Basics of Stochastic Analysis

©Timo Seppäläinen

This version November 16, 2014

Department of Mathematics, University of Wisconsin–Madison, Madison, Wisconsin 53706

E-mail address: seppalai@math.wisc.edu

This version November 16, 2014.

The author has been supported by NSF grants and by the Wisconsin Alumni Research Foundation.

ABSTRACT. This material is for a course on stochastic analysis at UW–Madison. The text covers the development of the stochastic integral of predictable processes with respect to cadlag semimartingale integrators, Itô's formula in an open domain in $\mathbf{R}^n$, an existence and uniqueness theorem for an equation of the type $dX = dH + F(t, X)\,dY$ where Y is a cadlag semimartingale, and local time and Girsanov's theorem for Brownian motion. There is also a chapter on the integral with respect to the white noise martingale measure and solving the stochastic heat equation with multiplicative noise.

The text is self-contained except for certain basics of integration theory and probability theory which are explained but not proved. In addition, the reader needs to accept without proof two basic martingale theorems: (i) the existence of quadratic variation for a cadlag local martingale; and (ii) the so-called fundamental theorem of local martingales that states the following: given a cadlag local martingale M and a positive constant c, M can be decomposed as $N + A$ where N and A are cadlag local martingales, jumps of N are bounded by c, and A has paths of bounded variation.

This text intends to provide a stepping stone to deeper books such as Karatzas-Shreve and Protter. The hope is that this material is accessible to students who do not have an ideal background in analysis and probability theory, and useful for instructors who (like the author) are not experts on stochastic analysis.

Contents

Chapter 1

Measures, Integrals, and Foundations of Probability Theory

In this chapter we sort out the integrals one typically encounters in courses on calculus, analysis, measure theory, probability theory and various applied subjects such as statistics and engineering. These are the Riemann integral, the Riemann-Stieltjes integral, the Lebesgue integral and the Lebesgue-Stieltjes integral. The starting point is the general Lebesgue integral on an abstract measure space. The other integrals are special cases, even though they have definitions that look different.

This chapter is not a complete treatment of the basics of measure theory. It provides a brief unified explanation for readers who have prior familiarity with various notions of integration. To avoid unduly burdening this chapter, many technical matters that we need later in the book have been relegated to the appendix. For details that we have omitted and for proofs the reader should turn to any of the standard textbook sources, such as Folland [8].

In the second part of the chapter we go over the measure-theoretic foundations of probability theory. Readers who know basic measure theory and measure-theoretic probability can safely skip this chapter.

1.1. Measure theory and integration

A *space* X is in general an arbitrary set. For integration, the space must have two additional structural elements, namely a σ-algebra and a measure.

1.1.1. σ-algebras. Suppose $\mathcal{A}$ is a collection of subsets of X. (Terms such as *class*, *collection* and *family* are used as synonyms for *set* to avoid speaking of "sets of sets," or even "sets of sets of sets.") Then $\mathcal{A}$ is a *σ-algebra* (also called a *σ-field*) if it has these properties:

(i) $X \in \mathcal{A}$ and $\emptyset \in \mathcal{A}$.

(ii) If $A \in \mathcal{A}$ then also $A^c \in \mathcal{A}$.

(iii) If $\{A_i\}$ is a sequence of sets in $\mathcal{A}$, then also their union $\bigcup_i A_i$ is an element of $\mathcal{A}$.

The restriction to countable unions in part (iii) is crucial. Unions over arbitrarily large collections of sets are not permitted. On the other hand, if part (iii) only permits finite unions, then $\mathcal{A}$ is called an *algebra* of sets, but this is not rich enough for developing a useful theory of integration.

A pair $(X, \mathcal{A})$ where X is a space and $\mathcal{A}$ is a σ-algebra on X is called a *measurable space*. The elements of $\mathcal{A}$ are called *measurable sets*. Suppose $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are two measurable spaces and $f : X \to Y$ is a map (another term for a function) from X into Y. Then f is *measurable* if for every $B \in \mathcal{B}$, the inverse image

$$f^{-1}(B) = \{x \in X : f(x) \in B\} = \{f \in B\}$$

lies in $\mathcal{A}$. Measurable functions are the fundamental object in measure theory. Measurability is preserved by composition $f \circ g$ of functions.

In finite or countable spaces the useful σ-algebra is usually the *power set* 2^X which is the collection of all subsets of X. Not so in uncountable spaces. Furthermore, the important σ-algebras are usually very complicated so that it is impossible to give a concise criterion for testing whether a given set is a member of the σ-algebra. The preferred way to define a σ-algebra is to *generate* it by a smaller collection of sets that can be explicitly described. This procedure is analogous to spanning a subspace of a vector space with a given set of vectors. Except that the generated σ-algebra usually lacks the kind of internal description that vector subspaces have as the set of finite linear combinations of basis vectors. Generation of σ-algebras is based on this lemma whose proof the reader should fill in as an exercise, if this material is new.

Lemma 1.1. *Let Γ be a family of σ-algebras on a space X. Then the intersection*

$$\mathcal{C} = \bigcap_{\mathcal{A} \in \Gamma} \mathcal{A}$$

is also a σ-algebra.

Let $\mathcal{E}$ be an arbitrary collection of subsets of X. The *σ-algebra generated by $\mathcal{E}$*, denoted by $\sigma(\mathcal{E})$, is by definition the intersection of all σ-algebras on

X that contain $\mathcal{E}$. This intersection is well-defined because there is always at least one σ-algebra on X that contains $\mathcal{E}$, namely the power set 2^X. An equivalent characterization of $\sigma(\mathcal{E})$ is that it satisfies these three properties: (i) $\sigma(\mathcal{E}) \supseteq \mathcal{E}$, (ii) $\sigma(\mathcal{E})$ is a σ-algebra on X, and (iii) if $\mathcal{B}$ is any σ-algebra on X that contains $\mathcal{E}$, then $\sigma(\mathcal{E}) \subseteq \mathcal{B}$. This last point justifies calling $\sigma(\mathcal{E})$ the *smallest* σ-algebra on X that contains $\mathcal{E}$.

A related notion is a σ-algebra generated by collection of functions. Suppose $(Y, \mathcal{H})$ is a measurable space, and Φ is a collection of functions from X into Y. Then the σ-algebra generated by Φ is defined by

$$\sigma(\Phi) = \sigma\big\{ \{f \in B\} : f \in \Phi, B \in \mathcal{H}\big\}. \tag{1.1}$$

$\sigma(\Phi)$ is the smallest σ-algebra that makes all the functions in Φ measurable.

Example 1.2 (Borel σ-algebras)**.** If X is a metric space, then the *Borel σ-field* $\mathcal{B}_X$ is the smallest σ-algebra on X that contains all open sets. The members of $\mathcal{B}_X$ are called *Borel sets.* We also write $\mathcal{B}(X)$ when subscripts become clumsy.

It is often technically convenient to have different generating sets for a particular σ-algebra. For example, the Borel σ-algebra $\mathcal{B}_{\mathbf{R}}$ of the real line is generated by either one of these classes of intervals:

$$\big\{(a, b] : -\infty < a < b < \infty\big\} \quad \text{and} \quad \big\{(-\infty, b) : -\infty < b < \infty\big\}.$$

We shall also need the Borel σ-algebra $\mathcal{B}_{[-\infty,\infty]}$ of the extended real line $[-\infty, \infty]$. We define this as the smallest σ-algebra that contains all Borel sets on the real line and the singletons $\{-\infty\}$ and $\{\infty\}$. This σ-algebra is also generated by the intervals $\{[-\infty, b] : b \in \mathbf{R}\}$. It is possible to define a metric on $[-\infty, \infty]$ such that this σ-algebra is the Borel σ-algebra determined by the metric.

When we speak of real-valued or extended real-valued measurable functions on an arbitrary measurable space $(X, \mathcal{A})$, we always have in mind the Borel σ-algebra on $\mathbf{R}$ and $[-\infty, \infty]$. One can then check that measurability is preserved by algebraic operations ($f \pm g$, fg, f/g, whenever these are well-defined) and by pointwise limits and suprema of sequences: if $\{f_n : n \in \mathbf{N}\}$ is a sequence of real-valued measurable functions, then for example the functions

$$g(x) = \sup_{n \in \mathbf{N}} f_n(x) \quad \text{and} \quad h(x) = \varlimsup_{n \to \infty} f_n(x)$$

are measurable. The set $\mathbf{N}$ above is the set of natural numbers $\mathbf{N} = \{1, 2, 3, \dots\}$. Thus measurability is a more robust property than other familiar types of regularity, such as continuity or differentiability, that are not in general preserved by pointwise limits.

If $(X, \mathcal{B}_X)$ is a metric space with its Borel σ-algebra, then every continuous function $f : X \to \mathbf{R}$ is measurable. The definition of continuity implies that for any open $G \subseteq \mathbf{R}$, $f^{-1}(G)$ is an open set in X, hence a member of $\mathcal{B}_X$. Since the open sets generate $\mathcal{B}_{\mathbf{R}}$, this suffices for concluding that $f^{-1}(B) \in \mathcal{B}_X$ for all Borel sets $B \subseteq \mathbf{R}$.

Example 1.3 (Product σ-algebras)**.** Of great importance for probability theory are product σ-algebras. Let $\mathcal{I}$ be an arbitrary index set, and for each $i \in \mathcal{I}$ let $(X_i, \mathcal{A}_i)$ be a measurable space. The *Cartesian product space* $X = \prod_{i \in \mathcal{I}} X_i$ is the space of all functions $x : \mathcal{I} \to \bigcup_{i \in \mathcal{I}} X_i$ such that $x(i) \in X_i$ for each i. Alternate notation for $x(i)$ is x_i. *Coordinate projection maps* on X are defined by $f_i(x) = x_i$, in other words f_i maps X onto X_i by extracting the i-coordinate of the $\mathcal{I}$-tuple x. The *product σ-algebra* $\bigotimes_{i \in \mathcal{I}} \mathcal{A}_i$ is by definition the σ-algebra generated by the coordinate projections $\{f_i : i \in \mathcal{I}\}$.

1.1.2. Measures. Let us move on to discuss the second fundamental ingredient of integration. Let $(X, \mathcal{A})$ be a measurable space. A *measure* is a function $\mu : \mathcal{A} \to [0, \infty]$ that satisfies these properties:

(i) $\mu(\emptyset) = 0$.

(ii) If $\{A_i\}$ is a sequence of sets in $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$ (pairwise disjoint is the term), then

$$\mu\Big(\bigcup_i A_i\Big) = \sum_i \mu(A_i).$$

Property (ii) is called *countable additivity*. It goes together with the fact that σ-algebras are closed under countable unions, so there is no issue about whether the union $\bigcup_i A_i$ is a member of $\mathcal{A}$. The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*.

If $\mu(X) < \infty$ then μ is a *finite measure*. If $\mu(X) = 1$ then μ is a *probability measure*. Infinite measures arise naturally. The ones we encounter satisfy a condition called σ-finiteness: μ is *σ-finite* if there exists a sequence of measurable sets $\{V_i\}$ such that $X = \bigcup V_i$ and $\mu(V_i) < \infty$ for all i. A measure defined on a Borel σ-algebra is called a *Borel measure*.

Example 1.4. Suppose $X = \{x_i : i \in \mathbf{N}\}$ is a countable space, and let $\{a_i : i \in \mathbf{N}\}$ be a sequence of nonnegative real numbers. Then

$$\mu(A) = \sum_{i : x_i \in A} a_i$$

defines a σ-finite (or finite, if $\sum a_i < \infty$) measure on the σ-algebra 2^X of all subsets of X.

