is the joint measurability, in (r,ω). This joint measurability is needed to make sense of pathwise integrals such as $\int_0^t u_r(\omega)\,dr$. We record two standard sufficient conditions, both established, e.g., in Karatzas–Shreve [6]:*

(i) *Every $\mathcal{F}_t$-adapted process with right-continuous paths is progressively measurable. In particular every simple process $u \in \mathcal{E}([0,\tau])$ (Definition 2.1.1) is progressively measurable, since its paths are right-continuous by construction.*

(ii) *Every deterministic, Borel-measurable function $h : [0,\tau] \to \mathbb{R}$, viewed as a process constant in ω, is progressively measurable with respect to any filtration.*

With progressive measurability in hand, we can now specify precisely the class of integrands to which we extend the Itô integral: processes that are progressively measurable and square-integrable in the sense of (2.12) below.

Definition 2.1.8: The space L_a^2

We denote by $L_a^2([0,\tau])$ the set of real-valued, progressively measurable processes $u = \{u_t\,;\, t \in [0,\tau]\}$ (Definition 2.1.6) such that

$$\|u\|_{L_a^2}^2 := \int_0^\tau \mathbb{E}\big[u_s^2\big]\, ds \,<\, \infty. \tag{2.12}$$

Remark 2.1.9. *By Remark 2.1.7, it is readily checked that $\mathcal{E}([0,\tau]) \subset L_a^2([0,\tau])$. Moreover, $\|\cdot\|_{L_a^2}$ is the norm induced on $L_a^2([0,\tau])$ by $L^2(\Omega \times [0,\tau])$, for $\Omega \times [0,\tau]$ equipped with the finite measure $\mathbb{P} \otimes \mathrm{Leb}$.*

Proposition 2.1.10: Density of simple processes

$\mathcal{E}([0,\tau])$ is dense in $L_a^2([0,\tau])$: for every $u \in L_a^2([0,\tau])$, there exists a sequence $(u^n)_{n\geq 0}$ in $\mathcal{E}([0,\tau])$ such that

$$\lim_{n\to\infty} \|u - u^n\|_{L_a^2} \,=\, 0. \tag{2.13}$$

We omit the proof of Proposition 2.1.10 in this generality, a standard truncation-and-mollification argument, and refer to Karatzas–Shreve [6] or Revuz–Yor [7]. We record instead, and prove, a concrete approximation valid under additional boundedness and continuity hypotheses, which covers every case we use in practice.

Proposition 2.1.11: Concrete approximation

Let $u \in L_a^2([0,\tau])$ be such that $|u_t| \leq M$ almost surely, for every $t \in [0,\tau]$ and some constant $M > 0$. We also assume that u has almost surely continuous paths. Let $(\pi^m)_{m\geq 1}$, $\pi^m = \{s_0^m, \dots, s_{k_m}^m\} \in \Pi_{0,\tau}$ (Definition 1.7.8), be a sequence of partitions with $|\pi^m| \to 0$, and set

$$u_t^m := \sum_{j=0}^{k_m-1} u_{s_j^m}\, \mathbf{1}_{[s_j^m, s_{j+1}^m)}(t), \qquad t \in [0,\tau]. \tag{2.14}$$

Then $u^m \in \mathcal{E}([0,\tau])$ for every $m \geq 1$, and

$$\lim_{m\to\infty} \|u - u^m\|_{L_a^2} \,=\, 0. \tag{2.15}$$

Proof. Step 1: $u^m \in \mathcal{E}([0,\tau])$. Each coefficient $u_{s_j^m}$ is $\mathcal{F}_{s_j^m}$-measurable, since u is adapted (Remark 2.1.7), and bounded by M, so (2.14) is of the form (2.1) relative to π^m.

Step 2: pointwise convergence. Fix ω in the almost sure event on which $t \mapsto u_t(\omega)$ is continuous. For $t \in [0,\tau)$ and $m \geq 1$, let $s_j^m = s_j^m(t)$ be the unique point of π^m with $t \in [s_j^m, s_{j+1}^m)$. Then $|t - s_j^m| \leq |\pi^m| \to 0$, so

$$u_t^m(\omega) \;=\; u_{s_j^m}(\omega) \;\longrightarrow\; u_t(\omega), \qquad m \to \infty, \tag{2.16}$$

by continuity of $u(\omega)$. This holds for every $t \in [0,\tau)$ and almost every ω, hence for Lebesgue-almost every $(t,\omega) \in [0,\tau] \times \Omega$.

Step 3: dominated convergence. By construction, $|u_t^m| \leq M$ for every $t \in [0,\tau]$ and $m \geq 1$, so $|u_t - u_t^m|^2 \leq 4M^2$, a constant integrable on the finite measure space $(\Omega \times [0,\tau], \mathbb{P} \otimes \mathrm{Leb})$. Combining this bound with the almost everywhere convergence of Step 2, dominated convergence gives

$$\|u - u^m\|_{L_a^2}^2 \;=\; \int_0^\tau \mathbb{E}\Big[|u_s - u_s^m|^2\Big]\, ds \;\longrightarrow\; 0, \qquad m \to \infty, \tag{2.17}$$

which proves our claim (2.15). ■

With Proposition 2.1.10 in hand, we extend the integral of Definition 2.1.3 from $\mathcal{E}([0,\tau])$ to all of $L_a^2([0,\tau])$, by continuity.

Proposition 2.1.12: Extension of the integral

Let $u \in L_a^2([0,\tau])$ and let $(u^n)_{n\geq 0}$ be a sequence in $\mathcal{E}([0,\tau])$ such that $\|u - u^n\|_{L_a^2} \to 0$ (see Proposition 2.1.10). Fix $s,t \in [0,\tau]$ with $s \leq t$. Then the following holds true:

(i) The sequence $(\mathcal{J}_{s,t}(u^n\, dW))_{n\geq 0}$ converges in $L^2(\Omega)$.

(ii) The limit in part (i) does not depend on the approximating sequence $(u^n)_{n\geq 0}$.

Proof. Step 1: convergence. By Remark 2.1.2 and linearity (property (ii) of Proposition 2.1.5), $u^n - u^p \in \mathcal{E}([0,\tau])$ and

$$\mathcal{J}_{s,t}\Big((u^n - u^p)\, dW\Big) = \mathcal{J}_{s,t}(u^n\, dW) - \mathcal{J}_{s,t}(u^p\, dW),$$

for every $n,p \geq 0$. The isometry (2.4) then gives

$$\mathbb{E}\Big[\Big(\mathcal{J}_{s,t}(u^n\, dW) - \mathcal{J}_{s,t}(u^p\, dW)\Big)^2\Big] \;=\; \int_s^t \mathbb{E}\Big[(u_r^n - u_r^p)^2\Big]\, dr \;\leq\; \|u^n - u^p\|_{L_a^2}^2. \tag{2.18}$$

In addition, relation (2.18) shows that $(\mathcal{J}_{s,t}(u^n\, dW))_{n\geq 0}$ is a Cauchy sequence in $L^2(\Omega)$. Completeness of $L^2(\Omega)$ thus easily yields part (i).

Step 2: independence of the sequence. Let $(v^n)_{n\geq 0}$ be another sequence in $\mathcal{E}([0,\tau])$ with $\|u - v^n\|_{L_a^2} \to 0$. Denote by $Z^{(u)}$ and $Z^{(v)}$ the respective $L^2(\Omega)$-limits of $\mathcal{J}_{s,t}(u^n\, dW)$ and $\mathcal{J}_{s,t}(v^n\, dW)$, granted by Step 1. Interlacing the two sequences into $w^{2n} := u^n$, $w^{2n+1} := v^n$ gives a single sequence in $\mathcal{E}([0,\tau])$ with $\|u - w^n\|_{L_a^2} \to 0$, so $(\mathcal{J}_{s,t}(w^n\, dW))_{n\geq 0}$ converges in $L^2(\Omega)$ by Step 1. Its two subsequences $(\mathcal{J}_{s,t}(u^n\, dW))_{n\geq 0}$ and $(\mathcal{J}_{s,t}(v^n\, dW))_{n\geq 0}$ then converge to the same limit, that is $Z^{(u)} = Z^{(v)}$ almost surely. ■

2.1.3 The stochastic integral

We can now state the definition toward which this section has been building.

Definition 2.1.13: Itô integral

Let $u \in L^2_a([0,\tau])$ (Definition 2.1.8) and let $s,t \in [0,\tau]$ with $s \le t$. The *Itô integral* of u against W, on $[s,t]$, is the element of $L^2(\Omega)$ given by

$$\mathcal{J}_{s,t}(u\,dW) \;:=\; L^2(\Omega)\text{-}\lim_{n\to\infty} \mathcal{J}_{s,t}(u^n\,dW), \tag{2.19}$$

for an arbitrary sequence $(u^n)_{n\ge 0}$ in $\mathcal{E}([0,\tau])$ with $\|u-u^n\|_{L^2_a} \to 0$ (Proposition 2.1.10). Recall that the limit exists and does not depend on the choice of $(u^n)_{n\ge 0}$ owing to Proposition 2.1.12.

Notation 2.1.14

From now on we write $\int_s^t u_r\,dW_r := \mathcal{J}_{s,t}(u\,dW)$, for $u \in L^2_a([0,\tau])$ (Definition 2.1.13) and $0 \le s \le t \le \tau$. We also abbreviate $\int_0^t u_r\,dW_r$ by $\int_0^t u\,dW$ when convenient.

Remark 2.1.15 (Additivity in time)**.** *Let $u \in L^2_a([0,\tau])$ and $0 \le s \le t \le r \le \tau$. Then, almost surely,*

$$\int_s^t u\,dW \;+\; \int_t^r u\,dW \;=\; \int_s^r u\,dW. \tag{2.20}$$

Indeed, for $u \in \mathcal{E}([0,\tau])$ simple, both sides of (2.20) are, by Definition 2.1.3, sums of the form (2.2) along the common refined partition $\pi \cup \{s,t,r\}$: the two sums on the left concatenate exactly at the index k with $r_k = t$, and together they reconstitute the single sum on the right. This proves (2.20) for simple integrands. For general $u \in L^2_a([0,\tau])$, approximate u by $u^n \in \mathcal{E}([0,\tau])$ as in Definition 2.1.13. By that same definition, the same sequence $(u^n)_{n\ge 0}$ gives $\int_s^t u^n\,dW \to \int_s^t u\,dW$, $\int_t^r u^n\,dW \to \int_t^r u\,dW$, and $\int_s^r u^n\,dW \to \int_s^r u\,dW$, all in $L^2(\Omega)$. Letting $n \to \infty$ in the simple-process identity (2.20) therefore extends it to u.

Remark 2.1.16 (Kiyosi Itô, 1915–2008)**.** *The construction of Definition 2.1.13 is due to the Japanese mathematician Kiyosi Itô, who introduced it in 1944. Working at the time for the Japanese National Statistical Bureau, and largely isolated from the international mathematical community during the Second World War, Itô developed the stochastic integral, and the differential calculus built on it in the rest of this chapter, essentially on his own.*

The properties of Proposition 2.1.5 pass to the limit unchanged. We recall them below for further use.

Proposition 2.1.17: Properties of the Itô integral

Let $u, v \in L^2_a([0,\tau])$ and $\alpha, \beta \in \mathbb{R}$. Fix $s, t \in [0,\tau]$ with $s \le t$, and recall that $\int_s^t u\,dW$ denotes the Itô integral of u against W on $[s,t]$, defined by Notation 2.1.14. Then, almost surely,

(i) $\int_t^t u\,dW = 0$.

(ii) $\int_s^t (\alpha u + \beta v)\,dW = \alpha \int_s^t u\,dW + \beta \int_s^t v\,dW$.

(iii) $\mathbb{E}\Big[\int_s^t u\,dW \,\Big|\, \mathcal{F}_s\Big] = 0$.

(iv) We have

$$\mathbb{E}\Big[\Big(\int_s^t u\,dW\Big)^2 \,\Big|\, \mathcal{F}_s\Big] = \int_s^t \mathbb{E}\Big[u_r^2 \,\Big|\, \mathcal{F}_s\Big]\,dr. \tag{2.21}$$

(v) In particular, the following holds true:

$$\mathbb{E}\Big[\Big(\int_s^t u\,dW\Big)^2\Big] = \int_s^t \mathbb{E}\Big[u_r^2\Big]\,dr. \tag{2.22}$$

Proof. Step 1: properties (i)-(iii). Fix approximating sequences $(u^n)_{n\ge0}$, $(v^n)_{n\ge0}$ in $\mathcal{E}([0,\tau])$ for u, v as in Definition 2.1.13. Property (i) holds for every u^n by Proposition 2.1.5. Therefore it also holds for its $L^2(\Omega)$-limit $\int_t^t u\,dW$. Property (ii) follows since $\alpha u^n + \beta v^n \to \alpha u + \beta v$ in L^2_a and, by linearity of the $L^2(\Omega)$-limit,

$$\begin{aligned}\int_s^t (\alpha u + \beta v)\,dW &= L^2(\Omega)\text{-}\lim_n \mathcal{J}_{s,t}\Big((\alpha u^n + \beta v^n)\,dW\Big) \\ &= L^2(\Omega)\text{-}\lim_n \Big(\alpha\,\mathcal{J}_{s,t}(u^n\,dW) + \beta\,\mathcal{J}_{s,t}(v^n\,dW)\Big),\end{aligned} \tag{2.23}$$

where we have invoked property (ii) of Proposition 2.1.5 for the simple processes u^n, v^n. The right-hand side of (2.23) equals $\alpha \int_s^t u\,dW + \beta \int_s^t v\,dW$, which gives property (ii). For property (iii), conditional expectation given $\mathcal{F}_s$ is a contraction on $L^2(\Omega)$, since

$$\|\mathbb{E}[X \,|\, \mathcal{F}_s]\|_{L^2(\Omega)} \le \|X\|_{L^2(\Omega)}$$

thanks to Jensen's inequality. Hence $\mathcal{J}_{s,t}(u^n\,dW) \to \int_s^t u\,dW$ in $L^2(\Omega)$ gives

$$\mathbb{E}[\mathcal{J}_{s,t}(u^n\,dW) \,|\, \mathcal{F}_s] \longrightarrow \mathbb{E}\Big[\int_s^t u\,dW \,|\, \mathcal{F}_s\Big], \quad \text{in } L^2(\Omega).$$

Since the left-hand side above is identically 0 by Proposition 2.1.5, so is the limit. This proves property (iii).

Step 2: unconditional isometry (v). We have already seen that

$$\mathcal{J}_{s,t}(u^n\,dW) \longrightarrow \int_s^t u\,dW, \quad \text{in } L^2(\Omega).$$

Therefore it is immediate to see that

$$\mathbb{E}[\mathcal{J}_{s,t}(u^n\,dW)^2] \longrightarrow \mathbb{E}[(\int_s^t u\,dW)^2].$$

Furthermore the relation $\|u^n - u\|_{L^2_a} \to 0$ gives, in the same way,

$$\int_s^t \mathbb{E}[(u_r^n)^2]\,dr \longrightarrow \int_s^t \mathbb{E}[u_r^2]\,dr.$$

Invoking the last 2 relations and passing to the limit in the identity (2.4) for u^n, we end up with (2.22).