The example shows that to define a measure on a countable space, one only needs to specify the measures of the singletons, and the rest follows by countable additivity. Again, things are more complicated in uncountable spaces. For example, we would like to have a measure m on the Borel sets of the real line with the property that the measure $m(I)$ of an interval I is the length of the interval. (This measure is known as the Lebesgue measure.) But then the measure of any singleton must be zero. So there is no way to construct the measure by starting with singletons.

Since it is impossible to describe every element of a σ-algebra, it is even more impossible to give a simple explicit formula for the measure of every measurable set. Consequently we need a theorem that "generates" a measure from some modest ingredients that can be explicitly written down. Here is a useful one.

First, a class $\mathcal{S}$ of subsets of X is a *semialgebra* if it has these properties:

(i) $\emptyset \in \mathcal{S}$

(ii) If $A, B \in \mathcal{S}$ then also $A \cap B \in \mathcal{S}$.

(iii) If $A \in \mathcal{S}$, then A^c is a finite disjoint union of elements of $\mathcal{S}$.

A good example on the real line to keep in mind is

$$\mathcal{S} = \{(a, b] : -\infty \le a \le b < \infty\} \cup \{(a, \infty) : -\infty \le a < \infty\}. \tag{1.2}$$

This semialgebra generates $\mathcal{B}_{\mathbf{R}}$.

Theorem 1.5. *Let $\mathcal{S}$ be a semialgebra, and $\mu_0 : \mathcal{S} \to [0, \infty]$ a function with these properties:*

(i) $\mu_0(\emptyset) = 0$

(ii) *If $A \in \mathcal{S}$ is a finite disjoint union of sets $B_1, \dots, B_n$ in $\mathcal{S}$, then $\mu_0(A) = \sum \mu_0(B_i)$.*

(iii) *If $A \in \mathcal{S}$ is a countable disjoint union of sets $B_1, B_2, \dots, B_n, \dots$ in $\mathcal{S}$, then $\mu_0(A) \le \sum \mu_0(B_i)$.*

Assume furthermore that there exists a sequence of sets $\{A_i\}$ in $\mathcal{S}$ such that $X = \bigcup A_i$ and $\mu_0(A_i) < \infty$ for all i. Then there exists a unique measure μ on the σ-algebra $\sigma(\mathcal{S})$ such that $\mu = \mu_0$ on $\mathcal{S}$.

This theorem is proved by first extending μ_0 to the algebra generated by $\mathcal{S}$, and by then using the so-called Carathéodory Extension Theorem to go from the algebra to the σ-algebra. With Theorem 1.5 we can describe a large class of measures on $\mathbf{R}$.

Example 1.6 (Lebesgue-Stieltjes measures)**.** Let F be a nondecreasing, right-continuous real-valued function on $\mathbf{R}$. For intervals $(a, b]$, define

$$\mu_0(a, b] = F(b) - F(a).$$

This will work also for $a = -\infty$ and $b = \infty$ if we define

$$F(\infty) = \lim_{x \nearrow \infty} F(x) \quad \text{and} \quad F(-\infty) = \lim_{y \searrow -\infty} F(y).$$

It is possible that $F(\infty) = \infty$ and $F(-\infty) = -\infty$, but this will not hurt the definition. One can show that μ_0 satisfies the hypotheses of Theorem 1.5, and consequently there exists a measure μ on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ that gives mass $F(b) - F(a)$ to each interval $(a, b]$. This measure is called the *Lebesgue-Stieltjes measure* of the function F, and we shall denote μ by Λ_F to indicate the connection with F.

The most important special case is *Lebesgue measure* which we shall denote by m, obtained by taking $F(x) = x$.

On the other hand, if μ is a Borel measure on $\mathbf{R}$ such that $\mu(B) < \infty$ for all bounded Borel sets, we can define a right-continuous nondecreasing function by

$$G(0) = 0, \text{ and } G(x) = \begin{cases} \mu(0, x], & x > 0 \\ -\mu(x, 0], & x < 0 \end{cases}$$

and then $\mu = \Lambda_G$. Thus Lebesgue-Stieltjes measures give us all the Borel measures that are finite on bounded sets.

1.1.3. The integral. Let $(X, \mathcal{A}, \mu)$ be a fixed measure space. To say that a function $f : X \to \mathbf{R}$ or $f : X \to [-\infty, \infty]$ is measurable is always interpreted with the Borel σ-algebra on $\mathbf{R}$ or $[-\infty, \infty]$. In either case, it suffices to check that $\{f \le t\} \in \mathcal{A}$ for each real t. The Lebesgue integral is defined in several stages, starting with cases for which the integral can be written out explicitly. This same pattern of proceeding from simple cases to general cases will also be used to define stochastic integrals.

Step 1. *Nonnegative measurable simple functions.* A nonnegative simple function is a function with finitely many distinct values $\alpha_1, \dots, \alpha_n \in [0, \infty)$. If we set $A_i = \{f = \alpha_i\}$, then we can write

$$f(x) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(x)$$

where

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

is the *indicator function* (also called *characteristic function*) of the set A. The integral $\int f \, d\mu$ is defined by

$$\int f \, d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

This sum is well-defined because f is measurable iff each A_i is measurable, and it is possible to add and multiply numbers in $[0, \infty]$. Note the convention $0 \cdot \infty = 0$.

Step 2. $[0, \infty]$-*valued measurable functions.* Let $f : X \to [0, \infty]$ be measurable. Then we define

$$\int f \, d\mu = \sup \left\{ \int g \, d\mu : \ g \text{ is a simple function such that } 0 \leq g \leq f \right\}.$$

This integral is a well-defined number in $[0, \infty]$.

Step 3. *General measurable functions.* Let $f : X \to [-\infty, \infty]$ be measurable. The positive and negative parts of f are $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$. $f^\pm$ are nonnegative functions, and satisfy $f = f^+ - f^-$ and $|f| = f^+ + f^-$. The integral of f is defined by

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$

provided at least one of the integrals on the right is finite.

These steps complete the construction of the integral. Along the way one proves that the integral has all the necessary properties, such as linearity

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu,$$

monotonicity:

$$f \leq g \text{ implies } \int f \, d\mu \leq \int g \, d\mu,$$

and the important inequality

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu.$$

Various notations are used for the integral $\int f \, d\mu$. Sometimes it is desirable to indicate the space over which one integrates by $\int_X f \, d\mu$. Then one can indicate integration over a subset A by defining

$$\int_A f \, d\mu = \int_X \mathbf{1}_A f \, d\mu.$$

To make the integration variable explicit, one can write

$$\int_X f(x) \, \mu(dx) \quad \text{or} \quad \int_X f(x) \, d\mu(x).$$

Since the integral is linear in both the function and the measure, the linear functional notation $\langle f, \mu \rangle$ is used. Sometimes the notation is simplified to $\mu(f)$, or even to μf.

A basic aspect of measure theory is that whatever happens on sets of measure zero is not visible. We say that a property holds *μ-almost everywhere* (or simply *almost everywhere* if the measure is clear from the context) if there exists a set $N \in \mathcal{A}$ such that $\mu(N) = 0$ (a *μ-null* set) and the property in question holds on the set N^c.

For example, we can define a measure μ on $\mathbf{R}$ which agrees with Lebesgue measure on $[0,1]$ and vanishes elsewhere by $\mu(B) = m(B \cap [0,1])$ for $B \in \mathcal{B}_{\mathbf{R}}$. Then if $g(x) = x$ on $\mathbf{R}$ while

$$f(x) = \begin{cases} \sin x, & x < 0 \\ x, & 0 \le x \le 1 \\ \cos x, & x > 1 \end{cases}$$

we can say that $f = g$ μ-almost everywhere.

The principal power of the Lebesgue integral derives from three fundamental convergence theorems which we state next. The value of an integral is not affected by changing the function on a null set. Therefore the hypotheses of the convergence theorems require only almost everywhere convergence.

Theorem 1.7. *(Fatou's lemma) Let $0 \le f_n \le \infty$ be measurable functions. Then*

$$\int \big(\varliminf_{n\to\infty} f_n \big)\, d\mu \le \varliminf_{n\to\infty} \int f_n \, d\mu.$$

Theorem 1.8. *(Monotone convergence theorem) Let f_n be nonnegative measurable functions, and assume $f_n \le f_{n+1}$ almost everywhere, for each n. Let $f = \lim_{n\to\infty} f_n$. This limit exists at least almost everywhere. Then*

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Theorem 1.9. *(Dominated convergence theorem) Let f_n be measurable functions, and assume the limit $f = \lim_{n\to\infty} f_n$ exists almost everywhere. Assume there exists a function $g \ge 0$ such that $|f_n| \le g$ almost everywhere for each n, and $\int g \, d\mu < \infty$. Then*

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Finding examples where the hypotheses and the conclusions of these theorems fail are excellent exercises. By the monotone convergence theorem we can give the following explicit limit expression for the integral $\int f\,d\mu$ of a $[0,\infty]$-valued function f. Define simple functions

$$f_n(x) = \sum_{k=0}^{2^n n-1} 2^{-n}k \cdot \mathbf{1}_{\{2^{-n}k \le f < 2^{-n}(k+1)\}}(x) + n \cdot \mathbf{1}_{\{f \ge n\}}(x). \tag{1.3}$$

Then $0 \le f_n(x) \nearrow f(x)$, and so by Theorem 1.8,

$$\begin{aligned} \int f\,d\mu &= \lim_{n\to\infty} \int f_n\,d\mu \\ &= \lim_{n\to\infty} \biggl\{ \sum_{k=0}^{2^n n-1} 2^{-n}k \cdot \mu\bigl\{\tfrac{k}{2^n} \le f < \tfrac{k+1}{2^n}\bigr\} + n \cdot \mu\{f \ge n\} \biggr\}. \end{aligned} \tag{1.4}$$

There is an abstract change of variables principle which is particularly important in probability. Suppose we have a measurable map $\psi : (X,\mathcal{A}) \to (Y,\mathcal{H})$ between two measurable spaces, and a measurable function $f : (Y,\mathcal{H}) \to (\mathbf{R},\mathcal{B}_{\mathbf{R}})$. If μ is a measure on $(X,\mathcal{A})$, we can define a measure ν on $(Y,\mathcal{H})$ by

$$\nu(U) = \mu(\psi^{-1}(U)) \quad \text{for } U \in \mathcal{H}.$$

In short, this connection is expressed by

$$\nu = \mu \circ \psi^{-1}.$$

If the integral of f over the measure space $(Y,\mathcal{H},\nu)$ exists, then the value of this integral is not changed if instead we integrate $f \circ \psi$ over $(X,\mathcal{A},\mu)$:

$$\int_Y f\,d\nu = \int_X (f \circ \psi)\,d\mu. \tag{1.5}$$

Note that the definition of ν already gives equality (1.5) for $f = \mathbf{1}_U$. The linearity of the integral then gives it for simple f. General $f \ge 0$ follow by monotone convergence, and finally general $f = f^+ - f^-$ by linearity again. This sequence of steps recurs often when an identity for integrals is to be proved.

1.1.4. Completion of measures. There are certain technical benefits to having the following property in a measure space $(X,\mathcal{A},\mu)$, called *completeness*: if $N \in \mathcal{A}$ satisfies $\mu(N) = 0$, then every subset of N is measurable (and then of course has measure zero). It turns out that this can always be arranged by a simple enlargement of the σ-algebra. Let

$$\begin{aligned} \bar{\mathcal{A}} = \{A \subseteq X :\ &\text{there exists } B, N \in \mathcal{A} \text{ and } F \subseteq N \\ &\qquad \text{such that } \mu(N) = 0 \text{ and } A = B \cup F\} \end{aligned}$$

and define $\bar\mu$ on $\bar{\mathcal{A}}$ by $\bar\mu(A) = \mu(B)$ when B has the relationship to A from above. Then one can check that $\mathcal{A} \subseteq \bar{\mathcal{A}}$, $(X,\bar{\mathcal{A}},\bar\mu)$ is a complete measure space, and $\bar\mu$ agrees with μ on $\mathcal{A}$.

An important example of this procedure is the extension of Lebesgue measure m from $\mathcal{B}_{\mathbf{R}}$ to a σ-algebra $\mathcal{L}_{\mathbf{R}}$ of the so-called *Lebesgue measurable sets*. $\mathcal{L}_{\mathbf{R}}$ is the completion of $\mathcal{B}_{\mathbf{R}}$ under Lebesgue measure, and it is strictly larger than $\mathcal{B}_{\mathbf{R}}$. Proving this latter fact is typically an exercise in real analysis (for example, Exercise 2.9 in [8]). For our purposes the Borel sets suffice as a domain for Lebesgue measure. In analysis literature the term Lebesgue measure usually refers to the completed measure.

1.1.5. The Riemann and Lebesgue integrals. In calculus we learn the Riemann integral. Suppose f is a bounded function on a compact interval $[a, b]$. Given a (finite) partition $\pi = \{a = s_0 < s_1 < \cdots < s_n = b\}$ of $[a, b]$ and some choice of points $x_i \in [s_i, s_{i+1}]$, we form the *Riemann sum*

$$S(\pi) = \sum_{i=0}^{n-1} f(x_i)(s_{i+1} - s_i).$$

We say f is *Riemann integrable* on $[a, b]$ if there is a number c such that the following is true: given $\varepsilon > 0$, there exists $\delta > 0$ such that $|c - S(\pi)| \le \varepsilon$ for every partition π with $\text{mesh}(\pi) = \max\{s_{i+1} - s_i\} \le \delta$ and for any choice of the points x_i in the Riemann sum. In other words, the Riemann sums converge to c as the mesh of the partition converges to zero. The limiting value is by definition the *Riemann integral* of f:

$$\int_a^b f(x)\,dx = c = \lim_{\text{mesh}(\pi)\to 0} S(\pi). \tag{1.6}$$

One can then prove that every continuous function is Riemann integrable.

The definition of the Riemann integral is fundamentally different from the definition of the Lebesgue integral. For the Riemann integral there is one recipe for all functions, instead of a step-by-step definition that proceeds from simple to complex cases. For the Riemann integral we partition the domain $[a, b]$, whereas the Lebesgue integral proceeds by partitioning the range of f, as formula (1.4) makes explicit. This latter difference is sometimes illustrated by counting the money in your pocket: the Riemann way picks one coin at a time from the pocket, adds its value to the total, and repeats this until all coins are counted. The Lebesgue way first partitions the coins into pennies, nickles, dimes, and quarters, and then counts the piles. As the coin-counting picture suggests, the Lebesgue way is more efficient (it leads to a more general integral with superior properties) but when both apply, the answers are the same. The precise relationship is the following, which also gives the exact domain of applicability of the Riemann integral.

Theorem 1.10. *Suppose f is a bounded function on $[a, b]$.*

(a) *If f is a Riemann integrable function on $[a, b]$, then f is Lebesgue measurable, and the Riemann integral of f coincides with the Lebesgue integral of f with respect to Lebesgue measure m on $[a, b]$.*

(b) *f is Riemann integrable iff the set of discontinuities of f has Lebesgue measure zero.*

Because of this theorem, the Riemann integral notation is routinely used for Lebesgue integrals on the real line. In other words, we write

$$\int_a^b f(x)\,dx \quad \text{instead of} \quad \int_{[a,b]} f\,dm$$

for a Borel or Lebesgue measurable function f on $[a, b]$, even if the function f is not Riemann integrable.

1.1.6. Function spaces. Various function spaces play an important role in analysis and in all the applied subjects that use analysis. One way to define such spaces is through integral norms. Let $(X, \mathcal{A}, \mu)$ be a measure space. For $1 \le p < \infty$, the space $L^p(\mu)$ is the set of all measurable $f : X \to \mathbf{R}$ such that $\int |f|^p\,d\mu < \infty$. The L^p norm on this space is defined by

$$\|f\|_p = \|f\|_{L^p(\mu)} = \left\{ \int |f|^p\,d\mu \right\}^{\frac{1}{p}}. \tag{1.7}$$

A function f is called *integrable* if $\int |f|\,d\mu < \infty$. This is synonymous with $f \in L^1(\mu)$.

There is also a norm corresponding to $p = \infty$, defined by

$$\|f\|_\infty = \|f\|_{L^\infty(\mu)} = \inf\bigl\{c \ge 0 : \mu\{|f| > c\} = 0\bigr\}. \tag{1.8}$$

This quantity is called the *essential supremum* of $|f|$. The inequality $|f(x)| \le \|f\|_\infty$ holds almost everywhere, but can fail on a null set of points x.

The $L^p(\mu)$ spaces, $1 \le p \le \infty$, are Banach spaces (see Appendix). The type of convergence in these spaces is called L^p convergence, so we say $f_n \to f$ in $L^p(\mu)$ if

$$\|f_n - f\|_{L^p(\mu)} \to 0 \quad \text{as } n \to \infty.$$

However, a problem is created by the innocuous property of a norm that requires $\|f\|_p = 0$ if and only if $f = 0$. For example, let the underlying measure space be the interval $[0, 1]$ with Lebesgue measure on Borel sets. From the definition of the L^p norm then follows that $\|f\|_p = 0$ if and only if $f = 0$ Lebesgue–almost everywhere. In other words, f can be nonzero on even infinitely many points as long as these points form a Lebesgue–null set, and still $\|f\|_p = 0$. An example of this would be the indicator function of the rationals in $[0, 1]$. So the disturbing situation is that *many* functions have zero norm, not just the identically zero function $f(x) \equiv 0$.

To resolve this, we apply the idea that whatever happens on null sets is not visible. We simply adopt the point of view that functions are equal if they differ only on a null set. The mathematically sophisticated way of phrasing this is that we regard elements of $L^p(\mu)$ as *equivalence classes* of functions. Particular functions that are almost everywhere equal are representatives of the same equivalence class. Fortunately, we do not have to change our language. We can go on regarding elements of $L^p(\mu)$ as functions, as long as we remember the convention concerning equality. This issue will appear again when we discuss spaces of stochastic processes.

1.1.7. Product measures. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two σ-finite measure spaces. The *product measure space* $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$ is defined as follows. $X \times Y$ is the Cartesian product space. $\mathcal{A} \otimes \mathcal{B}$ is the product σ-algebra. The product measure $\mu \otimes \nu$ is the unique measure on $\mathcal{A} \otimes \mathcal{B}$ that satisfies

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$$

for *measurable rectangles* $A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Measurable rectangles generate $\mathcal{A} \otimes \mathcal{B}$ and they form a semialgebra. The hypotheses of the Extension Theorem 1.5 can be checked, so the measure $\mu \otimes \nu$ is uniquely and well defined. This measure $\mu \otimes \nu$ is also σ-finite.

The *x-section* f_x of an $\mathcal{A} \otimes \mathcal{B}$-measurable function f is $f_x(y) = f(x, y)$. It is a $\mathcal{B}$-measurable function on Y. Furthermore, the integral of f_x over $(Y, \mathcal{B}, \nu)$ gives an $\mathcal{A}$-measurable function of x. Symmetric statements hold for the y-section $f_y(x) = f(x, y)$, a measurable function on X. This is part of the important Tonelli-Fubini theorem.

Theorem 1.11. *Suppose $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces.*

(a) *(Tonelli's theorem) Let f be a $[0, \infty]$-valued $\mathcal{A} \otimes \mathcal{B}$-measurable function on $X \times Y$. Then the functions $g(x) = \int_Y f_x \, d\nu$ and $h(y) = \int_X f_y \, d\mu$ are $[0, \infty]$-valued measurable functions on their respective spaces. Furthermore, f can be integrated by iterated integration:*

$$\begin{aligned} \int_{X \times Y} f \, d(\mu \otimes \nu) &= \int_X \left\{ \int_Y f(x, y) \, \nu(dy) \right\} \mu(dx) \\ &= \int_Y \left\{ \int_X f(x, y) \, \mu(dx) \right\} \nu(dy). \end{aligned} \tag{1.9}$$

(b) *(Fubini's theorem) Let $f \in L^1(\mu \otimes \nu)$. Then $f_x \in L^1(\nu)$ for μ-almost every x, $f_y \in L^1(\mu)$ for ν-almost every y, $g \in L^1(\mu)$ and $h \in L^1(\nu)$. Iterated integration is valid as in (1.9) above.*

The product measure construction and the theorem generalize naturally to products

$$\Big(\prod_{i=1}^{n} X_i \,,\ \bigotimes_{i=1}^{n} \mathcal{A}_i \,,\ \bigotimes_{i=1}^{n} \mu_i \Big)$$

of finitely many σ-finite measure spaces. Infinite products shall be discussed in conjunction with the construction problem of stochastic processes.

The part of the Tonelli-Fubini theorem often needed is that integrating away some variables from a product measurable function always leaves a function that is measurable in the remaining variables.

In multivariable calculus we learn the multivariate Riemann integral over n-dimensional rectangles,

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \dots, x_n)\, dx_n \cdots dx_2\, dx_1.$$

This is an integral with respect to n-dimensional Lebesgue measure on $\mathbf{R}^n$, which can be defined as the completion of the n-fold product of one-dimensional Lebesgue measures.

As a final technical point, consider metric spaces $X_1, X_2, \dots, X_n$, with product space $X = \prod X_i$. X has the product σ-algebra $\bigotimes \mathcal{B}_{X_i}$ of the Borel σ-algebras from the factor spaces. On the other hand, X is a metric space in its own right, and so has its own Borel σ-algebra $\mathcal{B}_X$. What is the relation between the two? The projection maps $(x_1, \dots, x_n) \mapsto x_i$ are continuous, hence $\mathcal{B}_X$-measurable. Since these maps generate $\bigotimes \mathcal{B}_{X_i}$, it follows that $\bigotimes \mathcal{B}_{X_i} \subseteq \mathcal{B}_X$. It turns out that if the X_i's are *separable* then equality holds: $\bigotimes \mathcal{B}_{X_i} = \mathcal{B}_X$. A separable metric space is one that has a countable dense set. An example of a countable dense set is the set of rational numbers in $\mathbf{R}$.

1.1.8. Signed measures. A finite signed measure μ on a measurable space $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to \mathbf{R}$ such that $\mu(\emptyset) = 0$, and

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1.10}$$

whenever $A = \bigcup A_i$ is a disjoint union. The series in (1.10) has to converge absolutely, meaning that

$$\sum_i |\mu(A_i)| < \infty.$$

Without absolute convergence the limit of the series $\sum \mu(A_i)$ would depend on the order of the terms. But this must not happen because rearranging the sets $A_1, A_2, A_3, \dots$ does not change their union.

More generally, a signed measure is allowed to take one of the values $\pm\infty$ but not both. Absolute convergence in (1.10) is then required if $\mu(A)$ is a finite real number. We shall use the term *measure* only when the signed measure takes only values in $[0,\infty]$. If this point needs emphasizing, we use the term *positive measure* as a synonym for measure.

For any signed measure ν, there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$. (The statement $\nu^+ \perp \nu^-$ reads "ν^+ and ν^- are *mutually singular*", and means that there exists a measurable set A such that $\nu^+(A) = \nu^-(A^c) = 0$.) The measure ν^+ is the *positive variation* of ν, ν^- is the *negative variation* of ν, and the pair ν^+, ν^- is the *Jordan decomposition* of ν. There exist measurable sets P and N such that $P \cup N = X$, $P \cap N = \emptyset$, and $\nu^+(A) = \nu(A \cap P)$ and $\nu^-(A) = -\nu(A \cap N)$. (P, N) is called the *Hahn decomposition* of ν. The *total variation* of ν is the positive measure $|\nu| = \nu^+ + \nu^-$. We say that the signed measure ν is σ-finite if $|\nu|$ is σ-finite.