Step 3: conditional isometry (iv). For $X, Y \in L^2(\Omega)$, the identity $X^2 - Y^2 = (X-Y)(X+Y)$ and the Cauchy–Schwarz inequality give

$$\mathbb{E}\Big[|X^2 - Y^2|\Big] \;\le\; \|X - Y\|_{L^2(\Omega)}\,\|X + Y\|_{L^2(\Omega)}, \tag{2.24}$$

so $X \mapsto X^2$ is continuous from $L^2(\Omega)$ into $L^1(\Omega)$. Applying (2.24) with $X = \mathcal{J}_{s,t}(u^n\,dW)$, $Y = \int_s^t u\,dW$, using that conditional expectation is a contraction on $L^1(\Omega)$, and substituting the left-hand side via the isometry (2.3) for u^n, gives

$$\int_s^t \mathbb{E}\Big[(u_r^n)^2 \,\Big|\, \mathcal{F}_s\Big]\,dr \overset{L^1(\Omega)}{\longrightarrow} \mathbb{E}\Big[\Big(\int_s^t u\,dW\Big)^2 \,\Big|\, \mathcal{F}_s\Big]. \tag{2.25}$$

It remains to show that the left-hand side of (2.25) also converges in $L^1(\Omega)$ to $\int_s^t \mathbb{E}[u_r^2 \,|\, \mathcal{F}_s]\,dr$. To this aim, for each fixed $r \in [s,t]$, applying (2.24) with $X = u_r^n$, $Y = u_r$, and then using the contraction of conditional expectation on $L^1(\Omega)$, we obtain

$$\mathbb{E}\Big[\Big|\mathbb{E}\Big[(u_r^n)^2 \,\Big|\, \mathcal{F}_s\Big] - \mathbb{E}\Big[u_r^2 \,\Big|\, \mathcal{F}_s\Big]\Big|\Big] \;\le\; \|u_r^n - u_r\|_{L^2(\Omega)}\,\|u_r^n + u_r\|_{L^2(\Omega)}. \tag{2.26}$$

Integrating (2.26) over $r \in [s,t]$, resorting to Fubini's theorem to exchange $\mathbb{E}$ and $\int$, and then applying Cauchy–Schwarz to the right-hand side, gives

$$\mathbb{E}\Big[\Big|\int_s^t \mathbb{E}\Big[(u_r^n)^2 \,\Big|\, \mathcal{F}_s\Big]\,dr - \int_s^t \mathbb{E}\Big[u_r^2 \,\Big|\, \mathcal{F}_s\Big]\,dr\Big|\Big] \;\le\; \|u^n - u\|_{L^2_a}\,\|u^n + u\|_{L^2_a}. \tag{2.27}$$

Since $\|u^n - u\|_{L^2_a} \to 0$ and $\|u^n + u\|_{L^2_a}$ stays bounded, the right-hand side of (2.27) tends to 0, so the left-hand side of (2.25) converges in $L^1(\Omega)$ to $\int_s^t \mathbb{E}[u_r^2 \,|\, \mathcal{F}_s]\,dr$. Comparing with the right-hand side of (2.25) gives (2.21). ∎

Remark 2.1.18 (Itô isometry)**.** *Relation* (2.22) *is known as the* Itô isometry*: for fixed* $t \in [0,\tau]$*, the map* $u \mapsto \int_0^t u\,dW$ *is a linear isometry from* $\big(L^2_a([0,t]), \|\cdot\|_{L^2_a}\big)$ *into* $L^2(\Omega)$*. It is the source of most* L^2*-estimates in stochastic calculus, and we invoke it repeatedly starting in the next chapter.*

2.1.4 The Wiener integral

We close this section with the special case of deterministic integrands, which produces Gaussian random variables amenable to explicit computation, and which reappears when we study the Gaussian structure of solutions of linear stochastic differential equations.

Proposition 2.1.19: Wiener integral

Let $h^1, h^2 \in L^2([0,\tau])$ be deterministic functions, viewed as elements of $L^2_a([0,\tau])$ (part (ii) of Remark 2.1.7), and set $W(h^i) := \int_0^\tau h^i_r \, dW_r$ (Notation 2.1.14), for $i = 1, 2$. Then:

(i) $(W(h^1), W(h^2))$ is a centered Gaussian vector in $\mathbb{R}^2$.

(ii) The covariance of this Gaussian vector reads

$$\mathbb{E}\Big[W(h^1)\, W(h^2)\Big] \;=\; \int_0^\tau h^1_r \, h^2_r \, dr. \tag{2.28}$$

Proof. Step 1: approximation. For $m \geq 1$, set $s_j = s^m_j := j\tau/m$, $0 \leq j \leq m$, and

$$h^{i,m} \;:=\; \sum_{j=0}^{m-1} h^i_{s_j} \, \mathbf{1}_{[s_j, s_{j+1})}, \qquad i = 1, 2. \tag{2.29}$$

Since $h^i \in L^2([0,\tau])$, we have $\|h^i - h^{i,m}\|_{L^2_a} \to 0$ as $m \to \infty$. Indeed, this is a slight elaboration of the deterministic case of Proposition 2.1.11. Hence by Definition 2.1.13, we have $W(h^{i,m}) \to W(h^i)$ in $L^2(\Omega)$.

Step 2: the approximating vector is Gaussian. By Definition 2.1.3,

$$W(h^{i,m}) \;=\; \sum_{j=0}^{m-1} h^i_{s_j} \, \delta W_{s_j, s_{j+1}}, \qquad i = 1, 2. \tag{2.30}$$

The increments $\delta W_{s_0 s_1}, \ldots, \delta W_{s_{m-1} s_m}$ are independent, centered Gaussian random variables (property (iii) of Definition 1.3.15), so every linear combination of them is Gaussian: in particular, every linear combination of $W(h^{1,m})$ and $W(h^{2,m})$ in (2.30) is such a linear combination, so $(W(h^{1,m}), W(h^{2,m}))$ is a centered Gaussian vector in $\mathbb{R}^2$ (Definition 1.4.1). Its covariance is

$$\mathbb{E}\Big[W(h^{1,m})\, W(h^{2,m})\Big] \;=\; \sum_{i,j=0}^{m-1} h^1_{s_i} \, h^2_{s_j} \, \mathbb{E}\Big[\delta W_{s_i s_{i+1}} \, \delta W_{s_j s_{j+1}}\Big]. \tag{2.31}$$

For $i < j$, $\delta W_{s_i s_{i+1}}$ is $\mathcal{F}_{s_j}$-measurable (Definition 1.3.15, property (v), since $s_{i+1} \leq s_j$), and $\delta W_{s_j s_{j+1}} \perp\!\!\!\perp \mathcal{F}_{s_j}$. So we easily get relation (1.54),

$$\mathbb{E}\Big[\delta W_{s_i s_{i+1}} \, \delta W_{s_j s_{j+1}}\Big] \;=\; 0, \tag{2.32}$$

and by symmetry the same holds for $i > j$. For $i = j$, $\mathbb{E}[\delta W^2_{s_i s_{i+1}}] = s_{i+1} - s_i$. Only the diagonal terms of (2.31) therefore survive, by (2.32), so that

$$\mathbb{E}\Big[W(h^{1,m})\, W(h^{2,m})\Big] \;=\; \sum_{i=0}^{m-1} h^1_{s_i} \, h^2_{s_i} \, (s_{i+1} - s_i) \;=\; \int_0^\tau h^{1,m}_r \, h^{2,m}_r \, dr. \tag{2.33}$$

Step 3: passage to the limit. By Step 1, $(W(h^{1,m}), W(h^{2,m})) \to (W(h^1), W(h^2))$ in $L^2(\Omega)$, hence in distribution. Hence the characteristic function of the centered Gaussian vector $(W(h^{1,m}), W(h^{2,m}))$, determined entirely by its covariance matrix (2.33), converges pointwise to that of $(W(h^1), W(h^2))$, which is therefore itself a centered Gaussian vector. This yields part (i) of our claims. In addition, passing to the limit in (2.33), using $L^2(\Omega)$-convergence on the left and $h^{i,m} \to h^i$ in $L^2([0,\tau])$ on the right, gives (2.28). ■

We have now constructed the random variable $\int_s^t u_r\, dW_r$, for every fixed $s \le t$ in $[0,\tau]$, together with its basic algebraic and probabilistic properties (Proposition 2.1.17). Viewing $t \mapsto \int_0^t u_r\, dW_r$ as a stochastic process raises a further, genuinely delicate point: does it admit a continuous modification, as W itself does? We answer this question together with the martingale property of the integral below. Our proof relies on the following classical inequality. We state it without proof and refer to Karatzas–Shreve [6] or Revuz–Yor [7] for more details.

Proposition 2.1.20: Doob's L^2 maximal inequality

Let $N = \{N_t\,;\, t \in [0,\tau]\}$ be a continuous $\mathcal{F}_t$-martingale (Definition 1.5.6) with $N_t \in L^2(\Omega)$ for every $t \in [0,\tau]$. Then

$$\mathbb{E}\Big[\sup_{t\in[0,\tau]} N_t^2\Big] \le 4\,\mathbb{E}\big[N_\tau^2\big]. \tag{2.34}$$

We are now ready to state the announced martingale property. This is the last property of the Itô integral that we record before turning to Itô's formula itself.

Corollary 2.1.21

Let $u \in L^2_a([0,\tau])$ (Definition 2.1.8) and let $M = \{M_t\,;\, t \in [0,\tau]\}$ be given by $M_t := \int_0^t u_r\, dW_r$ (Notation 2.1.14).

(i) M is an $\mathcal{F}_t$-martingale (Definition 1.5.6).

(ii) M admits a continuous modification, which we denote again by M, in keeping with the convention of Remark 1.7.5.

Proof. Step 1: integrability and adaptedness. For every $t \in [0,\tau]$, the isometry (2.22) gives

$$\mathbb{E}\big[M_t^2\big] = \int_0^t \mathbb{E}\big[u_r^2\big]\, dr \le \|u\|_{L^2_a}^2 < \infty,$$

so $M_t \in L^2(\Omega) \subset L^1(\Omega)$. Adaptedness of M follows from Definition 2.1.13: M_t is an $L^2(\Omega)$-limit of the $\mathcal{F}_t$-measurable random variables $\mathcal{J}_{0,t}(u^n\, dW)$ (each a finite sum of terms $\xi_i\, \delta W_{t_i t_{i+1}}$ with $t_{i+1} \le t$, by Definition 2.1.3), hence itself $\mathcal{F}_t$-measurable.

Step 2: the martingale increment relation. Fix $0 \le s < t \le \tau$. By the additivity relation (2.20), with $r := t$ there, $\delta M_{st} = M_t - M_s = \int_s^t u_r\, dW_r$ almost surely. Property (iii)

of Proposition 2.1.17 then gives

$$\mathbb{E}\Big[\delta M_{st}\,\Big|\,\mathcal{F}_s\Big] \;=\; \mathbb{E}\Big[\int_s^t u_r\,dW_r\,\Big|\,\mathcal{F}_s\Big] \;=\; 0.$$

Together with Step 1, this verifies the three conditions of Definition 1.5.6, so M is an $\mathcal{F}_t$-martingale. This proves item (i).

Step 3: continuous approximations and Doob's inequality. Let $(u^n)_{n\ge 0}$ be a sequence in $\mathcal{E}([0,\tau])$ with $\|u-u^n\|_{L^2_a}\to 0$ (Proposition 2.1.10). Being convergent, $(u^n)_{n\ge 0}$ is a Cauchy sequence in $L^2_a([0,\tau])$, so, passing to a subsequence if needed, we may assume

$$\|u^{n+1}-u^n\|_{L^2_a} \;\le\; 2^{-n}, \qquad n\ge 0. \tag{2.35}$$

Set $M^n_t := \mathcal{J}_{0,t}(u^n\,dW)$, $t\in[0,\tau]$ (Definition 2.1.3). On each subinterval of the partition underlying u^n, formula (2.2) exhibits $t\mapsto M^n_t$ as affine in W_t, so M^n is continuous on $[0,\tau]$, being piecewise affine in the continuous process W. Now observe that $u^{n+1}-u^n$ is an element of $\mathcal{E}([0,\tau])\subset L^2_a([0,\tau])$. Moreover, by linearity (property (ii) of Proposition 2.1.5), we have

$$M^{n+1}_t - M^n_t = \mathcal{J}_{0,t}\Big((u^{n+1}-u^n)\,dW\Big)$$

almost surely. Thus Steps 1 and 2 above, applied with $u^{n+1}-u^n$ in place of u, show that $M^{n+1}-M^n$ is a continuous $\mathcal{F}_t$-martingale. Doob's inequality (2.34), applied to $N := M^{n+1}-M^n$, together with the isometry (2.4) and (2.35), gives

$$\mathbb{E}\Big[\sup_{t\in[0,\tau]}\Big(M^{n+1}_t-M^n_t\Big)^2\Big] \;\le\; 4\,\mathbb{E}\Big[\Big(M^{n+1}_\tau-M^n_\tau\Big)^2\Big] \;=\; 4\,\|u^{n+1}-u^n\|^2_{L^2_a} \;\le\; 4^{1-n}. \tag{2.36}$$

Step 4: almost sure uniform convergence. By Chebyshev's inequality and (2.36),

$$\mathbb{P}\Big(\sup_{t\in[0,\tau]}\big|M^{n+1}_t-M^n_t\big|>2^{-n/2}\Big) \;\le\; 2^n\,\mathbb{E}\Big[\sup_{t\in[0,\tau]}\Big(M^{n+1}_t-M^n_t\Big)^2\Big] \;\le\; 2^{2-n},$$

a summable sequence in n. By the (first) Borel–Cantelli lemma, almost surely there is $n_0(\omega)$ such that $\sup_{t\in[0,\tau]}|M^{n+1}_t(\omega)-M^n_t(\omega)|\le 2^{-n/2}$ for every $n\ge n_0(\omega)$. Since $\sum_{n\ge 0}2^{-n/2}<\infty$, on this almost-sure event $(M^n(\omega))_{n\ge 0}$ is a uniformly Cauchy sequence of continuous functions on $[0,\tau]$, hence converges uniformly to a continuous limit $\widetilde{M}(\omega) := \{\widetilde{M}_t(\omega)\,;\,t\in[0,\tau]\}$.

Step 5: identification of the limit. Fix $t\in[0,\tau]$. By Step 4, $M^n_t\to\widetilde{M}_t$ almost surely, hence in probability. On the other hand, owing to Definition 2.1.13, we also have $M^n_t\to M_t$ in $L^2(\Omega)$ (hence in probability as well). Since a sequence can converge in probability to at most one limit, we get $\widetilde{M}_t = M_t$ almost surely. As $t\in[0,\tau]$ was arbitrary, $\widetilde{M}$ is a continuous modification of M. This proves item (ii). ∎

With the integral in place, together with its algebraic, isometric, and martingale properties, we can turn to Itô's formula itself: the chain rule of stochastic calculus.

2.2 Itô's formula

Ordinary calculus rests on the chain rule: if f is $\mathcal{C}^1$ and g is differentiable, then $f(g(t))$ is differentiable with $\frac{d}{dt} f(g(t)) = f'(g(t))\, g'(t)$. Theorem 1.7.13 rules out this argument for $f(W_t)$: W has no derivative to plug into a chain rule. Yet $f(W_t)$ is, in general, still a well-behaved process, and Itô's formula identifies exactly what replaces the chain rule in this setting. The extra term that appears, absent from the classical formula, is a direct consequence of the finite quadratic variation of W (Theorem 1.7.11): informally, $(dW_t)^2$ behaves like dt, so a second-order Taylor expansion of f contributes a first-order term in dt that a first-order expansion would miss.

2.2.1 Statement of the formula

We work with functions of both time and space, since this is the form in which Itô's formula is used throughout the rest of the course. We first label the regularity that such a function must have.

Definition 2.2.1: The class $\mathcal{C}^{1,2}_{b,\tau}$

Let $\tau > 0$ and $d \geq 1$. A function $f : [0,\tau] \times \mathbb{R}^d \to \mathbb{R}$ belongs to $\mathcal{C}^{1,2}_{b,\tau}$ if:

(i) $t \mapsto f(t,x)$ is $\mathcal{C}^1([0,\tau])$ for every $x \in \mathbb{R}^d$.

(ii) $x \mapsto f(t,x)$ is $\mathcal{C}^2(\mathbb{R}^d)$ for every $t \in [0,\tau]$.

(iii) f, together with the derivatives $\partial_t f$, $\partial_{x_j} f$, and $\partial^2_{x_j x_k} f$ for $1 \leq j,k \leq d$, is bounded on $[0,\tau] \times \mathbb{R}^d$.

With this class in hand, we can state Itô's formula. We begin with the real-valued case, which contains the essential content of the result and is the one we prove in full in the next subsection.

Theorem 2.2.2: Itô's formula, one-dimensional case

Let $\tau > 0$, let W be a real-valued Wiener process, and let $f \in \mathcal{C}^{1,2}_{b,\tau}$ (Definition 2.2.1, $d = 1$). Then, for every $t \in [0,\tau]$, almost surely,

$$\begin{aligned} f(t,W_t) &= f(0,0) + \int_0^t \partial_r f(r,W_r)\, dr \\ &\quad + \int_0^t f'(r,W_r)\, dW_r + \frac{1}{2}\int_0^t f''(r,W_r)\, dr, \end{aligned} \tag{2.37}$$

where f' and f'' denote the first and second derivatives of f in its space variable.

Remark 2.2.3. *Every term in (2.37) is well defined: the two ordinary (Lebesgue) integrals make sense pathwise since $r \mapsto \partial_r f(r,W_r)$ and $r \mapsto f''(r,W_r)$ are almost surely continuous*

and bounded on $[0,t]$*, and the stochastic integral makes sense by Definition 2.1.13, since* $r \mapsto f'(r,W_r)$ *is continuous, bounded, hence progressively measurable and in* $L^2_a([0,\tau])$ *(Remark 2.1.7 and Definition 2.1.8).*

The multidimensional version of the formula, for a $\mathbb{R}^d$-valued Wiener process, is obtained by the same argument applied one coordinate of W at a time. We record its statement, and comment on its proof, at the end of the next subsection.