Integration with respect to a signed measure is defined by

$$\int f\, d\nu = \int f\, d\nu^+ - \int f\, d\nu^- \tag{1.11}$$

whenever both integrals on the right are finite. A function f is *integrable* with respect to ν if it is integrable with respect to $|\nu|$. In other words, $L^1(\nu)$ is by definition $L^1(\,|\nu|\,)$. A useful inequality is

$$\left| \int f\, d\nu \right| \le \int |f|\, d|\nu| \tag{1.12}$$

valid for all f for which the integral on the right is finite.

Note for future reference that integrals with respect to $|\nu|$ can be expressed in terms of ν by

$$\int f\, d|\nu| = \int_P f\, d\nu^+ + \int_N f\, d\nu^- = \int (\mathbf{1}_P - \mathbf{1}_N) f\, d\nu. \tag{1.13}$$

1.1.9. BV functions and Lebesgue-Stieltjes integrals. Let F be a function on $[a,b]$. The *total variation function* of F is the function $V_F(x)$ defined on $[a,b]$ by

$$V_F(x) = \sup \left\{ \sum_{i=1}^n |F(s_i) - F(s_{i-1})| : a = s_0 < s_1 < \cdots < s_n = x \right\}. \tag{1.14}$$

The supremum above is taken over partitions of the interval $[a,x]$. F has *bounded variation* on $[a,b]$ if $V_F(b) < \infty$. $BV[a,b]$ denotes the space of functions with bounded variation on $[a,b]$ (BV functions).

V_F is a nondecreasing function with $V_F(a) = 0$. F is a BV function iff it is the difference of two bounded nondecreasing functions, and in case F

is BV, one way to write this decomposition is

$$F = \tfrac{1}{2}(V_F + F) - \tfrac{1}{2}(V_F - F)$$

(the *Jordan decomposition* of F). If F is BV and right-continuous, then also V_F is right-continuous.

Henceforth suppose F is BV and right-continuous on $[a, b]$. Then there is a unique signed Borel measure Λ_F on $(a, b]$ determined by

$$\Lambda_F(u, v] = F(v) - F(u), \quad a \le u < v \le b.$$

We can obtain this measure from our earlier definition of Lebesgue-Stieltjes measures of nondecreasing functions. Let $F = F_1 - F_2$ be the Jordan decomposition of F. Extend these functions outside $[a, b]$ by setting $F_i(x) = F_i(a)$ for $x < a$, and $F_i(x) = F_i(b)$ for $x > b$. Then $\Lambda_F = \Lambda_{F_1} - \Lambda_{F_2}$ is the Jordan decomposition of the measure Λ_F, where Λ_{F_1} and Λ_{F_2} are as constructed in Example 1.6. Furthermore, the total variation measure of Λ_F is

$$|\Lambda_F| = \Lambda_{F_1} + \Lambda_{F_2} = \Lambda_{V_F},$$

the Lebesgue-Stieltjes measure of the total variation function V_F. The integral of a bounded Borel function g on $(a, b]$ with respect to the measure Λ_F is of course denoted by

$$\int_{(a,b]} g \, d\Lambda_F \quad \text{but also by} \quad \int_{(a,b]} g(x) \, dF(x),$$

and the integral is called a *Lebesgue-Stieltjes integral*. We shall use both of these notations in the sequel. Especially when $\int g \, dF$ might be confused with a stochastic integral, we prefer $\int g \, d\Lambda_F$. For Lebesgue-Stieltjes integrals inequality (1.12) can be written in the form

$$\left| \int_{(a,b]} g(x) \, dF(x) \right| \le \int_{(a,b]} |g(x)| \, dV_F(x). \tag{1.15}$$

We consider Λ_F a measure on $(a, b]$ rather than $[a, b]$ because according to the connection between a right-continuous function and its Lebesgue-Stieltjes measure, the measure of the singleton $\{a\}$ is

$$\Lambda_F\{a\} = F(a) - F(a-). \tag{1.16}$$

This value is determined by how we choose to extend F to $x < a$, and so is not determined by the values on $[a, b]$.

Advanced calculus courses sometimes cover a related integral called the *Riemann-Stieltjes*, or the *Stieltjes* integral. This is a generalization of the

Riemann integral. For bounded functions g and F on $[a, b]$, the Riemann-Stieltjes integral $\int_a^b g\,dF$ is defined by

$$\int_a^b g\,dF = \lim_{\text{mesh}(\pi)\to 0} \sum_i g(x_i)\big(F(s_{i+1}) - F(s_i)\big)$$

if this limit exists. The notation and the interpretation of the limit is as in (1.6). One can prove that this limit exists for example if g is continuous and F is BV [16, page 282]. The next lemma gives a version of this limit that will be used frequently in the sequel. The left limit function is defined by $f(t-) = \lim_{s\nearrow t, s<t} f(s)$, by approaching t strictly from the left, provided these limits exist.

Lemma 1.12. *Let ν be a finite signed measure on $(0, T]$. Let f be a bounded Borel function on $[0, T]$ for which the left limit $f(t-)$ exists at all $0 < t \le T$. Let $\pi^n = \{0 = s_1^n < \cdots < s_{m(n)}^n = T\}$ be partitions of $[0, T]$ such that* $\text{mesh}(\pi^n) \to 0$. *Then*

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left| \sum_{i=0}^{m(n)-1} f(s_i^n)\nu(s_i^n \wedge t, s_{i+1}^n \wedge t] - \int_{(0,t]} f(s-)\,\nu(ds) \right| = 0.$$

In particular, for a right-continuous function $G \in BV[0, T]$,

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left| \sum_{i=0}^{m(n)-1} f(s_i^n)\big(G(s_{i+1}^n \wedge t) - G(s_i^n \wedge t)\big) - \int_{(0,t]} f(s-)\,dG(s) \right| = 0.$$

It is important here that f is evaluated at the left endpoint of the partition intervals $[s_i^n, s_{i+1}^n]$.

Proof. For each $0 \le t \le T$,

$$\begin{aligned}
&\left| \sum_{i=0}^{m(n)-1} f(s_i^n)\nu(s_i^n \wedge t, s_{i+1}^n \wedge t] - \int_{(0,t]} f(s-)\,\nu(ds) \right| \\
&\qquad \le \int_{(0,t]} \left| \sum_{i=0}^{m(n)-1} f(s_i^n)\mathbf{1}_{(s_i^n, s_{i+1}^n]}(s) - f(s-) \right| |\nu|(ds) \\
&\qquad \le \int_{(0,T]} \left| \sum_{i=0}^{m(n)-1} f(s_i^n)\mathbf{1}_{(s_i^n, s_{i+1}^n]}(s) - f(s-) \right| |\nu|(ds)
\end{aligned}$$

where the last inequality is simply a consequence of increasing the interval of integration to $(0, T]$. The last integral gives a bound that is uniform in t, and it vanishes as $n \to \infty$ by the dominated convergence theorem. $\square$

Example 1.13. A basic example is a step function. Let $\{x_i\}$ be a sequence of points in $\mathbf{R}$ (ordering of x_i's is immaterial), and $\{\alpha_i\}$ an absolutely summable sequence, which means $\sum|\alpha_i| < \infty$. Define

$$G(t) = \sum_{i:x_i \le t} \alpha_i \qquad \text{for } t \in \mathbf{R}.$$

From

$$|G(t) - G(s)| \le \sum_{i:s<x_i\le t} |\alpha_i| \qquad \text{for } s < t$$

one can show that G is right-continuous and a BV function on any subinterval of $\mathbf{R}$. Furthermore, the left limit $G(t-)$ exists and is given by

$$G(t-) = \sum_{i:x_i<t} \alpha_i \qquad \text{for each } t \in \mathbf{R}.$$

The Lebesgue-Stieltjes integral of a bounded Borel function f is

$$\int_{(0,T]} f\, dG = \sum_{i:0<x_i\le T} \alpha_i f(x_i). \tag{1.17}$$

To justify this, first take $f = \mathbf{1}_{(a,b]}$ with $0 \le a < b \le T$ and check that both sides equal $G(b) - G(a)$ (left-hand side by definition of the Lebesgue-Stieltjes measure). These intervals, together with the empty set, form a π-system that generates the Borel σ-algebra on $(0, T]$. Theorem B.4 can be applied to verify the identity for all bounded Borel functions f.

The above proof of (1.17) is a good example of a recurring theme. Suppose the goal is to prove an identity for a large class of objects (for example, (1.17) above is supposed to be valid for all bounded Borel functions f). Typically we can do an explicit verification for some special cases. If this class of special cases is rich enough, then we can hope to complete the proof by appealing to some general principle that extends the identity from the special class to the entire class. Examples of such general principles are Theorems B.3 and B.4 and Lemmas B.5 and B.6 in Appendix B.

1.1.10. Radon-Nikodym theorem. This theorem is among the most important in measure theory. We state it here because it gives us the existence of conditional expectations in the next section. First a definition. Suppose μ is a measure and ν a signed measure on a measurable space $(X, \mathcal{A})$. We say ν is *absolutely continuous* with respect to μ, abbreviated $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{A}$.

Theorem 1.14. *Let μ be a σ-finite measure and ν a σ-finite signed measure on a measurable space $(X, \mathcal{A})$. Assume ν is absolutely continuous with respect to μ. Then there exists a μ-almost everywhere unique $\mathcal{A}$-measurable* *function f such that at least one of $\int f^+\,d\mu$ and $\int f^-\,d\mu$ is finite, and for each $A \in \mathcal{A}$,*

$$\nu(A) = \int_A f\,d\mu. \tag{1.18}$$

Some remarks are in order. Since either $\int_A f^+\,d\mu$ or $\int_A f^-\,d\mu$ is finite, the integral $\int_A f\,d\mu$ has a well-defined value in $[-\infty, \infty]$. The equality of integrals (1.18) extends to measurable functions, so that

$$\int g\,d\nu = \int gf\,d\mu \tag{1.19}$$

for all $\mathcal{A}$-measurable functions g for which the integrals make sense. The precise sense in which f is unique is this: if $\tilde{f}$ also satisfies (1.18) for all $A \in \mathcal{A}$, then $\mu\{f \neq \tilde{f}\} = 0$.

The function f is the *Radon-Nikodym derivative* of ν with respect to μ, and denoted by $f = d\nu/d\mu$. The derivative notation is very suggestive. It leads to $d\nu = f\,d\mu$ which tells us how to do the substitution in the integral. Also, it suggests that

$$\frac{d\nu}{d\rho} \cdot \frac{d\rho}{d\mu} = \frac{d\nu}{d\mu} \tag{1.20}$$

which is a true theorem under the right assumptions: suppose ν is a signed measure, ρ and μ positive measures, all σ-finite, $\nu \ll \rho$ and $\rho \ll \mu$. Then

$$\int g\,d\nu = \int g \cdot \frac{d\nu}{d\rho}\,d\rho = \int g \cdot \frac{d\nu}{d\rho} \cdot \frac{d\rho}{d\mu}\,d\mu$$

by two applications of (1.19). Since the Radon-Nikodym derivative is unique, the equality above proves (1.20).

Here is a result that combines the Radon-Nikodym theorem with Lebesgue-Stieltjes integrals.

Lemma 1.15. *Suppose ν is a finite signed Borel measure on $[0,T]$ and $g \in L^1(\nu)$. Let*

$$F(t) = \int_{[0,t]} g(s)\,\nu(ds), \quad 0 \le t \le T.$$

Then F is a right-continuous BV function on $[0,T]$. The Lebesgue-Stieltjes integral of a bounded Borel function ϕ on $(0,T]$ satisfies

$$\int_{(0,T]} \phi(s)\,dF(s) = \int_{(0,T]} \phi(s)g(s)\,\nu(ds). \tag{1.21}$$

In abbreviated form, $dF = g\,d\nu$ and $g = d\Lambda_F/d\nu$ on $(0,T]$.