Theorem 2.2.4: Itô's formula

Let $\tau > 0$, let $W = (W^1,\dots,W^d)$ be a d-dimensional Wiener process (Definition 1.3.13), and let $f \in \mathcal{C}^{1,2}_{b,\tau}$ (Definition 2.2.1). Then, for every $t \in [0,\tau]$, almost surely,

$$\begin{aligned} f(t,W_t) &= f(0,0) + \int_0^t \partial_r f(r,W_r)\,dr + \sum_{j=1}^{d} \int_0^t \partial_{x_j} f(r,W_r)\,dW_r^j \\ &\quad + \frac{1}{2}\int_0^t \Delta f(r,W_r)\,dr, \end{aligned} \tag{2.38}$$

where $\Delta = \sum_{j=1}^d \partial^2_{x_j x_j}$ denotes the Laplace operator on $\mathbb{R}^d$.

Before turning to the proof, it is worth extracting the heuristic that makes (2.37) and (2.38) easy to remember, and that in fact drives the proof itself.

Remark 2.2.5 (Heuristic derivation)**.** *Consider the autonomous, real-valued case* $f = f(x)$*, so that* $\partial_r f \equiv 0$ *in (2.37). A formal Taylor expansion of* $f(W_t)$ *to second order gives*

$$\begin{aligned} df(W_t) &= f'(W_t)\,dW_t + \frac{1}{2}f''(W_t)\,(dW_t)^2 + \mathrm{o}\big((dW_t)^2\big) \\ &= f'(W_t)\,dW_t + \frac{1}{2}f''(W_t)\,dt + \mathrm{o}(dt), \end{aligned} \tag{2.39}$$

where the second line substitutes the heuristic identification $(dW_t)^2 \leftrightarrow dt$*, itself a restatement of the finite quadratic variation of* W *(Theorem 1.7.11) at the level of infinitesimals. Integrating (2.39) over* $[0,t]$ *recovers (2.37), in the autonomous case. The proof given below makes this heuristic rigorous, by controlling the error term* $\mathrm{o}(dt)$ *uniformly along a partition of* $[0,t]$*, in the limit of vanishing mesh.*

2.2.2 Proof of the one-dimensional formula

We prove Theorem 2.2.2 under two simplifying restrictions, which keep the argument in Remark 2.2.5 transparent while retaining its full content.

(i) We take $f = f(x)$ autonomous, that is $\partial_r f \equiv 0$ in (2.37). The time-dependent case then follows by combining the autonomous case, applied on each subinterval of a fine partition, with an ordinary (Lebesgue) first-order Taylor expansion of f in its time variable, whose contribution is exactly $\int_0^t \partial_r f(r, W_r)\, dr$.

(ii) We assume in addition that f''' exists and is bounded, beyond what Definition 2.2.1 requires. This extra derivative controls the modulus of continuity of f'' in Step 4 below. The general case $f \in \mathcal{C}^{1,2}_{b,\tau}$ follows by approximating f, uniformly on compacts together with its first two derivatives, by a sequence of functions with a bounded third derivative, and passing to the limit in the formula.

We omit the routine details of both reductions and refer to Karatzas–Shreve [6] for the fully general statement. Throughout the proof we fix $t \in [0, \tau]$ and $\gamma \in (0, 1/2)$, and write $c_\gamma := \|W\|_\gamma$ (Definition 1.7.1, on $[0, t]$), which is almost surely finite by Theorem 1.7.4.

Step 1: Taylor decomposition along a partition. Let $\pi = \{t_0, \dots, t_n\} \in \Pi_{0,t}$ (Definition 1.7.8). Taylor's theorem with Lagrange remainder, applied to f between W_{t_j} and $W_{t_{j+1}}$, gives, for every $0 \le j \le n-1$,

$$f(W_{t_{j+1}}) - f(W_{t_j}) \;=\; f'(W_{t_j})\, \delta W_{t_j t_{j+1}} \;+\; \frac{1}{2} f''(\xi_j) \left(\delta W_{t_j t_{j+1}}\right)^2, \tag{2.40}$$

for some (random) ξ_j lying between W_{t_j} and $W_{t_{j+1}}$. Summing (2.40) over $j = 0, \dots, n-1$, the left-hand side telescopes to $f(W_t) - f(W_0)$, and we obtain

$$f(W_t) - f(W_0) \;=\; J_t^{1,\pi} \;+\; \frac{1}{2} J_t^{2,\pi}, \tag{2.41}$$

where

$$\begin{aligned} J_t^{1,\pi} &:= \sum_{j=0}^{n-1} f'(W_{t_j})\, \delta W_{t_j t_{j+1}}, \\ J_t^{2,\pi} &:= \sum_{j=0}^{n-1} f''(\xi_j) \left(\delta W_{t_j t_{j+1}}\right)^2. \end{aligned} \tag{2.42}$$

Relation (2.41) holds for every partition $\pi \in \Pi_{0,t}$, with no limiting procedure yet involved: it is our starting point.

Step 2: strategy. Fix an arbitrary sequence of partitions $(\pi^m)_{m \ge 1}$, $\pi^m = \{t_0^m, \dots, t_{n_m}^m\} \in \Pi_{0,t}$, with $|\pi^m| \to 0$. We will produce a subsequence, still denoted $(\pi^m)_{m \ge 1}$ after passing to it, along which, almost surely,

$$\lim_{m \to \infty} J_t^{1,\pi^m} \;=\; \int_0^t f'(W_s)\, dW_s, \qquad \lim_{m \to \infty} J_t^{2,\pi^m} \;=\; \int_0^t f''(W_s)\, ds. \tag{2.43}$$

Passing to the limit in (2.41) along this subsequence then gives Theorem 2.2.2, in the autonomous case, at the fixed time t.

Step 3: the term J_t^{1,π^m}. The process $u := f'(W)$ is bounded (Definition 2.2.1) and has almost surely continuous paths, hence lies in $L_a^2([0,\tau])$ (Remark 2.1.7). With u^m denoting the piecewise-constant approximation of u along π^m, defined by (2.14), we have

$$J_t^{1,\pi^m} = \mathcal{J}_{0,t}(u^m\,dW)\,,$$

where we recall that $\mathcal{J}_{0,t}(u^m\,dW)$ is defined by (2.2). Proposition 2.1.11 gives $\|u - u^m\|_{L_a^2} \to 0$, so Definition 2.1.13 yields

$$L^2(\Omega)\text{-}\lim_{m\to\infty} J_t^{1,\pi^m} = \int_0^t f'(W_s)\,dW_s. \tag{2.44}$$

Passing to a subsequence, the limit (2.44) holds almost surely as well, which is the first relation in (2.43).

Step 4: the term J_t^{2,π^m}, strategy. Introduce the two auxiliary sums

$$J_t^{3,\pi} := \sum_{j=0}^{n-1} f''(W_{t_j})\left(\delta W_{t_j t_{j+1}}\right)^2, \qquad J_t^{4,\pi} := \sum_{j=0}^{n-1} f''(W_{t_j})\,(t_{j+1}-t_j), \tag{2.45}$$

obtained from $J_t^{2,\pi}$ by replacing ξ_j with the left endpoint W_{t_j}, respectively by further replacing $(\delta W_{t_j t_{j+1}})^2$ with its mean $(t_{j+1}-t_j)$. We show, along a further subsequence of $(\pi^m)_{m\geq 1}$, that almost surely

$$\lim_{m\to\infty}\left|J_t^{2,\pi^m} - J_t^{3,\pi^m}\right| = 0, \tag{2.46}$$

$$\lim_{m\to\infty}\left|J_t^{3,\pi^m} - J_t^{4,\pi^m}\right| = 0, \tag{2.47}$$

$$\lim_{m\to\infty} J_t^{4,\pi^m} = \int_0^t f''(W_s)\,ds. \tag{2.48}$$

Together, (2.46)–(2.48) give the second relation in (2.43), and complete Step 2.

Step 5: proof of (2.48). Fix ω in the almost sure event on which $s \mapsto W_s(\omega)$ is continuous (Theorem 1.7.4). Then $s \mapsto f''(W_s(\omega))$ is continuous on $[0,t]$, since f'' is continuous. By construction, $J_t^{4,\pi^m}(\omega)$ is exactly the Riemann sum of $s \mapsto f''(W_s(\omega))$ along π^m, evaluated at left endpoints. Since $|\pi^m| \to 0$, this Riemann sum converges to the corresponding Riemann integral, that is

$$\lim_{m\to\infty} J_t^{4,\pi^m}(\omega) = \int_0^t f''(W_s(\omega))\,ds,$$

which is (2.48). No further subsequence extraction is needed for this step: the convergence already holds along the full sequence $(\pi^m)_{m\geq 1}$.

Step 6: proof of (2.46). Set $S_{\pi^m} := \sum_{j=0}^{n_m-1}(\delta W_{t_j^m t_{j+1}^m})^2$. Equation (1.120) (with $a=0, b=t$) shows that $S_\pi \to t$ in $L^2(\Omega)$ along any sequence of partitions of $[0,t]$ with mesh going to 0, hence in particular along $(\pi^m)_{m\geq 1}$. Passing to a subsequence of $(\pi^m)_{m\geq 1}$, still denoted $(\pi^m)_{m\geq 1}$, we may therefore assume $S_{\pi^m} \to t$ almost surely. Recall that $c_\gamma = \|W\|_\gamma$, fixed

at the start of the proof, is almost surely finite by Theorem 1.7.4. Fix ω in the almost sure event

$$\{c_\gamma < \infty\} \cap \{S_{\pi^m} \to t\}.$$

Since f''' is bounded, f'' is Lipschitz with constant $\|f'''\|_\infty$. Moreover, ξ_j lies between W_{t_j} and $W_{t_{j+1}}$. Hence we get

$$\left|f''(W_{t_j}) - f''(\xi_j)\right| \le \|f'''\|_\infty \left|\delta W_{t_j t_{j+1}}\right| \le \|f'''\|_\infty c_\gamma \left|t_{j+1} - t_j\right|^\gamma \le \|f'''\|_\infty c_\gamma |\pi^m|^\gamma, \quad (2.49)$$

Combining (2.49) with the definitions of $J_t^{2,\pi}$ and $J_t^{3,\pi}$ in (2.42) and (2.45) thus gives

$$\left|J_t^{3,\pi^m}(\omega) - J_t^{2,\pi^m}(\omega)\right| \le \|f'''\|_\infty c_\gamma |\pi^m|^\gamma S_{\pi^m}(\omega) \xrightarrow[m\to\infty]{} \|f'''\|_\infty c_\gamma \cdot 0 \cdot t = 0, \quad (2.50)$$

since $|\pi^m| \to 0$ and $S_{\pi^m}(\omega) \to t$. This is (2.46).

Step 7: proof of (2.47). Set $Z_j := (\delta W_{t_j t_{j+1}})^2 - (t_{j+1} - t_j)$, so that, by (2.45),

$$J_t^{4,\pi} - J_t^{3,\pi} = -\sum_{j=0}^{n-1} f''(W_{t_j})\, Z_j. \quad (2.51)$$

We show $\lim_{m\to\infty} \mathbb{E}\Big[(J_t^{4,\pi^m} - J_t^{3,\pi^m})^2\Big] = 0$, which yields (2.47) almost surely along a further subsequence. To this aim, squaring (2.51) we obtain

$$\mathbb{E}\Big[(J_t^{4,\pi} - J_t^{3,\pi})^2\Big] = \sum_{j=0}^{n-1} \mathbb{E}\Big[f''(W_{t_j})^2 Z_j^2\Big] + 2 \sum_{0\le j<k\le n-1} \mathbb{E}\Big[f''(W_{t_j})\, Z_j\, f''(W_{t_k})\, Z_k\Big]. \quad (2.52)$$

For the cross terms, $j < k$, the random variable $f''(W_{t_j})\, Z_j\, f''(W_{t_k})$ is $\mathcal{F}_{t_k}$-measurable, and $\delta W_{t_k t_{k+1}} \perp\!\!\!\perp \mathcal{F}_{t_k}$. In addition, we have

$$\mathbb{E}[Z_k] = \mathbb{E}[(\delta W_{t_k t_{k+1}})^2] - (t_{k+1} - t_k) = 0.$$

Hence by the same conditioning argument as in (2.7)–(2.9), we end up with

$$\mathbb{E}\Big[f''(W_{t_j})\, Z_j\, f''(W_{t_k})\, Z_k\Big] = \mathbb{E}\Big[f''(W_{t_j})\, Z_j\, f''(W_{t_k})\, \mathbb{E}[Z_k \,|\, \mathcal{F}_{t_k}]\Big] = 0, \quad (2.53)$$

so the cross terms in (2.52) vanish identically. For the diagonal terms, observe that $f''(W_{t_j})^2$ is $\mathcal{F}_{t_j}$-measurable and $\delta W_{t_j t_{j+1}} \perp\!\!\!\perp \mathcal{F}_{t_j}$. Thereofre the same freezing argument as in (2.53) gives

$$\mathbb{E}[f''(W_{t_j})^2 Z_j^2] = \mathbb{E}[f''(W_{t_j})^2]\, \mathbb{E}[Z_j^2].$$

Moreover $\mathbb{E}[Z_j^2] = 2(t_{j+1} - t_j)^2$, by the same computation as for (1.118). Hence we have

$$\begin{aligned} \sum_{j=0}^{n-1} \mathbb{E}\Big[f''(W_{t_j})^2 Z_j^2\Big] &= 2 \sum_{j=0}^{n-1} \mathbb{E}\Big[f''(W_{t_j})^2\Big] (t_{j+1} - t_j)^2 \\ &\le 2 \|f''\|_\infty^2 |\pi| \sum_{j=0}^{n-1} (t_{j+1} - t_j) = 2 \|f''\|_\infty^2\, t\, |\pi|. \end{aligned} \quad (2.54)$$

Combining (2.52), (2.53), and (2.54), applied to $\pi = \pi^m$, gives

$$\mathbb{E}\Big[(J_t^{4,\pi^m} - J_t^{3,\pi^m})^2\Big] \le 2\,\|f''\|_\infty^2\, t\,|\pi^m| \xrightarrow[m\to\infty]{} 0,$$

which is the claimed $L^2(\Omega)$-convergence.

Step 8: conclusion. Steps 5–7 established (2.46)–(2.48) along a common subsequence of $(\pi^m)_{m\ge1}$. More specifically, Step 6 required only passing to the subsequence already extracted for $S_{\pi^m} \to t$, Step 7 required a further subsequence for the almost sure convergence $J_t^{4,\pi^m} - J_t^{3,\pi^m} \to 0$, and Step 5 held along every sequence, so it survives both extractions. Along this final subsequence, (2.43) holds almost surely. This proves Theorem 2.2.2, at the fixed time $t \in [0,\tau]$, in the autonomous, $\mathcal{C}_b^3$ case described at the start of the proof. ■

Remark 2.2.6 (The multidimensional case)**.** *The proof of Theorem 2.2.4, in the d-dimensional case, follows the same six steps, with two changes: the Taylor expansion (2.40) is taken to second order in $\mathbb{R}^d$, producing cross terms $\partial^2_{x_{j_1}x_{j_2}} f(W_{t_k})\,\delta W^{j_1}_{t_kt_{k+1}}\,\delta W^{j_2}_{t_kt_{k+1}}$ for $j_1 \ne j_2$, and the analogue of the quadratic variation computation of Step 7 shows that these cross terms vanish in the limit, rather than surviving as in (2.38): this uses the independence of the coordinates $W^1,\dots,W^d$ (Definition 1.3.13), in place of the identity $\mathbb{E}[Z_j^2] = 2(t_{j+1}-t_j)^2$ used above. We omit the bookkeeping and refer to Karatzas–Shreve [6].*

We close this section with a piece of history behind Itô's formula, no less remarkable than the mathematics itself.

Remark 2.2.7 (Wolfgang Doeblin, 1915–1940)**.** *Wolfgang Doeblin, son of the novelist Alfred Döblin, fled Nazi Germany for France in 1933 and later obtained French citizenship. He was already a leading probabilist by the outbreak of the Second World War, when he was mobilized into the French army as a telephone operator. In 1940, isolated and anticipating capture, he sent the French Academy of Sciences a sealed envelope, a* pli cacheté*, containing an outline of results that anticipated much of the theory built in this chapter, including a version of Itô's formula, several years before Itô's own work (Remark 2.1.16). Doeblin died soon after, at 25 years of age, on the collapsing French front. His envelope, unclaimed, was not opened until the year 2000, at the request of his family.*

2.3 Extensions and consequences of Itô's formula

Theorem 2.2.4 applies Itô's formula to functions of the Wiener process W itself. Many applications, starting with the stochastic differential equations of Chapter 3, instead require functions of a process X that is only *driven* by W, through both a drift term and a stochastic integral against W. This section extends Itô's formula to this broader class of processes, and draws two consequences: an infinitesimal, conditional-expectation form of the formula, which recovers and refines the generator computation of Remark 1.6.13, and its extension to the same broader class.