Proof. For right continuity of F, let $t_n \searrow t$. Then

$$|F(t_n) - F(t)| = \left| \int_{[0,T]} \mathbf{1}_{(t,t_n]} \cdot g \, d\nu \right| \le \int_{[0,T]} \mathbf{1}_{(t,t_n]} |g| \, d|\nu|.$$

The last integral vanishes as $t_n \searrow t$ because $\mathbf{1}_{(t,t_n]}(s) \to 0$ at each point s, and the integral converges by dominated convergence. Thus $F(t+) = F(t)$.

For any partition $0 = s_0 < s_1 < \cdots < s_n = T$,

$$\sum_i \bigl|F(s_{i+1}) - F(s_i)\bigr| = \sum_i \left| \int_{(s_i,s_{i+1}]} g \, d\nu \right| \le \sum_i \int_{(s_i,s_{i+1}]} |g| \, d|\nu|$$
$$= \int_{(0,T]} |g| \, d|\nu|.$$

By the assumption $g \in L^1(\nu)$ the last quantity above is a finite upper bound on the sums of F-increments over all partitions. Hence $F \in BV[0,T]$.

The last issue is the equality of the two measures Λ_F and $g \, d\nu$ on $(0,T]$. By Lemma B.5 it suffices to check the equality of the two measures for intervals $(a,b]$, because this class of intervals is closed under intersection and generates the Borel σ-algebra on $(0,T]$.

$$\Lambda_F(a,b] = F(b) - F(a) = \int_{[0,b]} g(s)\,\nu(ds) - \int_{[0,a]} g(s)\,\nu(ds) = \int_{(a,b]} g(s)\,\nu(ds).$$

This suffices. ∎

The conclusion (1.21) can be extended to $[0,T]$ if we define $F(0-) = 0$. For then

$$\Lambda_F\{0\} = F(0) - F(0-) = F(0) = g(0)\nu\{0\} = \int_{\{0\}} g \, d\nu.$$

On the other hand, the conclusion of the lemma on $(0,T]$ would not change if we defined $F(0) = 0$ and

$$F(t) = \int_{(0,t]} g(s)\,\nu(ds), \quad 0 < t \le T.$$

This changes F by a constant and hence does not affect its total variation or Lebesgue-Stieltjes measure.

1.2. Basic concepts of probability theory

This section summarizes the measure-theoretic foundations of probability theory. Matters related to stochastic processes will be treated in the next chapter.

1.2.1. Probability spaces, random variables and expectations. The foundations of probability are taken directly from measure theory, with notation and terminology adapted to probabilistic conventions. A *probability space* $(\Omega, \mathcal{F}, P)$ is a measure space with total mass $P(\Omega) = 1$. The probability space is supposed to model the random experiment or collection of experiments that we wish to analyze. The underlying space Ω is called the *sample space*, and its *sample points* $\omega \in \Omega$ are the elementary outcomes of the experiment. The measurable sets in $\mathcal{F}$ are called *events*. P is a *probability measure.* A *random variable* is a measurable function $X : \Omega \to S$ with values in some measurable space S. Most often $S = \mathbf{R}$. If $S = \mathbf{R}^d$ one can call X a *random vector*, and if S is a function space then X is a *random function.*

Here are some examples to illustrate the terminology.

Example 1.16. Consider the experiment of choosing randomly a person in a room of N people and registering his or her age in years. Then naturally Ω is the set of people in the room, $\mathcal{F}$ is the collection of all subsets of Ω, and $P\{\omega\} = 1/N$ for each person $\omega \in \Omega$. Let $X(\omega)$ be the age of person ω. Then X is a $\mathbf{Z}_+$-valued measurable function (random variable) on Ω.

Example 1.17. Consider the (thought) experiment of tossing a coin infinitely many times. Let us record the outcomes (heads and tails) as zeroes and ones. The sample space Ω is the space of sequences $\omega = (x_1, x_2, x_3, \dots)$ of zeroes and ones, or $\Omega = \{0, 1\}^{\mathbf{N}}$, where $\mathbf{N} = \{1, 2, 3, \dots\}$ is the set of natural numbers. The σ-algebra $\mathcal{F}$ on Ω is the product σ-algebra $\mathcal{B}^{\otimes \mathbf{N}}$ where each factor is the natural σ-algebra

$$\mathcal{B} = \big\{\emptyset, \{0\}, \{1\}, \{0, 1\}\big\}$$

on $\{0, 1\}$. To choose the appropriate probability measure on Ω, we need to make assumptions on the coin. Simplest would be to assume that successive coin tosses are independent (a term we discuss below) and fair (heads and tails equally likely). Let $\mathcal{S}$ be the class of events of the form

$$A = \{\omega : (x_1, \dots, x_n) = (a_1, \dots, a_n)\}$$

as n varies over $\mathbf{N}$ and $(a_1, \dots, a_n)$ varies over n-tuples of zeroes and ones. Include $\emptyset$ and Ω to make $\mathcal{S}$ a semialgebra. Our assumptions dictate that the probability of the event A should be $P_0(A) = 2^{-n}$. One needs to check that P_0 satisfies the hypotheses of Theorem 1.5. Then the mathematical machinery takes over and gives the existence of a unique probability measure P on $(\Omega, \mathcal{F})$ that agrees with P_0 on $\mathcal{S}$.

This is a mathematical model of a sequence of independent fair coin tosses. Natural random variables to define on Ω are first the coordinate variables $X_i(\omega) = x_i$, and then variables derived from these such as $S_n =$

$X_1 + \cdots + X_n$, the number of ones among the first n tosses. The random variables $\{X_i\}$ are an example of an *i.i.d.* sequence, which is short for *independent and identically distributed*.

The *expectation* of a real-valued random variable X is simply its Lebesgue integral over the probability space:

$$EX = \int_\Omega X\,dP.$$

The rules governing the existence of the expectation are exactly those inherited from measure theory. The spaces $L^p(P)$ are also defined as for general measure spaces.

The *probability distribution* (or simply *distribution*, also the term *law* is used) μ of a random variable X is the probability measure obtained when the probability measure P is transported to the real line via

$$\mu(B) = P\{X \in B\}, \quad B \in \mathcal{B}_{\mathbf{R}}.$$

The expression $\{X \in B\}$ is an abbreviation for the longer set expression $\{\omega \in \Omega : X(\omega) \in B\}$.

If h is a bounded Borel function on $\mathbf{R}$, then $h(X)$ is also a random variable (this means the composition $h \circ X$), and

$$Eh(X) = \int_\Omega h(X)\,dP = \int_{\mathbf{R}} h(x)\,\mu(dx). \tag{1.22}$$

This equality is an instance of the change of variables identity (1.5). Notice that we need not even specify the probability space to make this calculation. This is the way things usually work. There must always be a probability space underlying our reasoning, but when situations are simple we can ignore it and perform our calculations in familiar spaces such as the real line or Euclidean spaces.

The *(cumulative) distribution function* F of a random variable X is defined by $F(x) = P\{X \le x\}$. The distribution μ is the Lebesgue-Stieltjes measure of F. Using the notation of Lebesgue-Stieltjes integrals, (1.22) can be expressed as

$$Eh(X) = \int_{\mathbf{R}} h(x)\,dF(x). \tag{1.23}$$

This is the way expectations are expressed in probability and statistics books that avoid using measure theory, relying on the advanced calculus level understanding of the Stieltjes integral.

The *density function* f of a random variable X is the Radon-Nikodym derivative of its distribution with respect to Lebesgue measure, so $f = d\mu/dx$. It exists iff μ is absolutely continuous with respect to Lebesgue measure on

$\mathbf{R}$. When f exists, the distribution function F is differentiable Lebesgue–almost everywhere, and $F' = f$ Lebesgue–almost everywhere. The expectation can then be expressed as an integral with respect to Lebesgue measure:

$$Eh(X) = \int_{\mathbf{R}} h(x) f(x)\, dx. \tag{1.24}$$

This is the way most expectations are evaluated in practice. For example, if X is a rate λ exponential random variable ($X \sim \text{Exp}(\lambda)$ in symbols), then

$$Eh(X) = \int_0^\infty h(x) \lambda e^{-\lambda x}\, dx.$$

The concepts discussed above have natural extensions to $\mathbf{R}^d$-valued random vectors.

Example 1.18. Here are the most important probability densities.

(i) On any bounded interval $[a, b]$ of $\mathbf{R}$ there is the *uniform distribution* Unif$[a, b]$ with density $f(x) = (b-a)^{-1}\mathbf{1}_{[a,b]}(x)$. Whether the endpoints are included is immaterial because it does not affect the outcome of any calculation.

(ii) The exponential distribution mentioned above is a special case of the Gamma(α, λ) distribution on $\mathbf{R}_+$ with density $f(x) = \Gamma(\alpha)^{-1}(\lambda x)^{\alpha-1}\lambda e^{-\lambda x}$. The two parameters satisfy $\alpha, \lambda > 0$. The gamma function is $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$.

(iii) For $\alpha, \beta > 0$ the Beta(α, β) distribution on $(0, 1)$ has density $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$.

(iv) For a vector $\mathbf{v} \in \mathbf{R}^d$ ($d \geq 1$) and a symmetric, nonnegative definite, nonsingular $d \times d$ matrix Γ, the normal (or Gaussian) distribution $\mathcal{N}(\mathbf{v}, \Gamma)$ on $\mathbf{R}^d$ has density

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Gamma}} \exp\bigl(-\tfrac{1}{2}(\mathbf{x}-\mathbf{v})^T \Gamma^{-1}(\mathbf{x}-\mathbf{v})\bigr). \tag{1.25}$$

Above, and in general, we regard a d-vector $\mathbf{v}$ as a $d \times 1$ matrix with $1 \times d$ transpose $\mathbf{v}^T$.

If Γ is singular then the $\mathcal{N}(\mathbf{v}, \Gamma)$ distribution can be characterized by its characteristic function (the probabilistic term for Fourier transform)

$$E(e^{i\mathbf{s}^T X}) = \exp\bigl(i\mathbf{s}^T\mathbf{v} - \tfrac{1}{2}\mathbf{s}^T\Gamma\mathbf{s}\bigr), \quad \mathbf{s} \in \mathbf{R}^d, \tag{1.26}$$

where X represents the $\mathbf{R}^d$-valued $\mathcal{N}(\mathbf{v}, \Gamma)$-distributed random vector and i is the imaginary unit $\sqrt{-1}$. This probability distribution is supported on the image of $\mathbf{R}^d$ under Γ.

The $\mathcal{N}(0, 1)$ distribution on $\mathbf{R}$ with density $f(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is the *standard normal* distribution. When $d > 1$, $\mathcal{N}(\mathbf{v}, \Gamma)$ is called a *multivariate* *normal.* As is evident from the formulas, the distribution of a Gaussian vector X is determined by its mean vector $\mathbf{v} = EX$ and covariance matrix $\Gamma = E(XX^T)$.

One more terminological change as we switch from analysis to probability: *almost everywhere* (*a.e.*) becomes *almost surely* (*a.s.*). But of course there is no harm in using both.

Equality of random variables X and Y has the same meaning as equality for any functions: if X and Y are defined on the same sample space Ω then they are equal as functions if $X(\omega) = Y(\omega)$ for each $\omega \in \Omega$. Often we cannot really control what happens on null sets (sets of probability zero), so the more relevant notion is the almost sure equality: $X = Y$ a.s. if $P(X = Y) = 1$. We also talk about *equality in distribution* of X and Y which means that $P(X \in B) = P(Y \in B)$ for all measurable sets B in the (common) range space of X and Y. This is abbreviated $X \overset{d}{=} Y$ and makes sense even if X and Y are defined on different probability spaces.