2.3.1 Itô processes

Definition 2.3.1: Itô process

Let $\tau > 0$, $n, d \geq 1$, and let W be a d-dimensional Wiener process. An $\mathbb{R}^n$-valued process $X = \{X_t\,;\, t \in [0,\tau]\}$ is an *Itô process*, driven by W, if there exist $a \in \mathbb{R}^n$, a real bounded, adapted process $b = \{b^j\,;\, 1 \leq j \leq n\}$, and a process $\sigma = \{\sigma^{jk}\,;\, 1 \leq j \leq n,\ 1 \leq k \leq d\}$ with $\sigma^{jk} \in L^2_a([0,\tau])$ (Definition 2.1.8) for every j, k, such that

$$X^j_t \;=\; a^j \;+\; \int_0^t b^j_s\, ds \;+\; \sum_{k=1}^{d} \int_0^t \sigma^{jk}_s\, dW^k_s, \qquad t \in [0,\tau],\ 1 \leq j \leq n. \tag{2.55}$$

Remark 2.3.2. *The two families of terms in (2.55) have different natures. Namely the process $t \mapsto a^j + \int_0^t b^j_s\, ds$ has finite variation, being an ordinary (Lebesgue) integral, while $t \mapsto \sum_k \int_0^t \sigma^{jk}_s\, dW^k_s$ is, by Corollary 2.1.21 applied coordinatewise, a martingale. A process that decomposes as the sum of a finite-variation part and a martingale part, as in (2.55), is called a* semimartingale*: Itô processes are the semimartingales that appear throughout this course.*

We can now state the extension of Itô's formula to Itô processes. Its proof follows the same pattern as that of Theorem 2.2.2, with the finite-variation part b contributing an ordinary chain-rule term and the coordinates of σ recombining, through the same quadratic-variation computation as in Step 7 above, into the matrix $\sigma\sigma^T$. We state it without proof and refer to Karatzas–Shreve [6].

Theorem 2.3.3: Itô's formula for Itô processes

Let X be an Itô process driven by a d-dimensional Wiener process W, with data a, b, σ as in Definition 2.3.1, and let $f \in \mathcal{C}^{1,2}_{b,\tau}$ (Definition 2.2.1). Then, for every $t \in [0,\tau]$, almost surely,

$$\begin{aligned} f(t, X_t) \;=\; & f(0,a) \;+\; \int_0^t \partial_r f(r, X_r)\, dr \;+\; \sum_{j=1}^{n} \int_0^t \partial_{x_j} f(r, X_r)\, b^j_r\, dr \\ & + \sum_{j=1}^{n} \sum_{k=1}^{d} \int_0^t \partial_{x_j} f(r, X_r)\, \sigma^{jk}_r\, dW^k_r \\ & + \frac{1}{2} \sum_{j_1, j_2 = 1}^{n} \sum_{k=1}^{d} \int_0^t \partial^2_{x_{j_1} x_{j_2}} f(r, X_r)\, \sigma^{j_1 k}_r\, \sigma^{j_2 k}_r\, dr. \end{aligned} \tag{2.56}$$

Remark 2.3.4. *Theorem 2.2.4 is the special case $n = d$, $a = 0$, $b \equiv 0$, $\sigma \equiv \mathrm{Id}_d$ of Theorem 2.3.3. Indeed, in this case we have $X = W$ and $\sum_k \sigma^{j_1 k}\sigma^{j_2 k} = \delta_{j_1 j_2}$, so that the last line of (2.56) reduces to $\frac{1}{2}\int_0^t \Delta f(r, X_r)\, dr$.*

The boundedness of f and its derivatives, required in Definition 2.2.1, was used above only to control error terms uniformly along the underlying partition. It plays no role in

the identity itself, and both Itô's formula and the generator identity of the next subsection extend to functions that fail to be globally bounded, by a routine stopping-time argument.

Proposition 2.3.5: Localization

Theorems 2.2.2, 2.2.4, and 2.3.3 continue to hold for f merely of class $\mathcal{C}^{1,2}$ jointly in its time and space variables, without the global boundedness required by Definition 2.2.1. The same holds for Theorem 2.3.8 below, with $f \in \mathcal{C}^2(\mathbb{R}^n)$ in place of $f \in \mathcal{C}_b^2(\mathbb{R}^n)$.

Proof. Let X denote the underlying Itô process (or $X = W$, for Theorems 2.2.2 and 2.2.4), let $\chi_n : \mathbb{R}^d \to [0,1]$ be smooth, equal to 1 on the ball $B(0,n)$ and to 0 outside $B(0,n+1)$, with χ_n and its derivatives up to order two bounded uniformly in n, and set $f_n := \chi_n f$, an element of $\mathcal{C}_{b,\tau}^{1,2}$ (respectively $\mathcal{C}_b^2(\mathbb{R}^n)$ for Theorem 2.3.8). Let

$$T_n := \inf\{t \in [0,\tau]\,;\, |X_t| \geq n\} \wedge \tau,$$

which is a stopping time by Proposition 1.5.10. Since $f_n = f$ on $[0,\tau] \times B(0,n)$, the formula for f_n, evaluated up to time $t \wedge T_n$, coincides with the formula for f on $[0, T_n]$. Since X has continuous paths on the compact interval $[0,\tau]$, it is almost surely bounded there, so $T_n = \tau$ for n large enough, almost surely. Letting $n \to \infty$ extends the formula to all of $[0,\tau]$. We refer to Karatzas–Shreve [6] for the complete argument, and use this extension freely from here on, whenever f fails to be globally bounded. ∎

2.3.2 The infinitesimal generator of Brownian motion

Remark 1.6.13 identified the generator of the heat semigroup $\{p_t\}$ as $\mathcal{L} = \frac{1}{2}\Delta$, through an explicit differentiation of the Gaussian density defining $p_t f$. We now recover, and refine, the same identity directly from Itô's formula, as a statement about conditional expectations given $\mathcal{F}_s$ rather than about the unconditional semigroup. This form of the generator is the one that generalizes most readily to Itô processes, at the end of this section. It also reappears when we study stochastic differential equations in Chapter 3.

Proposition 2.3.6: Generator of Brownian motion

Let W be a d-dimensional Wiener process with respect to $(\mathcal{F}_t)_{t\geq 0}$, let $s \geq 0$, and let $f \in \mathcal{C}_b^2(\mathbb{R}^d)$. Then, almost surely,

$$\lim_{h\to 0^+} \frac{\mathbb{E}\big[\delta f(W)_{s,s+h} \,\big|\, \mathcal{F}_s\big]}{h} = \frac{1}{2}\Delta f(W_s). \tag{2.57}$$

Proof. Step 1: recasting Itô's formula. Apply Theorem 2.2.4 to the (autonomous) function f, and set

$$M_t := \sum_{j=1}^{d} \int_0^t \partial_{x_j} f(W_r)\, dW_r^j.$$

Since (2.38) reads $f(W_t) = f(W_0) + M_t + \frac{1}{2}\int_0^t \Delta f(W_r)\,dr$, taking increments between s and $s+h$ gives

$$\delta f(W)_{s,s+h} \;=\; \delta M_{s,s+h} \;+\; \frac{1}{2}\int_s^{s+h} \Delta f(W_r)\,dr, \tag{2.58}$$

the additivity of the stochastic integral M over $[0,s]\cup[s,s+h]$ following from Remark 2.1.15.

Step 2: conditional expectation. Each coordinate of M is a martingale, by Corollary 2.1.21 applied to $u := \partial_{x_j} f(W)$, which is bounded and continuous (Remark 2.2.3). By Definition 1.5.6, this gives

$$\mathbb{E}\Big[\delta M_{s,s+h}\,\Big|\,\mathcal{F}_s\Big] \;=\; 0.$$

Taking $\mathbb{E}[\,\cdot\,|\,\mathcal{F}_s]$ in (2.58) therefore yields

$$\mathbb{E}\Big[\delta f(W)_{s,s+h}\,\Big|\,\mathcal{F}_s\Big] \;=\; \frac{1}{2}\int_s^{s+h} \mathbb{E}\Big[\Delta f(W_r)\,\Big|\,\mathcal{F}_s\Big]\,dr. \tag{2.59}$$

Step 3: Limit procedure. The map $r \mapsto \Delta f(W_r)$ is bounded. We first show that its conditional expectation given $\mathcal{F}_s$ is continuous in $r \geq s$. Fix $r \geq s$ and set $t := r - s$. Since $\delta W_{s,s+t} \perp\!\!\!\perp \mathcal{F}_s$ with law $\mathcal{N}(0, t\,\mathrm{Id}_d)$ similarly to (1.84) we obtain

$$\mathbb{E}\Big[\Delta f(W_r)\,\Big|\,\mathcal{F}_s\Big] \;=\; \mathbb{E}\Big[\Delta f(W_s + \delta W_{s,s+t})\,\Big|\,\mathcal{F}_s\Big] \;=\; p_t(\Delta f)(W_s), \tag{2.60}$$

In addition, since Δf is bounded and continuous, $u \mapsto p_u(\Delta f)(x)$ is continuous at $u = 0$ for every x, so $r \mapsto p_{r-s}(\Delta f)(W_s) = \mathbb{E}[\Delta f(W_r)\,|\,\mathcal{F}_s]$ is continuous at $r = s$. Dividing (2.59) by h and letting $h \to 0^+$, this continuity gives

$$\lim_{h\to 0^+} \frac{1}{h}\int_s^{s+h} \mathbb{E}\Big[\Delta f(W_r)\,\Big|\,\mathcal{F}_s\Big]\,dr \;=\; \mathbb{E}\Big[\Delta f(W_s)\,\Big|\,\mathcal{F}_s\Big] \;=\; \Delta f(W_s),$$

the last equality since $\Delta f(W_s)$ is $\mathcal{F}_s$-measurable. Combined with (2.59), this proves (2.57). ∎

Remark 2.3.7. *Taking $\mathbb{E}[\,\cdot\,]$ (unconditionally) in (2.57) and interchanging the limit with the (unconditional) expectation, justified as in Step 3 above, recovers the relation*

$$\lim_{h\to 0^+} \frac{1}{h}\big(p_h f(x) - f(x)\big) = \frac{1}{2}\Delta f(x),$$

at $s = 0$ and $W_0 = x$. This is exactly the generator identity of Remark 1.6.13, now derived from Itô's formula rather than from the Gaussian density of p_t.

The same computation, with the drift and diffusion terms of an Itô process contributing their own pieces, extends Proposition 2.3.6 to Itô processes. We record the result and refer to Karatzas–Shreve [6] for the (routine) proof, identical to that of Proposition 2.3.6 once Theorem 2.3.3 is used in place of Theorem 2.2.4 in Step 1.

Theorem 2.3.8: Generator of an Itô process

Let X be an Itô process driven by a d-dimensional Wiener process W, with data a, b, σ as in Definition 2.3.1, and let $f \in \mathcal{C}_b^2(\mathbb{R}^n)$. Then, for every $s \in [0, \tau)$, almost surely,

$$\lim_{h \to 0^+} \frac{\mathbb{E}\left[\delta f(X)_{s,s+h} \,\middle|\, \mathcal{F}_s\right]}{h} = \sum_{j=1}^{n} \partial_{x_j} f(X_s)\, b_s^j + \frac{1}{2} \sum_{j_1, j_2 = 1}^{n} \sum_{k=1}^{d} \partial^2_{x_{j_1} x_{j_2}} f(X_s)\, \sigma_s^{j_1 k}\, \sigma_s^{j_2 k}. \tag{2.61}$$

Remark 2.3.9. *The right-hand side of* (2.61) *defines the* infinitesimal generator *of the Itô process X, applied to f and evaluated at X_s. It depends on X only through its drift b and diffusion matrix $\sigma\sigma^T$, and is a second-order differential operator in x, with coefficients that may themselves depend on s and ω through b_s and σ_s. When b and σ are deterministic functions of (s, X_s) alone, as will be the case for the stochastic differential equations studied in Chapter 3, we get a genuine generator of a Markov process (in the sense of Theorem 1.6.10). This generalizes Proposition 1.6.12 beyond Brownian motion.*

CHAPTER 3

Stochastic differential equations

Chapter 2 introduced Itô processes (Definition 2.3.1): processes that decompose as the sum of a Lebesgue integral and a stochastic integral against a Wiener process W, with drift and diffusion data that may depend on ω and s in a completely arbitrary, adapted way. This chapter specializes that picture in a single respect. From now on, the drift and diffusion at time s depend on ω only through the current value X_s of the process itself, via two fixed functions on $\mathbb{R}^n$. The resulting object is a *stochastic differential equation* (SDE): the stochastic analogue of an ordinary differential equation $\dot{x} = b(x)$, driven by an additional source of Gaussian noise.

This first section states the equation, its most immediate consequence (the infinitesimal drift and covariance identity of Proposition 3.1.4), and three examples to which we return throughout the chapter: geometric Brownian motion, the Ornstein–Uhlenbeck process, and a diffusion approximation of a branching population model. In each example a solution is exhibited explicitly, so existence is not an issue there. Section 3.2 isolates a further case that already goes beyond direct substitution: equations whose diffusion coefficient depends on time alone, not on the state, for which a simple change of variables reduces existence and uniqueness to classical, deterministic ordinary differential equation theory. The general question of existence and uniqueness, with a diffusion coefficient that genuinely depends on the state, is taken up in the section after that.

3.1 Introduction and examples

3.1.1 Stochastic differential equations

Throughout this section we fix $\tau > 0$, $n, d \geq 1$, and a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\tau]}, \mathbb{P})$ carrying a d-dimensional Wiener process W with respect to $(\mathcal{F}_t)_{t\in[0,\tau]}$ (Definition 1.3.13).

Definition 3.1.1: Solution of a stochastic differential equation

Let $\alpha \in \mathbb{R}^n$, and let $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma = (\sigma^1, \dots, \sigma^d) : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ be continuous. A continuous, $\mathcal{F}_t$-adapted process $X = \{X_t\,;\, t \in [0,\tau]\}$ with values in $\mathbb{R}^n$ is a *solution* of the stochastic differential equation with coefficients (α, b, σ), driven by W, if $\sigma^{jk}(X_\cdot) \in L^2_a([0,\tau])$ (Definition 2.1.8) for every $1 \leq j \leq n$, $1 \leq k \leq d$, and, almost surely, for every $t \in [0,\tau]$ and $1 \leq j \leq n$, the following relation holds true:

$$X^j_t \;=\; \alpha^j \;+\; \int_0^t b^j(X_s)\,ds \;+\; \sum_{k=1}^{d} \int_0^t \sigma^{jk}(X_s)\,dW^k_s. \tag{3.1}$$

We also write this in differential form as

$$dX_t \;=\; b(X_t)\,dt \;+\; \sigma(X_t)\,dW_t, \qquad X_0 = \alpha, \tag{3.2}$$

a shorthand for (3.1) with no independent meaning.

Remark 3.1.2 (Well-posedness)**.** *Every term in (3.1) is well defined. Since X and b are continuous, $s \mapsto b^j(X_s)$ is almost surely continuous, hence bounded, on the compact interval $[0,\tau]$, so the ordinary (Lebesgue) integral $\int_0^t b^j(X_s)\,ds$ makes sense path by path. The stochastic integral $\int_0^t \sigma^{jk}(X_s)\,dW^k_s$ is well defined by Definition 2.1.13, precisely because of the integrability condition $\sigma^{jk}(X_\cdot) \in L^2_a([0,\tau])$ imposed above.*

Remark 3.1.3 (Relation to Itô processes)**.** *When b and σ are, in addition, bounded, as in Proposition 3.1.4 below, $b(X_\cdot)$ is automatically a bounded adapted process, and $\sigma^{jk}(X_\cdot)$ automatically belongs to $L^2_a([0,\tau])$, being bounded on the finite interval $[0,\tau]$. In this case a solution X of (3.1) is an Itô process (Definition 2.3.1), with $b_s = b(X_s)$ and $\sigma_s = \sigma(X_s)$. In particular, by Remark 2.3.2, X decomposes as the sum of a finite-variation part and a martingale part, the latter a genuine martingale by Corollary 2.1.21. When b or σ fail to be bounded, as in the geometric Brownian motion and Ornstein–Uhlenbeck examples below, we verify these properties directly for each solution instead.*

We can now state, and prove, the basic link between the coefficients (b, σ) of a stochastic differential equation and the short-time behavior of its solution: the coefficients determine the conditional mean and covariance of X, to first order in time.

Proposition 3.1.4: Infinitesimal drift and covariance

Let b, σ be bounded and continuous, and let X be a solution of (3.1) (Definition 3.1.1). Set $a := \sigma\sigma^T : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and, for $s \in [0,\tau)$, $h \in (0, \tau - s]$, and $1 \le l, m \le n$,

$$\mathrm{Cov}_s\Big(X^l_{s+h}, X^m_{s+h}\Big) := \mathbb{E}\Big[X^l_{s+h}X^m_{s+h}\,\Big|\,\mathcal{F}_s\Big] - \mathbb{E}\Big[X^l_{s+h}\,\Big|\,\mathcal{F}_s\Big]\,\mathbb{E}\Big[X^m_{s+h}\,\Big|\,\mathcal{F}_s\Big]. \tag{3.3}$$

Then, for every $s \in [0,\tau)$ and $1 \le l, m \le n$, almost surely, we have

$$\begin{aligned} \lim_{h\to 0^+} \frac{\mathbb{E}[X^l_{s+h}\,|\,\mathcal{F}_s] - X^l_s}{h} &= b^l(X_s), \\ \lim_{h\to 0^+} \frac{\mathrm{Cov}_s(X^l_{s+h}, X^m_{s+h})}{h} &= a^{lm}(X_s). \end{aligned} \tag{3.4}$$

Proof. By Remark 3.1.3, X is an Itô process with $b_s = b(X_s)$, $\sigma_s = \sigma(X_s)$, so Theorem 2.3.8 applies to it.

Step 1: the drift identity. Fix $1 \le j \le n$ and apply Theorem 2.3.8, extended to unbounded f by Proposition 2.3.5, to the linear function $f(x) := x^j$. Having the relations $\partial_{x_k} f = \delta_{jk}$ and $\partial^2_{x_{k_1}x_{k_2}} f = 0$ in mind, and since $b_s = b(X_s)$, this gives (almost surely):

$$\lim_{h\to 0^+} \frac{\mathbb{E}[X^j_{s+h} - X^j_s\,|\,\mathcal{F}_s]}{h} = b^j(X_s). \tag{3.5}$$

Moreover X is adapted, X^j_s is $\mathcal{F}_s$-measurable, so $\mathbb{E}[X^j_s\,|\,\mathcal{F}_s] = X^j_s$ almost surely, and (3.5) is exactly the first line of (3.4).

Step 2: the quadratic identity. Fix $1 \le l, m \le n$ and apply Theorem 2.3.8, again extended by Proposition 2.3.5, to the quadratic function $f(x) := x^l x^m$. We compute $\partial_{x_k} f(x) = x^m\delta_{kl} + x^l\delta_{km}$ and $\partial^2_{x_{k_1}x_{k_2}} f = \delta_{k_1 l}\delta_{k_2 m} + \delta_{k_1 m}\delta_{k_2 l}$, so that

$$\begin{aligned} \sum_{k=1}^n \partial_{x_k} f(X_s)\, b^k_s &= X^m_s\, b^l(X_s) + X^l_s\, b^m(X_s), \\ \frac{1}{2}\sum_{k_1,k_2=1}^n \sum_{k=1}^d \partial^2_{x_{k_1}x_{k_2}} f(X_s)\, \sigma^{k_1 k}_s\, \sigma^{k_2 k}_s &= \sum_{k=1}^d \sigma^{lk}(X_s)\,\sigma^{mk}(X_s) = a^{lm}(X_s). \end{aligned}$$

Theorem 2.3.8 therefore gives, almost surely,

$$\lim_{h\to 0^+} \frac{\mathbb{E}[X^l_{s+h}X^m_{s+h} - X^l_s X^m_s\,|\,\mathcal{F}_s]}{h} = X^m_s\, b^l(X_s) + X^l_s\, b^m(X_s) + a^{lm}(X_s). \tag{3.6}$$

Step 3: the covariance identity. By Step 1, applied to both l and m, the two maps $h \mapsto \mathbb{E}[X^l_{s+h}\,|\,\mathcal{F}_s]$ and $h \mapsto \mathbb{E}[X^m_{s+h}\,|\,\mathcal{F}_s]$ have a right-derivative at $h = 0$, equal to $b^l(X_s)$

and $b^m(X_s)$ respectively, and both equal X_s^l, respectively X_s^m, at $h = 0$. The one-sided product rule therefore gives

$$\lim_{h\to 0^+} \frac{\mathbb{E}[X_{s+h}^l \,|\, \mathcal{F}_s]\,\mathbb{E}[X_{s+h}^m \,|\, \mathcal{F}_s] \;-\; X_s^l X_s^m}{h} \;=\; b^l(X_s)\,X_s^m \;+\; X_s^l\,b^m(X_s). \tag{3.7}$$

Subtracting (3.7) from (3.6) and recalling Definition (3.3) of Cov_s, we end up with

$$\lim_{h\to 0^+} \frac{\mathrm{Cov}_s(X_{s+h}^l, X_{s+h}^m)}{h} \;=\; a^{lm}(X_s),$$

the second line of (3.4). This finishes the proof. ∎

Remark 3.1.5 (Interpretation)**.** *Proposition 3.1.4 justifies the terminology attached to b and σ in Definition 3.1.1. Namely $b(X_s)$ is the* infinitesimal drift *of X at X_s: the first-order rate of change of its conditional mean. And $a(X_s) = \sigma\sigma^T(X_s)$ is the* infinitesimal covariance *of X at X_s: the first-order rate of change of its conditional covariance matrix. Equation (3.2) should be read through this lens. Namely b prescribes an average, ordinary differential drift, while σ prescribes the intensity of the Gaussian noise superimposed on it, via the driving Wiener process W.*

3.1.2 Examples

We now work through three examples of stochastic differential equations, in each of which $n = d = 1$ and a solution can be exhibited explicitly.

Proposition 3.1.6: Geometric Brownian motion

Let $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha \in \mathbb{R}$, and let W be a real-valued Wiener process. Set

$$X_t \;:=\; \alpha \exp\!\Big(\mu t + \sigma W_t\Big), \qquad t \in [0,\tau]. \tag{3.8}$$

Then X is a solution (Definition 3.1.1, $n = d = 1$) of (3.2) with coefficients

$$b(x) \;=\; \Big(\mu + \frac{\sigma^2}{2}\Big)x, \qquad \sigma(x) \;=\; \sigma x. \tag{3.9}$$

Proof. Step 1: continuity and adaptedness. The map $(t,w) \mapsto \alpha\exp(\mu t + \sigma w)$ is continuous on $[0,\tau]\times\mathbb{R}$, and $t \mapsto (t, W_t)$ is continuous and $\mathcal{F}_t$-adapted, so X, as their composition, is itself continuous and $\mathcal{F}_t$-adapted.

Step 2: the integrability condition. Since $W_s \sim \mathcal{N}(0,s)$ (Definition 1.3.2), its moment generating function gives $\mathbb{E}[e^{2\sigma W_s}] = e^{2\sigma^2 s}$, so

$$\mathbb{E}\Big[X_s^2\Big] \;=\; \alpha^2 e^{2\mu s}\,\mathbb{E}\Big[e^{2\sigma W_s}\Big] \;=\; \alpha^2 e^{(2\mu+2\sigma^2)s}, \tag{3.10}$$

a continuous, hence bounded, function of $s \in [0,\tau]$. Consequently $\int_0^\tau \mathbb{E}[\sigma(X_s)^2]\,ds = \sigma^2 \int_0^\tau \mathbb{E}[X_s^2]\,ds < \infty$, so $\sigma(X_\cdot) = \sigma X_\cdot \in L_a^2([0,\tau])$ (Definition 2.1.8).

Step 3: Itô's formula. Set $f(t,w) := \alpha \exp(\mu t + \sigma w)$, and apply Theorem 2.2.2 to f along W, extended to this unbounded f by Proposition 2.3.5. We compute $\partial_t f = \mu f$, $f' = \sigma f$, and $f'' = \sigma^2 f$, and $f(0,0) = \alpha$, $f(t,W_t) = X_t$, so (2.37) gives

$$\begin{aligned} X_t &= \alpha + \int_0^t \mu X_r\,dr + \int_0^t \sigma X_r\,dW_r + \frac{1}{2}\int_0^t \sigma^2 X_r\,dr \\ &= \alpha + \int_0^t \Big(\mu + \frac{\sigma^2}{2}\Big) X_r\,dr + \int_0^t \sigma X_r\,dW_r, \end{aligned}$$

which is exactly (3.1) with the coefficients (3.9). ■

Remark 3.1.7 (Interpretation)**.** *In the differential form (3.2), the coefficients (3.9) identify $b(X_s) = (\mu + \sigma^2/2)X_s$ and $\sigma(X_s)^2 = \sigma^2 X_s^2$ as the infinitesimal drift and variance of X at X_s. Note that Proposition 3.1.4 does not apply directly here, since b and σ in (3.9) are unbounded. However, the proof of Proposition 3.1.6 obtains the same two identities directly, by Itô's formula (rather than through the boundedness hypothesis of Proposition 3.1.4). Three features of (3.8) make it a standard model for an asset price.*

(i) *$X_t \geq 0$ for every $t \in [0,\tau]$, since X is defined as an exponential.*

(ii) *The infinitesimal drift $(\mu + \sigma^2/2)X_t$ is linear in the current price X_t.*

(iii) *The infinitesimal standard deviation is proportional to X_t, so that absolute fluctuations grow with the price level.*

These features are the basis of the Black–Scholes model of option pricing.

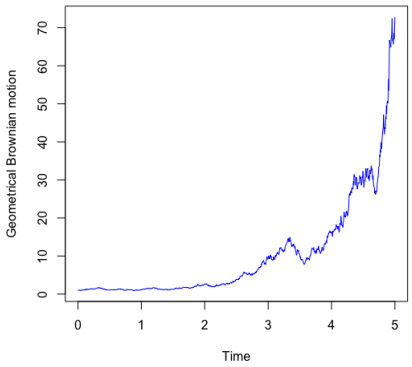


Figure 3.1 – A simulated path of geometric Brownian motion (3.8), with $\mu = 1$, $\sigma = 1/2$, $\alpha = 1$.

Proposition 3.1.8: Ornstein–Uhlenbeck process

Let $\alpha, \sigma > 0$, $x_0 \in \mathbb{R}$, and let W be a real-valued Wiener process. Set

$$X_t := e^{-\alpha t}\Big(x_0 + \sigma \int_0^t e^{\alpha s}\, dW_s\Big), \qquad t \in [0,\tau]. \tag{3.11}$$

Then X is a solution (Definition 3.1.1, $n = d = 1$) of

$$dX_t = -\alpha X_t\, dt + \sigma\, dW_t, \qquad X_0 = x_0, \tag{3.12}$$

and, for every $t \in [0,\tau]$, we have

$$X_t \sim \mathcal{N}\Big(x_0 e^{-\alpha t}, \sigma_t^2\Big), \quad \text{with} \quad \sigma_t^2 := \frac{\sigma^2\big(1 - e^{-2\alpha t}\big)}{2\alpha}. \tag{3.13}$$

Proof. Step 1: an auxiliary Itô process. The deterministic function $s \mapsto \sigma e^{\alpha s}$ is continuous, hence bounded, on $[0,\tau]$, so it belongs to $L_a^2([0,\tau])$ (Remark 2.1.7). The process

$$Y_t := x_0 + \sigma \int_0^t e^{\alpha s}\, dW_s, \qquad t \in [0,\tau],$$

is therefore an Itô process (Definition 2.3.1), with vanishing drift and diffusion coefficient $\sigma_s = \sigma e^{\alpha s}$. Furthermore, we obviously have $X_t = e^{-\alpha t} Y_t$.

Step 2: Itô's formula. Set $f(t,y) := e^{-\alpha t} y$, and apply Theorem 2.3.3 to f and Y, extended to this unbounded f by Proposition 2.3.5. We compute $\partial_t f(t,y) = -\alpha e^{-\alpha t} y$, $\partial_y f(t,y) = e^{-\alpha t}$, and $\partial_y^2 f = 0$, and $f(0, Y_0) = x_0$. Owing to (2.56), we get

$$\begin{aligned} X_t = f(t, Y_t) &= x_0 - \alpha \int_0^t e^{-\alpha r} Y_r\, dr + \int_0^t e^{-\alpha r}\, dY_r \\ &= x_0 - \alpha \int_0^t X_r\, dr + \int_0^t e^{-\alpha r}\, \sigma e^{\alpha r}\, dW_r \\ &= x_0 - \alpha \int_0^t X_r\, dr + \sigma \int_0^t dW_r, \end{aligned}$$

where we have used $e^{-\alpha r} Y_r = X_r$ on the second line. This is exactly (3.1) for the dynamics (3.12). Since $\sigma(x) \equiv \sigma$ is trivially bounded and continuous, Remark 3.1.3 gives $\sigma(X_\cdot) \in L_a^2([0,\tau])$, so X solves (3.12) in the sense of Definition 3.1.1.

Step 3: the law of X_t. Fix $t \in [0,\tau]$. The function $h_t := e^{\alpha(\cdot)} \mathbf{1}_{[0,t]}$ belongs to $L^2([0,\tau])$, so Proposition 2.1.19, with $h^1 = h^2 = h_t$, shows that $\int_0^t e^{\alpha s}\, dW_s = \int_0^\tau h_t(s)\, dW_s$ is a centered Gaussian random variable, with variance

$$\mathbb{E}\Big[\Big(\int_0^t e^{\alpha s}\, dW_s\Big)^2\Big] = \int_0^\tau h_t(s)^2\, ds = \int_0^t e^{2\alpha s}\, ds = \frac{e^{2\alpha t} - 1}{2\alpha}. \tag{3.14}$$

By (3.11), X_t is the image of this Gaussian random variable under the affine map $z \mapsto e^{-\alpha t}(x_0 + \sigma z)$, hence itself Gaussian, with mean $x_0 e^{-\alpha t}$ and (thanks to (3.14)) a variance given by

$$\sigma^2 e^{-2\alpha t} \frac{e^{2\alpha t} - 1}{2\alpha} = \frac{\sigma^2\left(1 - e^{-2\alpha t}\right)}{2\alpha},$$

which is (3.13). ■

Remark 3.1.9 (Interpretation)**.** *Equation* (3.12) *models the velocity of a Brownian particle subject to friction, of coefficient* $\alpha > 0$*. The drift* $-\alpha X_t$ *pulls* X *back toward* 0 *at a rate proportional to its current displacement, while the term* $\sigma\, dW_t$ *superimposes a constant intensity of Gaussian noise. As in Remark 3.1.7, Proposition 3.1.4 does not apply directly to* (3.12)*, since* $b(x) = -\alpha x$ *is unbounded on* $\mathbb{R}$*. By* (3.13)*, the law of* X_t *converges, as* $t \to \infty$*, to* $\mathcal{N}(0, \sigma^2/(2\alpha))$*, independently of the initial condition* x_0*: an instance of the equilibrium behavior that friction, balanced against noise, produces.*

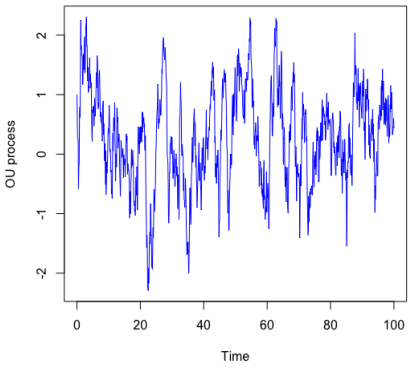


Figure 3.2 – A simulated path of the Ornstein–Uhlenbeck process (3.11), with $\alpha = 1$, $\sigma = 1$, $x_0 = 1$. The stationary nature of the process, described in Remark 3.1.9, is already apparent from the picture: the path fluctuates around 0 with a range that settles down, rather than drifting or spreading out over time.

3.1.3 A diffusion limit of a branching population model

The two previous examples exhibit an explicit solution of a stochastic differential equation directly. This last example goes the other way: it derives a stochastic differential equation as the limit of a family of elementary, discrete population models. The derivation given here is heuristic, and we state it for its own interest, since it explains where the square-root diffusion coefficient below comes from. We refer to Ethier–Kurtz [3] for the rigorous justification of the limit.