1.2.2. Convergence of random variables.

Here is a list of ways in which random variables can converge. Except for convergence in distribution, they are direct adaptations of the corresponding modes of convergence from analysis.

Definition 1.19. Let $\{X_n\}$ be a sequence of random variables and X a random variable, all real-valued.

(a) $X_n \to X$ *almost surely* if

$$P\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\} = 1.$$

(b) $X_n \to X$ *in probability* if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\Big\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\Big\} = 0.$$

(c) $X_n \to X$ *in* L^p for $1 \leq p < \infty$ if

$$\lim_{n\to\infty} E\big\{|X_n(\omega) - X(\omega)|^p\big\} = 0.$$

(d) $X_n \to X$ *in distribution* (also called *weakly*) if

$$\lim_{n\to\infty} P\{X_n \leq x\} = P\{X \leq x\}$$

for each x at which $F(x) = P\{X \leq x\}$ is continuous.

Convergence types (a)–(c) require that all the random variables are defined on the same probability space, but (d) does not.

The definition of weak convergence above is a specialization to the real line of the general definition, which is this: let $\{\mu_n\}$ and μ be Borel probability measures on a metric space S. Then $\mu_n \to \mu$ weakly if

$$\int_S g\, d\mu_n \longrightarrow \int_S g\, d\mu$$

for all bounded continuous functions g on S. Random variables converge weakly if their distributions do in the sense above. Commonly used notation for weak convergence is $X_n \Rightarrow X$.

Here is a summary of the relationships between the different types of convergence. We need one new definition: a sequence $\{X_n\}$ of random variables is *uniformly integrable* if

$$\lim_{M\to\infty} \sup_{n\in\mathbf{N}} E\big[\,|X_n| \cdot \mathbf{1}_{\{|X_n|\ge M\}}\big] = 0. \tag{1.27}$$

Theorem 1.20. *Let $\{X_n\}$ and X be real-valued random variables on a common probability space.*

(i) *If $X_n \to X$ almost surely or in L^p for some $1 \le p < \infty$, then $X_n \to X$ in probability.*

(ii) *If $X_n \to X$ in probability, then $X_n \to X$ weakly.*

(iii) *If $X_n \to X$ in probability, then there exists a subsequence X_{n_k} such that $X_{n_k} \to X$ almost surely.*

(iv) *Suppose $X_n \to X$ in probability. Then $X_n \to X$ in L^1 iff $\{X_n\}$ is uniformly integrable.*

1.2.3. Independence and conditioning. Fix a probability space $(\Omega, \mathcal{F}, P)$. In probability theory, σ-algebras represent information. $\mathcal{F}$ represents all the information about the experiment, and sub-σ-algebras $\mathcal{A}$ of $\mathcal{F}$ represent partial information. "Knowing σ-algebra $\mathcal{A}$" means knowing for each event $A \in \mathcal{A}$ whether A happened or not. A common way to create sub-σ-algebras is to generate them with random variables. If X is a random variable on Ω, then the σ-algebra generated by X is denoted by $\sigma(X)$ and it is given by the collection of inverse images of Borel sets: $\sigma(X) = \big\{\{X \in B\} : B \in \mathcal{B}_{\mathbf{R}}\big\}$. Measurability of X is exactly the same as $\sigma(X) \subseteq \mathcal{F}$.

Knowing the actual value of X is the same as knowing whether $\{X \in B\}$ happened for each $B \in \mathcal{B}_{\mathbf{R}}$. But of course there may be many sample points ω that have the same values for X, so knowing X does not allow us to determine which outcome ω actually happened. In this sense $\sigma\{X\}$ represents partial information. Here is an elementary example.

Example 1.21. Suppose we flip a coin twice. The sample space is $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$ and the generic sample point $\omega = (\omega_1, \omega_2) \in$

$\{0,1\}^2$. Let $X(\omega) = \omega_1 + \omega_2$. Then $\sigma\{X\}$ consists of $\emptyset$, $\{(0,0)\}$, $\{(0,1),(1,0)\}$, $\{(1,1)\}$ and all their unions. But knowing $\sigma\{X\}$ cannot distinguish whether outcome $(0,1)$ or $(1,0)$ happened. In English: if you tell me only that one flip came out heads, I don't know if it was the first or the second flip.

Elementary probability courses define that two events A and B are independent if $P(A \cap B) = P(A)P(B)$. The conditional probability of A, given B, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.28}$$

provided $P(B) > 0$. Thus the independence of A and B can be equivalently expressed as $P(A|B) = P(A)$. This reveals the meaning of independence: knowing that B happened (in other words, conditioning on B) does not change our probability for A.

One technical reason we need to go beyond these elementary definitions is that we need to routinely condition on events of probability zero. For example, suppose X and Y are independent random variables, both uniformly distributed on $[0,1]$, and we set $Z = X + Y$. Then we would all agree that $P(Z \ge \frac{1}{2} | Y = \frac{1}{3}) = \frac{5}{6}$, yet since $P(Y = \frac{1}{3}) = 0$ this conditional probability cannot be defined in the above manner.

The general definition of independence, from which various other definitions follow as special cases, is for the independence of σ-algebras.

Definition 1.22. Let $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ be sub-σ-algebras of $\mathcal{F}$. Then $\mathcal{A}_1$, $\mathcal{A}_2$, $\dots$, $\mathcal{A}_n$ are *mutually independent* (or simply *independent*) if, for every choice of events $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, $\dots$, $A_n \in \mathcal{A}_n$,

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n). \tag{1.29}$$

An arbitrary collection $\{\mathcal{A}_i : i \in \mathcal{I}\}$ of sub-σ-algebras of $\mathcal{F}$ is independent if each finite subcollection is independent.

The more concrete notions of independence of random variables and independence of events derive from the above definition.

Definition 1.23. A collection of random variables $\{X_i : i \in \mathcal{I}\}$ on a probability space $(\Omega, \mathcal{F}, P)$ is *independent* if the σ-algebras $\{\sigma(X_i) : i \in \mathcal{I}\}$ generated by the individual random variables are independent. Equivalently, for any finite set of distinct indices $i_1, i_2, \dots, i_n$ and any measurable sets $B_1, B_2, \dots, B_n$ from the range spaces of the random variables, we have

$$P\{X_{i_1} \in B_1, X_{i_2} \in B_2, \dots, X_{i_n} \in B_n\} = \prod_{k=1}^{n} P\{X_{i_k} \in B_k\}. \tag{1.30}$$

Finally, events $\{A_i : i \in \mathcal{I}\}$ are independent if the corresponding indicator random variables $\{\mathbf{1}_{A_i} : i \in \mathcal{I}\}$ are independent.

Some remarks about the definitions. The product property extends to all expectations that are well-defined. If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are independent σ-algebras, and $Z_1, \ldots, Z_n$ are integrable random variables such that Z_i is $\mathcal{A}_i$-measurable ($1 \le i \le n$) and the product $Z_1 Z_2 \cdots Z_n$ is integrable, then

$$E\big[Z_1 Z_2 \cdots Z_n\big] = EZ_1 \cdot EZ_2 \cdots EZ_n. \tag{1.31}$$

If the variables Z_i are $[0, \infty]$-valued then this identity holds regardless of integrability because the expectations are limits of expectations of truncated random variables.

Independence is closely tied with the notion of product measure. Let μ be the distribution of the random vector $X = (X_1, X_2, \ldots, X_n)$ on $\mathbf{R}^n$, and let μ_i be the distribution of component X_i on $\mathbf{R}$. Then the variables $X_1, X_2, \ldots, X_n$ are independent iff $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$.

Further specialization yields properties familiar from elementary probability. For example, if the random vector (X, Y) has a density $f(x, y)$ on $\mathbf{R}^2$, then X and Y are independent iff $f(x, y) = f_X(x) f_Y(y)$ where f_X and f_Y are the *marginal densities* of X and Y. Also, it is enough to check properties (1.29) and (1.30) for classes of sets that are closed under intersections and generate the σ-algebras in question. (A consequence of the so-called π-λ theorem, see Lemma B.5 in the Appendix.) Hence we get the familiar criterion for independence in terms of cumulative distribution functions:

$$P\{X_{i_1} \le t_1, X_{i_2} \le t_2, \ldots, X_{i_n} \le t_n\} = \prod_{k=1}^{n} P\{X_{i_k} \le t_k\}. \tag{1.32}$$

Independence is a special property, and always useful when it is present. The key tool for handling *dependence* (that is, lack of independence) is the notion of conditional expectation. It is a nontrivial concept, but fundamental to just about everything that follows in this book.

Definition 1.24. Let $X \in L^1(P)$ and let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$. The *conditional expectation of X, given $\mathcal{A}$*, is the integrable, $\mathcal{A}$-measurable random variable Y that satisfies

$$\int_A X\,dP = \int_A Y\,dP \quad \text{for all } A \in \mathcal{A}. \tag{1.33}$$

The notation for the conditional expectation is $Y(\omega) = E(X|\mathcal{A})(\omega)$. It is almost surely unique, in other words, if $\widetilde{Y}$ is $\mathcal{A}$-measurable and satisfies (1.33), then $P\{Y = \widetilde{Y}\} = 1$.

Justification of the definition. The existence of the conditional expectation follows from the Radon-Nikodyn theorem. Define a finite signed measure ν on $(\Omega, \mathcal{A})$ by

$$\nu(A) = \int_A X\,dP, \quad A \in \mathcal{A}.$$

$P(A) = 0$ implies $\nu(A) = 0$, and so $\nu \ll P$. By the Radon-Nikodym theorem there exists a Radon-Nikodym derivative $Y = d\nu/dP$ which is $\mathcal{A}$-measurable and satisfies

$$\int_A Y\,dP = \nu(A) = \int_A X\,dP \quad \text{for all } A \in \mathcal{A}.$$

Y is integrable because

$$\begin{aligned}\int_\Omega Y^+\,dP = \int_{\{Y\ge 0\}} Y\,dP = \nu\{Y \ge 0\} = \int_{\{Y\ge 0\}} X\,dP\\ \le \int_\Omega |X|\,dP < \infty\end{aligned}$$

and a similar bound can be given for $\int_\Omega Y^-\,dP$.

To prove uniqueness, suppose $\widetilde{Y}$ satisfies the same properties as Y. Let $A = \{Y \ge \widetilde{Y}\}$. This is an $\mathcal{A}$-measurable event. On A, $Y - \widetilde{Y} = (Y - \widetilde{Y})^+$, while on A^c, $(Y - \widetilde{Y})^+ = 0$. Consequently

$$\begin{aligned}\int (Y - \widetilde{Y})^+\,dP = \int_A (Y - \widetilde{Y})\,dP = \int_A Y\,dP - \int_A \widetilde{Y}\,dP\\ = \int_A X\,dP - \int_A X\,dP = 0.\end{aligned}$$

The integral of a nonnegative function vanishes iff the function vanishes almost everywhere. Thus $(Y - \widetilde{Y})^+ = 0$ almost surely. A similar argument shows $(Y - \widetilde{Y})^- = 0$, and so $|Y - \widetilde{Y}| = 0$ almost surely. □

The defining property (1.33) of the conditional expectation $E(X|\mathcal{A})$ extends to

$$\int_\Omega ZX\,dP = \int_\Omega Z\,E(X|\mathcal{A})\,dP \tag{1.34}$$

for any bounded $\mathcal{A}$-measurable random variable Z. Boundedness of Z guarantees that ZX and $Z\,E(X|\mathcal{A})$ are integrable for an integrable random variable X.