Fix $n \geq 1$ and $\alpha > 0$, and consider a Galton–Watson branching process started from $\lfloor n\alpha \rfloor$ individuals. In this model each individual, independently, has a random number of

offspring distributed as a random variable Q_n, with

$$\mathbb{E}[Q_n] \;=\; 1 + \frac{\beta}{n}, \qquad \mathrm{Var}(Q_n) \;=\; \sigma^2, \tag{3.15}$$

for fixed $\beta \in \mathbb{R}$ and $\sigma^2 > 0$. Denote by Z_m^n the population size at generation $m \geq 0$, so $Z_0^n = \lfloor n\alpha \rfloor$, and set $X_t^n := Z_{\lfloor nt \rfloor}^n / n$, a rescaled, continuous-time population size. Summing the mean and variance of $\lfloor n\alpha \rfloor$ independent, identically distributed offspring counts, and rescaling by (3.15), gives

$$\mathbb{E}\Big[X_{1/n}^n - \alpha \,\Big|\, X_0^n = \alpha\Big] \;\approx\; \alpha\beta\,\frac{1}{n}, \qquad \mathrm{Var}\Big(X_{1/n}^n \,\Big|\, X_0^n = \alpha\Big) \;\approx\; \alpha\sigma^2\,\frac{1}{n}. \tag{3.16}$$

We can now compare (3.16) with Proposition 3.1.4. Namely over a time step of length $1/n$, the process X^n has infinitesimal drift $\approx \beta X_0^n$ and infinitesimal variance $\approx \sigma^2 X_0^n$, matching the coefficients

$$b(x) \;=\; \beta x, \qquad \sigma(x) \;=\; \sigma\sqrt{x}, \qquad x \geq 0. \tag{3.17}$$

This suggests, and one can show rigorously [3], that X^n converges in law, as $n \to \infty$, to a process X solving

$$dX_t \;=\; \beta X_t\, dt \;+\; \sigma\sqrt{X_t}\, dW_t, \qquad X_0 = \alpha, \tag{3.18}$$

the *Feller diffusion*. Unlike (3.9) and (3.12), the coefficients (3.17) are not defined (let alone continuous or bounded) on all of $\mathbb{R}$. More specifically σ is only defined on $[0, \infty)$, and fails to be Lipschitz continuous at $x = 0$. Existence and uniqueness of a solution of (3.18), in particular the fact that X stays nonnegative for all times, is a delicate question, outside the scope of Definition 3.1.1 as stated. We do not pursue it further here.

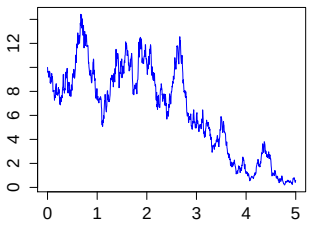


Figure 3.3 – A simulated path of the Feller diffusion (3.18), with $\beta = 0.02$, $\sigma = 2$.

These three examples, linear coefficients solved by direct substitution on the one hand, and the Feller diffusion obtained only as a heuristic limit on the other, already suggest that existence of a solution is not automatic, and that different classes of coefficients call for different tools. This is the subject of the rest of the chapter.

3.2 Equations with additive noise

The two solved examples of the previous section, geometric Brownian motion and the Ornstein–Uhlenbeck process, share a feature that made their explicit solution possible: their coefficients are *linear* in X_s, so that an integrating factor turns the equation into something solvable by direct substitution. This trick breaks down as soon as the drift is no longer linear. There is, however, a different case in which an explicit reduction remains available even for a fully nonlinear drift: equations in which the diffusion coefficient depends on time alone, not on the state X_t. Because the noise term does not see X_t, it can be computed on its own, ahead of time, and subtracted off. The resulting change of variables turns the equation into an ordinary differential equation whose right-hand side depends on ω only through this now-explicit, subtracted noise term, and can therefore be solved path by path with classical, deterministic tools. An equation transformed in this way is called a *random ordinary differential equation*, and the underlying change of variables a special case of the *Doss–Sussmann transformation* [1, 8]. We refer to Han and Kloeden [4] for a systematic treatment.

Definition 3.2.1: Solution of an equation with additive noise

Let $f : [0,\tau] \times \mathbb{R}^n \to \mathbb{R}^n$ and $g : [0,\tau] \to \mathbb{R}^{n\times d}$ be continuous. A continuous, $\mathcal{F}_t$-adapted process $X = \{X_t\,;\, t \in [0,\tau]\}$ with values in $\mathbb{R}^n$ solves the equation with *additive noise* of coefficients (α, f, g), driven by W, if, almost surely, for every $t \in [0,\tau]$,

$$X_t \;=\; \alpha \;+\; \int_0^t f(s, X_s)\, ds \;+\; \int_0^t g(s)\, dW_s. \tag{3.19}$$

Remark 3.2.2 (Well-posedness)**.** *Every term in (3.19) is well defined, exactly as in Remark 3.1.2. That is since X and f are continuous, $s \mapsto f(s, X_s)$ is almost surely continuous, so the ordinary integral $\int_0^t f(s, X_s)\, ds$ makes sense path by path. Furthermore each entry of g, being deterministic and continuous, is bounded on $[0,\tau]$, hence automatically belongs to $L^2_a([0,\tau])$ (Definition 2.1.8). Therefore the stochastic integral $\int_0^t g(s)\, dW_s$ requires no further hypothesis to make sense, unlike the general case of Definition 3.1.1.*

Notation 3.2.3: Additive noise process

Given a continuous function $g : [0,\tau] \to \mathbb{R}^{n\times d}$, we write

$$A_t \;:=\; \int_0^t g(s)\, dW_s, \qquad t \in [0,\tau], \tag{3.20}$$

for the $\mathbb{R}^n$-valued process with components $A^j_t := \sum_{k=1}^d \int_0^t g^{jk}(s)\, dW^k_s$, $1 \le j \le n$ (Notation 2.1.14).

Corollary 3.2.4: Properties of the additive noise process

The process $A = \{A_t\,;\, t \in [0,\tau]\}$ of Notation 3.2.3 admits a continuous, $\mathcal{F}_t$-adapted modification. For this modification, A is a centered Gaussian process (see Definition 1.4.1).

Proof. Step 1: continuity and adaptedness. Fix $1 \le j \le n$. Each entry g^{jk}, $1 \le k \le d$, is deterministic and continuous, hence lies in $L^2([0,\tau])$, viewed as an element of $L^2_a([0,\tau])$ (part (ii) of Remark 2.1.7). By Corollary 2.1.21, applied to the scalar Wiener process W^k, the process $\int_0^\cdot g^{jk}(s)\,dW^k_s$ is $\mathcal{F}_t$-adapted and admits a continuous modification. So does the finite sum $A^j = \sum_{k=1}^d \int_0^\cdot g^{jk}(s)\,dW^k_s$, and hence $A = (A^1,\dots,A^n)$.

Step 2: Gaussianity. Fix $m \ge 1$ and $t_1,\dots,t_m \in [0,\tau]$. By definition, showing that $(A_{t_1},\dots,A_{t_m})$ is a centered Gaussian vector in $\mathbb{R}^{nm}$ amounts to showing that every finite linear combination of its entries,

$$\sum_{i=1}^m c_i\, A^{j_i}_{t_i}, \qquad 1 \le j_1,\dots,j_m \le n, \quad c_1,\dots,c_m \in \mathbb{R},$$

is a centered Gaussian random variable. Fix such $j_1,\dots,j_m$ and $c_1,\dots,c_m$. Grouping terms by k, we have

$$\sum_{i=1}^m c_i\, A^{j_i}_{t_i} \;=\; \sum_{k=1}^d \int_0^\tau \Big(\sum_{i=1}^m c_i\, g^{j_i k}(s)\, \mathbf{1}_{[0,t_i]}(s)\Big)\, dW^k_s.$$

Note that the right hand side above has to be seen as a finite sum, over k, of Wiener integrals of the deterministic, square-integrable function $h^k(s) := \sum_i c_i\, g^{j_i k}(s)\, \mathbf{1}_{[0,t_i]}(s)$. Those Wiener integrals are computed against the independent Wiener processes $W^1,\dots,W^d$ (Remark 1.3.14). Each term $\int_0^\tau h^k(s)\,dW^k_s$ is a centered Gaussian random variable, by part (i) of Proposition 2.1.19. A sum of independent centered Gaussian random variables is itself centered Gaussian, so $\sum_i c_i\, A^{j_i}_{t_i}$ is centered Gaussian. Since $j_1,\dots,j_m$ and $c_1,\dots,c_m$ were arbitrary, $(A_{t_1},\dots,A_{t_m})$ is a centered Gaussian vector in $\mathbb{R}^{nm}$. Since m and $t_1,\dots,t_m$ were themselves arbitrary, this proves that A is a Gaussian process. ∎

With A now known to admit a continuous, adapted, Gaussian modification, we turn to existence and uniqueness for (3.19). For this we label a standard assumption on the drift f.

Assumption 3.2.5: Global Lipschitz drift

There is $L \geq 0$ such that

$$|f(t,x) - f(t,y)| \leq L\,|x-y|, \qquad t \in [0,\tau],\ x,y \in \mathbb{R}^n. \tag{3.21}$$

Under this assumption we can use deterministic type arguments to solve (3.19), as the following proposition shows.

Proposition 3.2.6: Reduction to a random ordinary differential equation

Let $g : [0,\tau] \to \mathbb{R}^{n\times d}$ be continuous, let f be as in Definition 3.2.1 and satisfy Assumption 3.2.5, and let A denote the continuous, adapted modification of Corollary 3.2.4. Then there is one, and, up to indistinguishability (Definition 1.2.2), only one, continuous $\mathcal{F}_t$-adapted process X solving (3.19). It is given by

$$X_t \;=\; A_t \,+\, Y_t, \qquad t \in [0,\tau], \tag{3.22}$$

where, almost surely, $Y = \{Y_t\,;\, t \in [0,\tau]\}$ is the unique continuous solution of

$$Y_t \;=\; \alpha \,+\, \int_0^t f(s, Y_s + A_s)\,ds, \qquad t \in [0,\tau]. \tag{3.23}$$

Remark 3.2.7. *Equation (3.23) is formally the ordinary differential equation*

$$\dot{Y}_t = f(t, Y_t + A_t), \qquad Y_0 = \alpha,$$

with random, but path-by-path continuous and deterministic, right-hand side.

Proof of Proposition 3.2.6. Step 1: reduction. Suppose X solves (3.19), and set $Y_t := X_t - A_t$. Subtracting (3.20) from (3.19) gives, almost surely, for every $t \in [0,\tau]$,

$$Y_t \;=\; \alpha \,+\, \int_0^t f(s, X_s)\,ds \;=\; \alpha \,+\, \int_0^t f(s, Y_s + A_s)\,ds,$$

which is (3.23). Conversely, if Y solves (3.23), then $X := Y + A$ solves (3.19), by the same computation read backwards. So X solves (3.19) if and only if $Y := X - A$ solves (3.23), and it suffices to prove existence and pathwise uniqueness for Y.

Step 2: pathwise existence and uniqueness. Fix ω in the almost-sure event on which $A(\cdot,\omega)$ is continuous (Corollary 3.2.4), and set $F^\omega(t,y) := f(t, y + A_t(\omega))$, for $(t,y) \in [0,\tau]\times\mathbb{R}^n$. Since f is continuous and $A(\cdot,\omega)$ is continuous, F^ω is continuous on $[0,\tau]\times\mathbb{R}^n$, and, by Assumption 3.2.5,

$$|F^\omega(t,y) - F^\omega(t,y')| \;=\; |f(t, y + A_t(\omega)) - f(t, y' + A_t(\omega))| \;\leq\; L\,|y - y'|,$$

for every $t \in [0,\tau]$ and $y, y' \in \mathbb{R}^n$, so F^ω is globally Lipschitz in y, uniformly in t, with the same constant L as f. By the classical Cauchy–Lipschitz theorem, for an ordinary differential equation with right-hand side jointly continuous in t and globally Lipschitz in y uniformly in t (see e.g. Hartman [5]), the equation $Y_t(\omega) = \alpha + \int_0^t F^\omega(s, Y_s(\omega))\,ds$ has a unique continuous solution on the whole interval $[0,\tau]$, for every such ω. This gives existence of a continuous process Y solving (3.23). It also gives uniqueness up to indistinguishability: if Y and Y' are two continuous processes solving (3.23), then, for every ω in the intersection of their two almost-sure defining events, $Y(\cdot,\omega)$ and $Y'(\cdot,\omega)$ are both continuous solutions of the same deterministic equation above. Therefore they coincide, by the uniqueness we just proved. So $Y = Y'$ outside a negligible set, simultaneously for every $t \in [0,\tau]$. Otherwise stated, Y and Y' are indistinguishable.

Step 3: adaptedness. We check that the solution Y of Step 2 is $\mathcal{F}_t$-adapted, so that $X = Y + A$ is a legitimate solution of (3.19) in the sense of Definition 3.2.1. The Cauchy–Lipschitz theorem constructs Y as the almost sure, uniform-in-t limit of the successive approximations $Y_t^{(0)} := \alpha$,

$$Y_t^{(k+1)} \;:=\; \alpha \;+\; \int_0^t f\Big(s, Y_s^{(k)} + A_s\Big)\,ds, \qquad k \geq 0.$$

By induction on k: $Y^{(0)}$ is constant, hence adapted, and if $Y^{(k)}$ is continuous and adapted, then so is $s \mapsto f(s, Y_s^{(k)} + A_s)$, being a continuous function of the continuous, adapted processes $Y^{(k)}$ and A, so its integral up to time t is $\mathcal{F}_t$-measurable, exactly as in Remark 3.2.2, and $Y^{(k+1)}$ is again continuous and adapted. Hence every $Y_t^{(k)}$ is $\mathcal{F}_t$-measurable, and so is its almost sure limit Y_t, which proves that Y is $\mathcal{F}_t$-adapted. Together with Step 2, this shows that $X = Y + A$ is a continuous, $\mathcal{F}_t$-adapted solution of (3.19). This finishes the proof. ∎

Remark 3.2.8 (Generative diffusion models)**.** *Equation* (3.19)*, in the special case where f is linear in x, is exactly the forward noising process at the heart of score-based generative models, taken up in Chapter* **??***. There, having g depend on time alone is what keeps the conditional law of X_t given X_0 explicitly Gaussian, as follows from Corollary 3.2.4 once the linear ordinary differential equation* (3.23) *for Y is solved by an integrating factor, as in the previous section, and this explicit Gaussian law is essential to training such models efficiently.*

Proposition 3.2.6 covers every case in which the diffusion coefficient does not depend on the state, but says nothing when σ genuinely depends on X_t, as in geometric Brownian motion or the Ornstein–Uhlenbeck process of the previous section. Handling that general case, without an explicit change of variables to fall back on, is the subject of this section.

3.3 Existence and uniqueness

When σ depends on the state X_t, the noise term of (3.1) can no longer be computed ahead of time and subtracted off, so the reduction of Section 3.2 is no longer available. We instead

adapt, to this stochastic setting, the classical strategy for ordinary differential equations already used in the proof of Proposition 3.2.6: under a global Lipschitz assumption on the coefficients, a Picard iteration converges to a solution, and the same Lipschitz control forces any two solutions to coincide.

3.3.1 Statement of the theorem

Throughout this section we keep the stochastic basis and the d-dimensional Wiener process W fixed as in Subsection 3.1.1, and we return to equation (3.1), with coefficients (α, b, σ) as in Definition 3.1.1. Since the basis, and in particular the filtration $(\mathcal{F}_t)_{t\in[0,\tau]}$, is fixed once and for all, two notions specific to this setting are worth making precise before we state the main result: what it means for a solution to be built from W alone, and what it means for two solutions to coincide.

Definition 3.3.1: Strong solution

Let $(\mathcal{F}_t^W)_{t\in[0,\tau]}$ denote the (completed) natural filtration of W (Remark 1.2.7), so that $\mathcal{F}_t^W \subset \mathcal{F}_t$ for every $t \in [0,\tau]$. We say that a solution X of (3.1) (Definition 3.1.1) is a *strong solution* if X is, in addition, $\mathcal{F}_t^W$-adapted.

Remark 3.3.2. *Equivalently, X_t is, for every $t \in [0,\tau]$, a measurable function of the initial condition α and of the path $\{W_s\,;\, 0 \le s \le t\}$ alone, not of any further randomness that the ambient filtration $(\mathcal{F}_t)_{t\in[0,\tau]}$ might carry. When b or σ fail to be Lipschitz, existence may hold only after enlarging the probability space itself (a* weak *solution), without X being a strong solution in this sense. This phenomenon does not occur under the hypotheses of this section.*

Definition 3.3.3: Pathwise uniqueness

We say that (3.1) enjoys *pathwise uniqueness* if, whenever X^1 and X^2 are two solutions of (3.1) (Definition 3.1.1) with the same coefficients (α, b, σ), X^1 and X^2 are indistinguishable (Definition 1.2.2):

$$\mathbb{P}\Big(X_t^1 = X_t^2 \text{ for every } t \in [0,\tau]\Big) \;=\; 1. \tag{3.24}$$

We can now state the hypothesis under which we prove existence and pathwise uniqueness.

Assumption 3.3.4: Global Lipschitz coefficients

There exist constants $L_b, L_\sigma \ge 0$ such that, for every $x, y \in \mathbb{R}^n$,

$$|b(x) - b(y)| \;\le\; L_b\,|x-y|, \qquad |\sigma^j(x) - \sigma^j(y)| \;\le\; L_\sigma\,|x-y|, \quad 1 \le j \le d. \tag{3.25}$$

The natural space in which to measure solutions, and the distance between them, is the following.

Definition 3.3.5: The space $L^2(\Omega;\mathcal{C}([0,\tau];\mathbb{R}^n))$

Let $\mathcal{C}([0,\tau];\mathbb{R}^n)$ denote the space of continuous functions $f:[0,\tau]\to\mathbb{R}^n$, equipped with the sup norm $\|f\|_{\infty,\tau} := \sup_{t\in[0,\tau]}|f_t|$ (as in Definition 1.4.6). We denote by $L^2(\Omega;\mathcal{C}([0,\tau];\mathbb{R}^n))$, abbreviated $\mathcal{H}_\tau$, the set of $\mathbb{R}^n$-valued, continuous, $\mathcal{F}_t$-adapted processes $Z=\{Z_t\,;\,t\in[0,\tau]\}$ such that

$$\|Z\|_{\mathcal{H}_\tau}^2 := \mathbb{E}\Big[\|Z\|_{\infty,\tau}^2\Big] = \mathbb{E}\Big[\sup_{t\in[0,\tau]}|Z_t|^2\Big] < \infty. \tag{3.26}$$

Theorem 3.3.6: Existence and pathwise uniqueness

Let $\alpha\in\mathbb{R}^n$ and let b,σ satisfy Assumption 3.3.4. Then equation (3.1), with coefficients (α,b,σ), admits a strong solution X (Definition 3.3.1) belonging to $\mathcal{H}_\tau$ (Definition 3.3.5), and it enjoys pathwise uniqueness (Definition 3.3.3) among solutions belonging to $\mathcal{H}_\tau$.

The rest of this section is devoted to the proof of Theorem 3.3.6, in four steps: we set up the Picard iteration and its basic contraction estimate, we use that estimate to show the iteration converges, we identify its limit as a solution, and we use the same contraction estimate a second time, together with a deterministic comparison lemma, to obtain pathwise uniqueness.

3.3.2 The Picard iteration

As in the classical Picard–Lindelöf proof of the Cauchy–Lipschitz theorem invoked in the previous section (Proposition 3.2.6), we construct the solution X as the limit of a sequence of successive approximations, each obtained by feeding the previous one into the right side of (3.1).

Notation 3.3.7: Picard map and Picard iterations

Given $Y \in \mathcal{H}_\tau$ (Definition 3.3.5), we set $\Gamma(Y) := \widetilde{Y}$, the $\mathbb{R}^n$-valued process defined by

$$\widetilde{Y}_t := \alpha + \int_0^t b(Y_s)\,ds + \sum_{j=1}^{d} \int_0^t \sigma^j(Y_s)\,dW_s^j, \qquad t \in [0,\tau]. \tag{3.27}$$

Starting from $X_t^{(0)} := \alpha$, $t \in [0,\tau]$, we define the *Picard iterations* $\{X^{(n)}\,;\, n \geq 0\}$ by

$$X^{(n+1)} := \Gamma\big(X^{(n)}\big), \qquad n \geq 0. \tag{3.28}$$

By construction, and comparing (3.27) with (3.1), a process $X \in \mathcal{H}_\tau$ solves (3.1) if and only if $X = \Gamma(X)$: existence and uniqueness for (3.1) in $\mathcal{H}_\tau$ amount to existence and uniqueness of a fixed point of Γ.

The next lemma is the technical heart of the section: it bounds the effect of Γ on the distance between two processes, in $\mathcal{H}_\tau$, by the same distance, up to a constant that is linear in the elapsed time. It will be used twice, once to show the Picard iterations converge, and once to prove pathwise uniqueness.

Lemma 3.3.8: Contraction bound for Γ

Under Assumption 3.3.4, let $Y, Z \in \mathcal{H}_\tau$ and set $\widetilde{Y} := \Gamma(Y)$, $\widetilde{Z} := \Gamma(Z)$ (Notation 3.3.7). With

$$c_{L,\tau,d} := 2\tau L_b^2 + 8d^2 L_\sigma^2, \tag{3.29}$$

we have, for every $t \in [0,\tau]$,

$$\mathbb{E}\Big[\sup_{s\in[0,t]} |\widetilde{Y}_s - \widetilde{Z}_s|^2\Big] \leq c_{L,\tau,d} \int_0^t \mathbb{E}\big[|Y_r - Z_r|^2\big]\,dr. \tag{3.30}$$

Proof. Step 1: decomposition. Fix $t \in [0,\tau]$. Subtracting the two instances of (3.27) gives

$$\widetilde{Y}_t - \widetilde{Z}_t = \int_0^t \big[b(Y_s) - b(Z_s)\big]\,ds + \sum_{j=1}^{d} \int_0^t \big[\sigma^j(Y_s) - \sigma^j(Z_s)\big]\,dW_s^j \equiv J_{t,1} + J_{t,2}. \tag{3.31}$$

The elementary bound $|u+v|^2 \leq 2|u|^2 + 2|v|^2$ then gives, for every $s \in [0,\tau]$,

$$\sup_{r\in[0,s]} |\widetilde{Y}_r - \widetilde{Z}_r|^2 \leq 2 \sup_{r\in[0,s]} |J_{r,1}|^2 + 2 \sup_{r\in[0,s]} |J_{r,2}|^2. \tag{3.32}$$

Step 2: the Lebesgue-integral term $J_{t,1}$. By the Cauchy–Schwarz inequality on $[0,r] \subset [0,\tau]$, and the Lipschitz bound on b (Assumption 3.3.4),

$$|J_{r,1}|^2 \leq r \int_0^r |b(Y_s) - b(Z_s)|^2\,ds \leq \tau L_b^2 \int_0^r |Y_s - Z_s|^2\,ds, \qquad r \in [0,\tau].$$

The right side of the above equation is nondecreasing in r. Thus for every $t \in [0,\tau]$, we get

$$\mathbb{E}\Big[\sup_{r\in[0,t]}|J_{r,1}|^2\Big] \le \tau L_b^2 \int_0^t \mathbb{E}\big[|Y_s - Z_s|^2\big]\,ds. \tag{3.33}$$

Step 3: the stochastic-integral term $J_{t,2}$. Fix $1 \le j \le d$ and $1 \le l \le n$, and set

$$M_t^{j,l} := \int_0^t \big[\sigma^{lj}(Y_s) - \sigma^{lj}(Z_s)\big]\,dW_s^j.$$

Since $Y, Z \in \mathcal{H}_\tau$ and $|\sigma^{lj}(x) - \sigma^{lj}(y)| \le |\sigma^j(x) - \sigma^j(y)| \le L_\sigma|x-y|$ by Assumption 3.3.4, the integrand $\sigma^{lj}(Y_\cdot) - \sigma^{lj}(Z_\cdot)$ belongs to $L_a^2([0,\tau])$ (Definition 2.1.8). Therefore Corollary 2.1.21 applies: $M^{j,l}$ is a continuous $\mathcal{F}_t$-martingale. Proposition 2.1.20, applied on $[0,t]$, together with the isometry (2.22) and the Lipschitz bound on σ^j, gives

$$\mathbb{E}\Big[\sup_{r\in[0,t]}(M_r^{j,l})^2\Big] \le 4\,\mathbb{E}\big[(M_t^{j,l})^2\big] = 4\int_0^t \mathbb{E}\big[(\sigma^{lj}(Y_s) - \sigma^{lj}(Z_s))^2\big]\,ds.$$

Writing $J_{t,2} = \sum_{j=1}^d M_t^j$, with $M_t^j := (M_t^{j,1}, \dots, M_t^{j,n})$, the bound $|\sum_j v_j|^2 \le d\sum_j |v_j|^2$ and $|M_r^j|^2 = \sum_{l=1}^n (M_r^{j,l})^2$ give

$$\begin{aligned}\mathbb{E}\Big[\sup_{r\in[0,t]}|J_{r,2}|^2\Big] &\le d\sum_{j=1}^d \mathbb{E}\Big[\sup_{r\in[0,t]}|M_r^j|^2\Big] \le d\sum_{j=1}^d\sum_{l=1}^n \mathbb{E}\Big[\sup_{r\in[0,t]}(M_r^{j,l})^2\Big] \\ &\le 4d\sum_{j=1}^d \int_0^t \mathbb{E}\big[|\sigma^j(Y_s) - \sigma^j(Z_s)|^2\big]\,ds \le 4d^2L_\sigma^2\int_0^t \mathbb{E}\big[|Y_s - Z_s|^2\big]\,ds.\end{aligned}$$

Step 4: conclusion. Combining (3.32) at $s = t$ with the bounds of Steps 2 and 3, we end up with

$$\mathbb{E}\Big[\sup_{r\in[0,t]}|\widetilde{Y}_r - \widetilde{Z}_r|^2\Big] \le 2\tau L_b^2\int_0^t \mathbb{E}\big[|Y_s - Z_s|^2\big]\,ds + 8d^2L_\sigma^2\int_0^t \mathbb{E}\big[|Y_s - Z_s|^2\big]\,ds.$$

We have thus achieved our claim (3.30), recalling the definition (3.29) of $c_{L,\tau,d}$. ∎

Remark 3.3.9 (Γ is well defined on $\mathcal{H}_\tau$)**.** *Let $Y \in \mathcal{H}_\tau$. Owing to Assumption 3.3.4, we have*

$$|b(y)| \le |b(0)| + L_b|y|, \quad \text{and} \quad |\sigma^j(y)| \le |\sigma^j(0)| + L_\sigma|y|,$$

for every $y \in \mathbb{R}^n$ and $1 \le j \le d$, so $b(Y_\cdot)$. In addition, every process of the form $\sigma^j(Y_\cdot)$ is continuous, $\mathcal{F}_t$-adapted, and belongs to $L_a^2([0,\tau])$ componentwise, since $Y \in \mathcal{H}_\tau$ (see Notation 3.3.7). Every term on the right side of (3.27) is therefore well defined, exactly as in Remark 3.1.2, and $\Gamma(Y)$ is itself continuous and $\mathcal{F}_t$-adapted, by Corollary 2.1.21. Applying Lemma 3.3.8 with $Z \equiv 0$, and observing that the process $\Gamma(0)$ has $\|\Gamma(0)\|_{\mathcal{H}_\tau} < \infty$, we obtain that $\Gamma(Y) \in \mathcal{H}_\tau$. In particular, since $X^{(0)} \equiv \alpha \in \mathcal{H}_\tau$, induction on n shows that $X^{(n)} \in \mathcal{H}_\tau$ for every $n \ge 0$.

3.3.3 Convergence of the Picard iterations

We now use Lemma 3.3.8 to show that the Picard iterations converge, first almost surely and uniformly on $[0,\tau]$, then in $\mathcal{H}_\tau$.

Lemma 3.3.10: Bound for the increments of the Picard iterations

Let $\{X^{(n)}\,;\, n\geq 0\}$ be the Picard iterations (Notation 3.3.7), and set, for $n\geq 1$ and $t\in[0,\tau]$,

$$\Delta_n(t) \;:=\; \mathbb{E}\Big[\sup_{s\in[0,t]} |X_s^{(n)} - X_s^{(n-1)}|^2\Big]. \tag{3.34}$$

Then, with $c_{L,\tau,d}$ as in (3.29), the following upper bound holds true:

$$\Delta_n(t) \;\leq\; \frac{c_1\,(c_{L,\tau,d})^{n-1}\,t^n}{n!}, \qquad t\in[0,\tau],\ n\geq 1, \tag{3.35}$$

where $c_1 := 2\tau|b(\alpha)|^2 + 8d\sum_{j=1}^d |\sigma^j(\alpha)|^2$. In particular,

$$\sup_{t\in[0,\tau]} \Delta_n(t) \;\leq\; \frac{c_1\,(c_{L,\tau,d})^{n-1}\,\tau^n}{n!}.$$

Proof. Step 1: the case $n=1$. By (3.27) with $Y\equiv\alpha$ (a deterministic constant), it is readily checked that

$$X_t^{(1)} - X_t^{(0)} \;=\; b(\alpha)\,t \;+\; \sum_{j=1}^d \sigma^j(\alpha)\,W_t^j.$$

The deterministic term contributes $\sup_{s\leq t}|b(\alpha)s|^2 = |b(\alpha)|^2t^2 \leq |b(\alpha)|^2\tau t$. For the stochastic term, $t\mapsto \sigma^j(\alpha)W_t^j$ is, for each j, a continuous $\mathcal{F}_t$-martingale with a deterministic, hence square-integrable, coefficient, so Proposition 2.1.20 gives

$$\mathbb{E}[\sup_{s\leq t}|\sigma^j(\alpha)W_s^j|^2] \leq 4|\sigma^j(\alpha)|^2\mathbb{E}[(W_t^j)^2] = 4|\sigma^j(\alpha)|^2 t.$$

Arguing as in Step 1 of the proof of Lemma 3.3.8 (the elementary bound $|u+v|^2\leq 2|u|^2+2|v|^2$, and $|\sum_j v_j|^2\leq d\sum_j|v_j|^2$ for the sum over j), we obtain

$$\Delta_1(t) \;\leq\; 2|b(\alpha)|^2\tau t \;+\; 8d\,t\sum_{j=1}^d |\sigma^j(\alpha)|^2 \;=\; c_1\,t,$$

which is (3.35) for $n=1$.

Step 2: induction. Let $n\geq 1$ and suppose (3.35) holds for n, for every $t\in[0,\tau]$. Thanks to (3.28), we trivially check that

$$X^{(n+1)} - X^{(n)} = \Gamma(X^{(n)}) - \Gamma(X^{(n-1)}).$$

Therefore Lemma 3.3.8, applied with $Y = X^{(n)}$, $Z = X^{(n-1)}$ (both in $\mathcal{H}_\tau$ by Remark 3.3.9), gives

$$\Delta_{n+1}(t) \;\le\; c_{L,\tau,d} \int_0^t \mathbb{E}\big[|X_r^{(n)} - X_r^{(n-1)}|^2\big]\, dr \;\le\; c_{L,\tau,d} \int_0^t \Delta_n(r)\, dr,$$

where the last inequality bounds $\mathbb{E}[|X_r^{(n)} - X_r^{(n-1)}|^2]$ by

$$\mathbb{E}[\sup_{s\le r} |X_s^{(n)} - X_s^{(n-1)}|^2] = \Delta_n(r).$$

Now by the induction hypothesis, it holds

$$\Delta_{n+1}(t) \;\le\; c_{L,\tau,d} \int_0^t \frac{c_1\,(c_{L,\tau,d})^{n-1}\, r^n}{n!}\, dr \;=\; \frac{c_1\,(c_{L,\tau,d})^n\, t^{n+1}}{(n+1)!},$$

which is (3.35) for $n+1$. This finishes the induction, and the proof. ∎

Lemma 3.3.11: Almost sure convergence of the Picard iterations

Let $\{X^{(n)}\,;\, n \ge 0\}$ be the Picard iterations (Notation 3.3.7). Then, almost surely, $X^{(n)}$ converges, uniformly on $[0,\tau]$, to a continuous process $X = \{X_t\,;\, t \in [0,\tau]\}$.

Proof. Step 1: reduction to a series. Telescoping,

$$X^{(n)} \;=\; X^{(0)} \;+\; \sum_{k=0}^{n-1} \big(X^{(k+1)} - X^{(k)}\big), \qquad n \ge 1,$$

so $(X^{(n)})_{n\ge0}$ converges uniformly on $[0,\tau]$ as soon as the series $\sum_{k\ge0} \|X^{(k+1)} - X^{(k)}\|_{\infty,\tau}$ converges: on the event where it does, the partial sums form a uniformly Cauchy sequence of continuous functions on $[0,\tau]$, which converges uniformly to a continuous limit.