Some notational conventions. When $X = \mathbf{1}_B$ is the indicator random variable of an event B, we can write $P(B|\mathcal{A})$ for $E(\mathbf{1}_B|\mathcal{A})$. When the conditioning σ-algebra is generated by a random variable Y, so $\mathcal{A} = \sigma\{Y\}$, we can write $E(X|Y)$ instead of $E(X|\sigma\{Y\})$.

Sometimes one also sees the conditional expectation $E(X|Y = y)$, regarded as a function of $y \in \mathbf{R}$ (assuming now that Y is real-valued). This is defined by an additional step. Since $E(X|Y)$ is $\sigma\{Y\}$-measurable, there exists a Borel function h such that $E(X|Y) = h(Y)$. This is an instance of a general exercise according to which every $\sigma\{Y\}$-measurable random variable is a Borel function of Y. Then one uses h to define $E(X|Y = y) = h(y)$. This conditional expectation works with integrals on the real line with respect to the distribution μ_Y of Y: for any $B \in \mathcal{B}_{\mathbf{R}}$,

$$E[\mathbf{1}_B(Y)X] = \int_B E(X|Y = y)\,\mu_Y(dy). \tag{1.35}$$

The definition of the conditional expectation is abstract, and it takes practice to get used to the idea of conditional probabilities and expectations as random variables rather than as numbers. The task is to familiarize oneself with this concept by working with it. Eventually one will understand how it actually does everything we need. The typical way to find conditional expectations is to make an educated guess, based on an intuitive understanding of the situation, and then verify the definition. The $\mathcal{A}$-measurability is usually built into the guess, so what needs to be checked is (1.33). Whatever its manifestation, conditional expectation always involves averaging over some portion of the sample space. This is especially clear in this simplest of examples.

Example 1.25. Let A be an event such that $0 < P(A) < 1$, and $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$. Then

$$E(X|\mathcal{A})(\omega) = \frac{E(\mathbf{1}_A X)}{P(A)} \cdot \mathbf{1}_A(\omega) + \frac{E(\mathbf{1}_{A^c} X)}{P(A^c)} \cdot \mathbf{1}_{A^c}(\omega). \tag{1.36}$$

Let us check (1.33) for A. Let Y denote the right-hand side of (1.36). Then

$$\begin{aligned}\int_A Y\,dP &= \int_A \frac{E(\mathbf{1}_A X)}{P(A)}\mathbf{1}_A(\omega)\,P(d\omega) = \frac{E(\mathbf{1}_A X)}{P(A)} \int_A \mathbf{1}_A(\omega)\,P(d\omega)\\ &= E(\mathbf{1}_A X) = \int_A X\,dP.\end{aligned}$$

A similar calculation checks $\int_{A^c} Y\,dP = \int_{A^c} X\,dP$, and adding these together gives the integral over Ω. $\emptyset$ is of course trivial, since any integral over $\emptyset$ equals zero. See Exercise 1.15 for a generalization of this.

Here is a concrete case. Let $X \sim \text{Exp}(\lambda)$, and suppose we are allowed to know whether $X \le c$ or $X > c$. What is our updated expectation for X? To model this, let us take $(\Omega, \mathcal{F}, P) = (\mathbf{R}_+, \mathcal{B}_{\mathbf{R}_+}, \mu)$ with $\mu(dx) = \lambda e^{-\lambda x} dx$, the identity random variable $X(\omega) = \omega$, and the sub-σ-algebra

$\mathcal{A} = \{\Omega, \emptyset, [0,c], (c,\infty)\}$. Following (1.36), for $\omega \in (c,\infty)$ we compute

$$E(X|\mathcal{A})(\omega) = \frac{E(\mathbf{1}_{\{X>c\}}X)}{P(X>c)} = \frac{1}{\mu(c,\infty)} \int_c^\infty x\lambda e^{-\lambda x}\,dx$$
$$= e^{\lambda c}(ce^{-\lambda c} + \lambda^{-1}e^{-\lambda c}) = c + \lambda^{-1}.$$

You probably knew the answer from the memoryless property of the exponential distribution? If not, see Exercise 1.13. Exercise 1.16 asks you to complete this example.

The next theorem lists the main properties of the conditional expectation. Equalities and inequalities for conditional expectations are almost sure statements because the conditional expectation itself is defined only up to null sets. So each statement below except (i) comes with an "a.s." modifier.

Theorem 1.26. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X and Y integrable random variables on Ω, and $\mathcal{A}$ and $\mathcal{B}$ sub-σ-fields of $\mathcal{F}$.*

(i) $E[E(X|\mathcal{A})] = EX$.

(ii) $E[\alpha X + \beta Y|\mathcal{A}] = \alpha E[X|\mathcal{A}] + \beta E[Y|\mathcal{A}]$ *for* $\alpha, \beta \in \mathbf{R}$.

(iii) *If* $X \ge Y$ *then* $E[X|\mathcal{A}] \ge E[Y|\mathcal{A}]$.

(iv) *If X is $\mathcal{A}$-measurable, then* $E[X|\mathcal{A}] = X$.

(v) *If X is $\mathcal{A}$-measurable and XY is integrable, then*

$$E[XY|\mathcal{A}] = X \cdot E[Y|\mathcal{A}]. \tag{1.37}$$

(vi) *If X is independent of $\mathcal{A}$ (which means that the σ-algebras $\sigma\{X\}$ and $\mathcal{A}$ are independent), then* $E[X|\mathcal{A}] = EX$.

(vii) *If* $\mathcal{A} \subseteq \mathcal{B}$, *then*

$$E\{E(X|\mathcal{A})\,|\mathcal{B}\} = E\{E(X|\mathcal{B})\,|\mathcal{A}\} = E[X|\mathcal{A}].$$

(viii) *If* $\mathcal{A} \subseteq \mathcal{B}$ *and $E(X|\mathcal{B})$ is $\mathcal{A}$-measurable, then* $E(X|\mathcal{B}) = E(X|\mathcal{A})$.

(ix) *(Jensen's inequality) Suppose f is a convex function on (a,b), $-\infty \le a < b \le \infty$. This means that*

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y) \quad \text{for } x, y \in (a,b) \text{ and } 0 < \theta < 1.$$

Assume $P\{a < X < b\} = 1$. *Then*

$$f\big(E[X|\mathcal{A}]\big) \le E\big[f(X) \bigm| \mathcal{A}\big] \tag{1.38}$$

provided the conditional expectations are well defined.

(x) *Suppose X is a random variable with values in a measurable space $(S_1, \mathcal{H}_1)$, Y is a random variable with values in a measurable space $(S_2, \mathcal{H}_2)$,* *and $\phi : S_1 \times S_2 \to \mathbf{R}$ is a measurable function such that $\phi(X, Y)$ is integrable. Assume that X is $\mathcal{A}$-measurable while Y is independent of $\mathcal{A}$. Then*

$$E[\phi(X,Y)|\mathcal{A}](\omega) = \int_\Omega \phi(X(\omega), Y(\widetilde{\omega}))\, P(d\widetilde{\omega}). \tag{1.39}$$

Proof. The proofs must appeal to the definition of the conditional expectation. We leave them mostly as exercises or to be looked up in any graduate probability textbook. Let us prove (v) and (vii) as examples.

Proof of part (v). We need to check that $X \cdot E[Y|\mathcal{A}]$ satisfies the definition of $E[XY|\mathcal{A}]$. The $\mathcal{A}$-measurability of $X \cdot E[Y|\mathcal{A}]$ is true because X is $\mathcal{A}$-measurable by assumption, $E[Y|\mathcal{A}]$ is $\mathcal{A}$-measurable by definition, and multiplication preserves $\mathcal{A}$-measurability. Then we need to check that

$$E\big(\mathbf{1}_A XY\big) = E\big(\mathbf{1}_A\, X\, E[Y|\mathcal{A}]\big) \tag{1.40}$$

for an arbitrary $A \in \mathcal{A}$. If X were bounded, this would be a special case of (1.34) with Z replaced by $\mathbf{1}_A X$. For the general case we need to check the integrability of $X\, E[Y|\mathcal{A}]$ before we can honestly write down the right-hand side of (1.40).

Let us assume first that both X and Y are nonnegative. Then also $E(Y|\mathcal{A}) \geq 0$ by (iii), because $E(0|\mathcal{A}) = 0$ by (iv). Let $X^{(k)} = X \wedge k$ be a truncation of X. We can apply (1.34) to get

$$E\big[\mathbf{1}_A X^{(k)} Y\big] = E\big[\mathbf{1}_A X^{(k)} E(Y|\mathcal{A})\big]. \tag{1.41}$$

Inside both expectations we have nonnegative random variables that increase with k. By the monotone convergence theorem we can let $k \to \infty$ and recover (1.40) in the limit, for nonnegative X and Y. In particular, this tells us that, at least if $X, Y \geq 0$, the integrability of X, Y and XY imply that $X\, E(Y|\mathcal{A})$ is integrable.

Now decompose $X = X^+ - X^-$ and $Y = Y^+ - Y^-$. By property (ii),

$$E(Y|\mathcal{A}) = E(Y^+|\mathcal{A}) - E(Y^-|\mathcal{A}).$$

The left-hand side of (1.40) becomes

$$E\big[\mathbf{1}_A X^+ Y^+\big] - E\big[\mathbf{1}_A X^- Y^+\big] - E\big[\mathbf{1}_A X^+ Y^-\big] + E\big[\mathbf{1}_A X^- Y^-\big].$$

The integrability assumption is true for all pairs $X^\pm Y^\pm$ and $X^\pm Y^\mp$, so to each term above we can apply the case of (1.40) already proved for nonnegative random variables. The expression becomes

$$E\big[\mathbf{1}_A X^+ E(Y^+|\mathcal{A})\big] - E\big[\mathbf{1}_A X^- E(Y^+|\mathcal{A})\big] - E\big[\mathbf{1}_A X^+ E(Y^-|\mathcal{A})\big] + E\big[\mathbf{1}_A X^- E(Y^-|\mathcal{A})\big].$$

For integrable random variables, a sum of expectations can be combined into an expectation of a sum. Consequently the sum above becomes the right-hand side of (1.40). This completes the proof of part (v).

Proof of part (vii). That $E\{E(X|\mathcal{A})\,|\mathcal{B}\} = E[X|\mathcal{A}]$ follows from part (iv). To prove $E\{E(X|\mathcal{B})\,|\mathcal{A}\} = E[X|\mathcal{A}]$, we show that $E[X|\mathcal{A}]$ satisfies the definition of $E\{E(X|\mathcal{B})\,|\mathcal{A}\}$. Again the measurability is not a problem. Then we need to check that for any $A \in \mathcal{A}$,

$$\int_A E(X|\mathcal{B})\,dP = \int_A E(X|\mathcal{A})\,dP.$$

This is true because A lies in both $\mathcal{A}$ and $\mathcal{B}$, so both sides equal $\int_A X\,dP$. □

There is a geometric way of looking at $E(X|\mathcal{A})$ as the solution to an optimization or estimation problem. Assume that $X \in L^2(P)$. Then what is the best $\mathcal{A}$-measurable estimate of X in the mean-square sense? In other words, find the $\mathcal{A}$-measurable random variable $Z \in L^2(P)$ that minimizes $E[(X-Z)^2]$. This is a "geometric view" of $E(X|\mathcal{A})$ because it involves projecting X orthogonally to the subspace $L^2(\Omega, \mathcal{A}, P)$ of $\mathcal{A}$-measurable L^2-random variables.

Proposition 1.27. *Let $X \in L^2(P)$. Then $E(X|\mathcal{A}) \in L^2(\Omega, \mathcal{A}, P)$. For all $Z \in L^2(\Omega, \mathcal{A}, P)$,*

$$E\big[(X - E[X|\mathcal{A}])^2\big] \le E\big[(X-Z)^2\big]$$

with equality iff $Z = E(X|\mathcal{A})$.