Step 2: Borel–Cantelli. Let $A_n := (\|X^{(n)} - X^{(n-1)}\|_{\infty,\tau} \ge 2^{-n})$. By Chebyshev's inequality and Lemma 3.3.10, applied at $t = \tau$,

$$\mathbb{P}(A_n) \;\le\; 4^n\, \Delta_n(\tau) \;\le\; 4^n\, \frac{c_1\,(c_{L,\tau,d})^{n-1}\, \tau^n}{n!} \;=\; \frac{c_1}{c_{L,\tau,d}} \cdot \frac{(4c_{L,\tau,d}\tau)^n}{n!}.$$

Since $\sum_{n\ge1} (4c_{L,\tau,d}\tau)^n/n! = \exp(4c_{L,\tau,d}\tau) - 1 < \infty$, we get $\sum_{n\ge1} \mathbb{P}(A_n) < \infty$, so, by the (first) Borel–Cantelli lemma, $\mathbb{P}(\limsup_n A_n) = 0$.

Step 3: conclusion. On the almost-sure event where only finitely many A_n occur, there is $n_0(\omega)$ such that $\|X^{(n)}(\omega) - X^{(n-1)}(\omega)\|_{\infty,\tau} < 2^{-n}$ for every $n \ge n_0(\omega)$, so $\sum_{n\ge0} \|X^{(n+1)}(\omega) - X^{(n)}(\omega)\|_{\infty,\tau} < \infty$ by comparison with the convergent geometric series $\sum_n 2^{-n}$. By Step 1, this proves almost-sure uniform convergence of $X^{(n)}$ to a continuous limit X. ∎

Lemma 3.3.12: $\mathcal{H}_\tau$ convergence of the Picard iterations

Let X be the almost-sure uniform limit of $\{X^{(n)}\,;\, n \geq 0\}$ given by Lemma 3.3.11. Then $X \in \mathcal{H}_\tau$ and

$$\lim_{n\to\infty} \mathbb{E}\Big[\|X^{(n)} - X\|_{\infty,\tau}^2\Big] = 0. \tag{3.36}$$

Proof. Step 1: a Cauchy criterion. For $n > m \geq 0$, the triangle inequality for $\|\cdot\|_{\infty,\tau}$, followed by the triangle inequality in $L^2(\Omega)$, gives

$$\Big\| \|X^{(n)} - X^{(m)}\|_{\infty,\tau} \Big\|_{L^2(\Omega)} \leq \sum_{k=m}^{n-1} \Big\| \|X^{(k+1)} - X^{(k)}\|_{\infty,\tau} \Big\|_{L^2(\Omega)} = \sum_{k=m}^{n-1} \Delta_{k+1}(\tau)^{1/2}. \tag{3.37}$$

Now we have seen in Lemma 3.3.10 that

$$\Delta_{k+1}(\tau)^{1/2} \leq \Big(c_1 (c_{L,\tau,d})^k \tau^{k+1}/(k+1)!\Big)^{1/2},$$

a term whose series in k converges. Hence the right side of (3.37) tends to 0 as $m, n \to \infty$. This is exactly the statement that $(\,\|X^{(n)} - X^{(m)}\|_{\infty,\tau}\,)_{n\geq 0}$ is Cauchy in $L^2(\Omega)$, jointly in m, n.

Step 2: passing to the limit. Fix $n \geq 0$. Almost surely, $\|X^{(n)} - X^{(m)}\|_{\infty,\tau} \to \|X^{(n)} - X\|_{\infty,\tau}$ as $m \to \infty$, by Lemma 3.3.11. Fatou's lemma, applied to $(\|X^{(n)} - X^{(m)}\|_{\infty,\tau}^2)_{m\geq 0}$ as $m \to \infty$, together with Step 1, then gives

$$\mathbb{E}\Big[\|X^{(n)} - X\|_{\infty,\tau}^2\Big] \leq \liminf_{m\to\infty} \mathbb{E}\Big[\|X^{(n)} - X^{(m)}\|_{\infty,\tau}^2\Big] \leq \Big(\sum_{k=n}^{\infty} \Delta_{k+1}(\tau)^{1/2}\Big)^2 \xrightarrow[n\to\infty]{} 0,$$

again using the convergence of $\sum_k \Delta_{k+1}(\tau)^{1/2}$ from Step 1. This proves (3.36). Finally, $\|X\|_{\mathcal{H}_\tau} \leq \|X^{(0)}\|_{\mathcal{H}_\tau} + \|X^{(0)} - X\|_{\mathcal{H}_\tau} < \infty$, and X is continuous and $\mathcal{F}_t$-adapted (Lemma 3.3.11), so $X \in \mathcal{H}_\tau$. ∎

Remark 3.3.13 (Role of Lemma 3.3.11)**.** *Bound (3.37), from Step 1 of the proof above, only shows that $(X^{(n)})_{n\geq 0}$ is Cauchy for the norm of $\mathcal{H}_\tau$. This does not, by itself, produce a limit process: doing so directly would require completeness of $\mathcal{H}_\tau$, a fact we have not established. Instead we reused the explicit almost-sure, uniform limit X already constructed in Lemma 3.3.11, by an elementary Borel–Cantelli argument. Then Step 2 of the proof above only had to check, via Fatou's lemma, that this particular, genuinely continuous X is also the $\mathcal{H}_\tau$-limit of $(X^{(n)})_{n\geq 0}$. Lemma 3.3.11 is thus not merely a convenience: it stands in for the completeness argument that Definition 3.3.5 does not supply on its own.*

3.3.4 Proof of the theorem

It remains to identify the limit X of Lemma 3.3.12 as a (strong) solution of (3.1), and to prove pathwise uniqueness.

Lemma 3.3.14: The limit solves the equation

Let $X \in \mathcal{H}_\tau$ be the limit of Lemma 3.3.12. Then $X = \Gamma(X)$, where Γ is introduced in Notation 3.3.7. Hence X is a solution of (3.1) as given in Definition 3.1.1.

Proof. Set $\widehat{X} := \Gamma(X) \in \mathcal{H}_\tau$ (Remark 3.3.9). By Lemma 3.3.8, applied with $Y = X$, $Z = X^{(n)}$, at $t = \tau$,

$$\begin{aligned} \|\widehat{X} - X^{(n+1)}\|_{\mathcal{H}_\tau}^2 &= \mathbb{E}\Big[\|\Gamma(X) - \Gamma(X^{(n)})\|_{\infty,\tau}^2\Big] \\ &\le c_{L,\tau,d} \int_0^\tau \mathbb{E}\Big[|X_s - X_s^{(n)}|^2\Big]\, ds \\ &\le c_{L,\tau,d}\, \tau\, \|X - X^{(n)}\|_{\mathcal{H}_\tau}^2, \end{aligned}$$

which tends to 0 as $n \to \infty$ by Lemma 3.3.12. Next we invoke the triangle inequality for $\|\cdot\|_{\mathcal{H}_\tau}$:

$$\|\widehat{X} - X\|_{\mathcal{H}_\tau} \le \|\widehat{X} - X^{(n+1)}\|_{\mathcal{H}_\tau} + \|X^{(n+1)} - X\|_{\mathcal{H}_\tau},$$

which yields $\|\widehat{X} - X\|_{\mathcal{H}_\tau} = 0$, by letting $n \to \infty$ and invoking Lemma 3.3.12 once more for the second term. Hence

$$\mathbb{E}[\sup_{t\in[0,\tau]} |\widehat{X}_t - X_t|^2] = 0,$$

so $\widehat{X}_t = X_t$ for every $t \in [0,\tau]$ simultaneously, almost surely. We have thus proved that $X = \Gamma(X)$. ∎

Remark 3.3.15 (X is a strong solution)**.** *Every Picard iteration $X^{(n)}$ is built, through (3.27)–(3.28), from the deterministic α and from Lebesgue and Itô integrals of functions of W against W itself. An immediate induction on n, using Definition 2.1.13 and Remark 1.2.7, shows that $X_t^{(n)}$ is $\mathcal{F}_t^W$-measurable for every $t \in [0,\tau]$ and $n \ge 0$. By Lemma 3.3.11, X_t is, for every t, an almost-sure limit of $\mathcal{F}_t^W$-measurable random variables, hence itself measurable with respect to the completion of $\mathcal{F}_t^W$. So the solution X of Lemma 3.3.14 is a strong solution (Definition 3.3.1).*

Existence is now proved. For uniqueness we first label a purely deterministic comparison lemma, which is classical in the theory of ordinary differential equations. As we will see, it turns the recursive bound (3.30) into a genuine uniqueness statement.

Lemma 3.3.16: Gronwall's lemma

Let $\varphi : [0,\tau] \to \mathbb{R}_+$ be continuous, and suppose there are constants $c, \kappa \ge 0$ such that

$$\varphi_t \le c + \kappa \int_0^t \varphi_s\, ds, \qquad t \in [0,\tau]. \tag{3.38}$$

Then

$$\varphi_t \le c \exp(\kappa t), \qquad t \in [0,\tau]. \tag{3.39}$$

Proof. Step 1: a majorizing function. Fix $\epsilon > 0$ and set $\psi_t := (c+\epsilon)\exp(\kappa t)$, $t \in [0,\tau]$. By (3.38) at $t = 0$, $\varphi_0 \le c < c + \epsilon = \psi_0$. If $\varphi_t < \psi_t$ for every $t \in [0,\tau]$, we are done, letting $\epsilon \to 0^+$ at the end of the proof. Otherwise, set $t^* := \inf\{t \in [0,\tau]\,;\, \varphi_t \ge \psi_t\}$. Since $\varphi_0 < \psi_0$ and both are continuous, we must have $t^* > 0$. Hence invoking continuity again, we get $\varphi_{t^*} = \psi_{t^*}$ while $\varphi_s < \psi_s$ for every $s < t^*$.

Step 2: contradiction. Using (3.38) at $t = t^*$, and $\varphi_s < \psi_s$ for $s < t^*$ from Step 1,

$$\psi_{t^*} \;=\; c + \epsilon + \kappa \int_0^{t^*} \psi_s \, ds \;>\; c + \kappa \int_0^{t^*} \psi_s \, ds \;\ge\; c + \kappa \int_0^{t^*} \varphi_s \, ds \;\ge\; \varphi_{t^*}.$$

This contradicts $\varphi_{t^*} = \psi_{t^*}$ from Step 1.

Step 3: conclusion. The contradiction in Step 2 rules out the second case of Step 1, so $\varphi_t < (c+\epsilon)\exp(\kappa t)$ for every $t \in [0,\tau]$ and every $\epsilon > 0$. Letting $\epsilon \to 0^+$ gives (3.39). ■

Proposition 3.3.17: Pathwise uniqueness

Under Assumption 3.3.4, equation (3.1) enjoys pathwise uniqueness (Definition 3.3.3) among solutions belonging to $\mathcal{H}_\tau$.

Proof. Step 1: a continuous comparison function. Let $X^1, X^2 \in \mathcal{H}_\tau$ solve (3.1) with the same coefficients (α, b, σ), so that $X^1 = \Gamma(X^1)$ and $X^2 = \Gamma(X^2)$ (Notation 3.3.7). Set

$$\varphi_t := \mathbb{E}[\sup_{s\in[0,t]} |X^1_s - X^2_s|^2],$$

for all $t \in [0,\tau]$, and observe that φ is a finite function since $X^1, X^2 \in \mathcal{H}_\tau$. Pathwise, $t \mapsto \sup_{s\in[0,t]} |X^1_s(\omega) - X^2_s(\omega)|$ is continuous, being the running maximum of a continuous function. Furthermore, it is dominated by the integrable random variable $2\|X^1\|^2_{\infty,\tau} + 2\|X^2\|^2_{\infty,\tau}$. Dominated convergence therefore shows that φ is continuous on $[0,\tau]$.

Step 2: Gronwall. Since $X^1 = \Gamma(X^1)$ and $X^2 = \Gamma(X^2)$, Lemma 3.3.8, applied with $Y = X^1$, $Z = X^2$, gives, for every $t \in [0,\tau]$,

$$\varphi_t \;\le\; c_{L,\tau,d} \int_0^t \mathbb{E}\Big[|X^1_s - X^2_s|^2\Big]\, ds \;\le\; c_{L,\tau,d} \int_0^t \varphi_s \, ds,$$

which is (3.38) with $c = 0$ and $\kappa = c_{L,\tau,d}$. Gronwall's lemma (Lemma 3.3.16) then gives $\varphi_t \le 0 \cdot \exp(c_{L,\tau,d} t) = 0$ for every $t \in [0,\tau]$.

Step 3: conclusion. As a direct application of Step 2, we get $\varphi_\tau = 0$. Otherwise stated, we have

$$\mathbb{E}[\sup_{t\in[0,\tau]} |X^1_t - X^2_t|^2] = 0.$$

This yields $X^1_t = X^2_t$ for every $t \in [0,\tau]$ simultaneously, almost surely: X^1 and X^2 are indistinguishable. ■

Proof of Theorem 3.3.6. Let $X \in \mathcal{H}_\tau$ be the almost-sure uniform limit of the Picard iterations, given by Lemma 3.3.11 and Lemma 3.3.12. By Lemma 3.3.14, X solves (3.1), and, by Remark 3.3.15, X is a strong solution (Definition 3.3.1). This proves existence. Pathwise uniqueness among solutions in $\mathcal{H}_\tau$ is Proposition 3.3.17. This finishes the proof. ■

Remark 3.3.18 (Further generality)**.** *We proved pathwise uniqueness among solutions belonging to $\mathcal{H}_\tau$, a restriction not present in Definition 3.3.3. This restriction is removable, at the cost of a further localization argument: any solution of (3.1) in the sense of Definition 3.1.1 can be shown, via a stopping-time truncation, to coincide with a solution in $\mathcal{H}_\tau$ on an increasing sequence of events covering Ω, which suffices to extend Proposition 3.3.17 to full generality. Theorem 3.3.6 itself extends well beyond the global Lipschitz Assumption 3.3.4; we record two such extensions without proof, referring to Karatzas–Shreve [6] or Revuz–Yor [7] for details.*

(i) *Coefficients b, σ merely locally Lipschitz, with a linear growth bound $|b(x)| + |\sigma(x)| \leq c(1 + |x|)$ for some constant $c > 0$: existence and pathwise uniqueness still hold on all of $[0, \tau]$, the local Lipschitz property giving a solution up to a stopping time through the same Picard scheme applied on balls of increasing radius. Then the linear growth bound rules out explosion of the solution in finite time.*

(ii) *In dimension $n = d = 1$: b Lipschitz and σ merely γ-Hölder continuous with $\gamma \geq 1/2$ still guarantees existence and pathwise uniqueness, through a genuinely different, more delicate argument than the contraction scheme of this section.*

With Theorem 3.3.6, the well-posedness toolkit of this chapter is complete: linear coefficients solved directly by an integrating factor (Subsection 3.1.2), additive noise reduced to a random ordinary differential equation (Proposition 3.2.6), and now genuinely state-dependent, Lipschitz diffusion coefficients handled by a Picard iteration. Chapter **??** builds on exactly this last, most general case: score-based generative models run a stochastic differential equation of the form (3.1) *forward*, from data to noise, and then ask whether the process can be run *backward*, from noise back to data, along a differently defined but related stochastic differential equation. The existence and uniqueness theory of this chapter is what makes both directions of that construction meaningful.

Bibliography

[1] Halim Doss. Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 13(2):99–125, 1977.

[2] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010. ISBN 978-0-521-76539-8. doi: 10.1017/CBO9780511779398. URL https://doi.org/10.1017/CBO9780511779398.

[3] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. ISBN 0-471-08186-8. doi: 10.1002/9780470316658. URL https://doi.org/10.1002/9780470316658. Characterization and convergence.

[4] Xiaoying Han and Peter E. Kloeden. *Random ordinary differential equations and their numerical solution*, volume 85 of *Probability Theory and Stochastic Modelling*. Springer, Singapore, 2017. ISBN 978-981-10-6264-3. doi: 10.1007/978-981-10-6265-0. URL https://doi.org/10.1007/978-981-10-6265-0.

[5] Philip Hartman. *Ordinary differential equations*, volume 38 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. ISBN 0-89871-510-5. doi: 10.1137/1.9780898719222. URL https://doi.org/10.1137/1.9780898719222. Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)], With a foreword by Peter Bates.

[6] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991. ISBN 0-387-97655-8. doi: 10.1007/978-1-4612-0949-2. URL https://doi.org/10.1007/978-1-4612-0949-2.

[7] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7. doi: 10.1007/978-3-662-06400-9. URL https://doi.org/10.1007/978-3-662-06400-9.

[8] Héctor J. Sussmann. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab.*, 6(1):19–41, 1978. doi: 10.1214/aop/1176995608. URL https://doi.org/10.1214/aop/1176995608.