Proof. By Jensen's inequality,

$$E\big\{E[X|\mathcal{A}]^2\big\} \le E\big\{E[X^2|\mathcal{A}]\big\} = E\big\{X^2\big\}.$$

Consequently $E[X|\mathcal{A}]$ is in $L^2(P)$.

$$\begin{aligned} E\big[(X-Z)^2\big] &= E\big[(X - E[X|\mathcal{A}] + E[X|\mathcal{A}] - Z)^2\big] = E\big[(X - E[X|\mathcal{A}])^2\big] \\ &\quad + 2E\big[(X - E[X|\mathcal{A}])(E[X|\mathcal{A}] - Z)\big] + E\big[(E[X|\mathcal{A}] - Z)^2\big] \\ &= E\big[(X - E[X|\mathcal{A}])^2\big] + E\big[(E[X|\mathcal{A}] - Z)^2\big]. \end{aligned}$$

The cross term of the square vanishes because $E[X|\mathcal{A}] - Z$ is $\mathcal{A}$-measurable, and this justifies the last equality. The last line is minimized by the unique choice $Z = E[X|\mathcal{A}]$. □

1.2.4. Construction of probability spaces. In addition to the usual construction issues of measures that we discussed before, in probability theory we need to construct stochastic processes which are infinite collections $\{X_t : t \in \mathcal{I}\}$ of random variables. Often the naturally available ingredients for the construction are the finite-dimensional distributions of the process.

These are the joint distributions of finite vectors $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ of random variables. Kolmogorov's Extension Theorem, whose proof is based on the general machinery for extending measures, states that the finite-dimensional distributions are all we need. The natural home for the construction is a product space.

To formulate the hypotheses below, we need to consider permutations acting on n-tuples of indices from an index set $\mathcal{I}$ and on n-vectors from a product space. A *permutation* π is a bijective map of $\{1, 2, \dots, n\}$ onto itself, for some finite n. If $\mathbf{s} = (s_1, s_2, \dots, s_n)$ and $\mathbf{t} = (t_1, t_2, \dots, t_n)$ are n-tuples, then $\mathbf{t} = \pi\mathbf{s}$ means that $(t_1, t_2, \dots, t_n) = (s_{\pi(1)}, s_{\pi(2)}, \dots, s_{\pi(n)})$. The action of π on any n-vector is defined similarly, by permuting the coordinates.

Here is the setting for the theorem. $\mathcal{I}$ is an arbitrary index set, and for each $t \in \mathcal{I}$, $(\Omega_t, \mathcal{B}_t)$ is a complete, separable metric space together with its Borel σ-algebra. Let $\Omega = \prod \Omega_t$ be the product space and $\mathcal{B} = \bigotimes \mathcal{B}_t$ the product σ-algebra. A generic element of Ω is written $\omega = (\omega_t)_{t\in\mathcal{I}}$.

Theorem 1.28 (Kolmogorov's extension theorem)**.** *Suppose that for each ordered n-tuple $\mathbf{t} = (t_1, t_2, \dots, t_n)$ of distinct indices we are given a probability measure $Q_{\mathbf{t}}$ on the product space $(\Omega^{\mathbf{t}}, \mathcal{B}^{\mathbf{t}}) = \left(\prod_{k=1}^n \Omega_{t_k}, \bigotimes_{k=1}^n \mathcal{B}_{t_k}\right)$, for all $n \ge 1$. We assume two properties that make $\{Q_{\mathbf{t}}\}$ a consistent family of finite-dimensional distributions:*

- (i) *If $\mathbf{t} = \pi\mathbf{s}$, then $Q_{\mathbf{t}} = Q_{\mathbf{s}} \circ \pi^{-1}$.*
- (ii) *If $\mathbf{t} = (t_1, t_2, \dots, t_{n-1}, t_n)$ and $\mathbf{s} = (t_1, t_2, \dots, t_{n-1})$, then for $A \in \mathcal{B}^{\mathbf{s}}$, $Q_{\mathbf{s}}(A) = Q_{\mathbf{t}}(A \times \Omega_{t_n})$.*

Then there exists a probability measure P on $(\Omega, \mathcal{B})$ whose finite-dimensional marginal distributions are given by $\{Q_{\mathbf{t}}\}$. In other words, for any $\mathbf{t} = (t_1, t_2, \dots, t_n)$ and any $B \in \mathcal{B}^{\mathbf{t}}$,

$$P\{\omega \in \Omega : (\omega_{t_1}, \omega_{t_2}, \dots, \omega_{t_n}) \in B\} = Q_{\mathbf{t}}(B). \tag{1.42}$$

We refer the reader to [4, Chapter 12] for a proof of Kolmogorov's theorem in this generality. [5] gives a proof for the case where $\mathcal{I}$ is countable and $\Omega_t = \mathbf{R}$ for each t. The main idea of the proof is no different for the more abstract result.

To construct a stochastic process $(X_t)_{t\in\mathcal{I}}$ with prescribed finite-dimensional marginals

$$Q_{\mathbf{t}}(A) = P\{(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in A\},$$

apply Kolmogorov's theorem to these inputs, take $(\Omega, \mathcal{B}, P)$ as the probability space, and for $\omega = (\omega_t)_{t\in\mathcal{I}} \in \Omega$ define the coordinate random variables $X_t(\omega) = \omega_t$. When $\mathcal{I}$ is countable, such as $\mathbf{Z}_+$, this strategy is perfectly adequate, and gives for example all infinite sequences of i.i.d. random variables. When $\mathcal{I}$ is uncountable, such as $\mathbf{R}_+$, we typically want something more from the construction than merely the correct distributions, namely some regularity for the random function $t \mapsto X_t(\omega)$. This is called *path regularity.* This issue is addressed in the next chapter.

We will not discuss the proof of Kolmogorov's theorem. Let us observe that hypotheses (i) and (ii) are necessary for the existence of P, so nothing unnecessary is assumed in the theorem. Property (ii) is immediate from (1.42) because $(\omega_{t_1}, \omega_{t_2}, \dots, \omega_{t_n-1}) \in A$ iff $(\omega_{t_1}, \omega_{t_2}, \dots, \omega_{t_n}) \in A \times \Omega_{t_n}$. Property (i) is also clear on intuitive grounds because all it says is that if the coordinates are permuted, their distribution gets permuted too. Here is a rigorous justification. Take a bounded measurable function f on $\Omega^{\mathbf{t}}$. Note that $f \circ \pi$ is then a function on $\Omega^{\mathbf{s}}$, because

$$\begin{aligned} \omega = (\omega_1, \dots, \omega_n) \in \Omega^{\mathbf{s}} &\iff \omega_i \in \Omega_{s_i} \ (1 \le i \le n) \\ &\iff \omega_{\pi(i)} \in \Omega_{t_i} \ (1 \le i \le n) \\ &\iff \pi\omega = (\omega_{\pi(1)}, \dots, \omega_{\pi(n)}) \in \Omega^{\mathbf{t}}. \end{aligned}$$

Compute as follows, assuming P exists:

$$\begin{aligned} \int_{\Omega^{\mathbf{t}}} f \, dQ_{\mathbf{t}} &= \int_{\Omega} f(\omega_{t_1}, \omega_{t_2}, \dots, \omega_{t_n}) \, P(d\omega) \\ &= \int_{\Omega} f(\omega_{s_{\pi(1)}}, \omega_{s_{\pi(2)}}, \dots, \omega_{s_{\pi(n)}}) \, P(d\omega) \\ &= \int_{\Omega} (f \circ \pi)(\omega_{s_1}, \omega_{s_2}, \dots, \omega_{s_n}) \, P(d\omega) \\ &= \int_{\Omega^{\mathbf{s}}} (f \circ \pi) \, dQ_{\mathbf{s}} = \int_{\Omega^{\mathbf{t}}} f \, d(Q_{\mathbf{s}} \circ \pi^{-1}). \end{aligned}$$

Since f is arbitrary, this says that $Q_{\mathbf{t}} = Q_{\mathbf{s}} \circ \pi^{-1}$.

Exercises

Exercise 1.1. (a) Suppose F is continuous on $[0, T]$, has a continuous derivative $F' = f$ on $(0, T)$, and $\int_0^T |f(s)| \, ds < \infty$ (equivalently, $f \in L^1(0, T)$). What formula does Lemma 1.15 give for $\int_{(0,T]} \phi(s) \, dF(s)$?

(b) Let F be as in part (a), $t_0 \in (0, T]$, $c \in \mathbf{R}$ and define

$$G(t) = \begin{cases} F(t) & t \in [0, t_0) \\ F(t) + c & t \in [t_0, T]. \end{cases}$$

Give the formula for $\int_{(0,T]} \phi(s) \, dG(s)$ in terms of ϕ, f, t_0 and c.

Exercise 1.2. In case you wish to check your understanding of the Lebesgue-Stieltjes integral, compute $\int_{(0,3a]} x\,dF(x)$ with $a > 0$ and

$$F(x) = \begin{cases} \pi, & 0 \le x < a \\ 4 + a - x, & a \le x < 2a \\ (x - 2a)^2, & x \ge 2a. \end{cases}$$

You should get $8a^3/3 + a^2/2 - (4 + \pi)a$.

Exercise 1.3. Let V be a continuous, nondecreasing function on $\mathbf{R}$ and Λ its Lebesgue-Stieltjes measure. Say t is a *point of strict increase* for V if $V(s) < V(t) < V(u)$ for all $s < t$ and all $u > t$. Let I be the set of such points. Show that I is a Borel set and $\Lambda(I^c) = 0$.

Hint. For rationals q let $a(q) = \inf\{s \in [q - 1, q] : V(s) = V(q)\}$ and $b(q) = \sup\{s \in [q, q+1] : V(s) = V(q)\}$. Considering q such that $a(q) < b(q)$ show that I^c is a countable union of closed intervals with zero Λ-measure. The restriction of $a(q)$ and $b(q)$ to $[q - 1, q + 1]$ is there only to guarantee that $a(q)$ and $b(q)$ are finite.

Exercise 1.4. Show that in the definition (1.14) of total variation one cannot in general replace the supremum over partitions by the limit as the mesh of the partition tends to zero. (How about the indicator function of a single point?) But if F has some regularity, for example right-continuity, then the supremum can be replaced by the limit as $\text{mesh}(\pi) \to 0$.

Exercise 1.5. Let G be a continuous BV function on $[0, T]$ with Lebesgue-Stieltjes measure Λ_G on $(0, T]$, and let h be a Borel measurable function on $[0, T]$ that is integrable with respect to Λ_G. Define a function F on $[0, T]$ by $F(0) = 0$ and $F(t) = \int_{(0,t]} h(s)\,dG(s)$ for $t > 0$. Building on Lemma 1.15 and its proof, and remembering also (1.16), show that F is a continuous BV function on $[0, T]$. Note in particular the special case $F(t) = \int_0^t h(s)\,ds$.

Exercise 1.6. For a simple example of the failure of uncountable additivity for probabilities, let X be a $[0, 1]$-valued uniformly distributed random variable on $(\Omega, \mathcal{F}, P)$. Then $P(X = s) = 0$ for each individual $s \in [0, 1]$ but the union of these events over all s is Ω.

Exercise 1.7. Here is a useful formula for computing expectations. Suppose X is a nonnegative random variable, and h is a nondecreasing function on $\mathbf{R}_+$ such that $h(0) = 0$ and h is absolutely continuous on each bounded interval. (This last hypothesis is for ensuring that $h(a) = \int_0^a h'(s)\,ds$ for all $0 \le a < \infty$.) Then

$$Eh(X) = \int_0^\infty h'(s) P[X > s]\,ds. \tag{1.43}$$

Exercise 1.8. Suppose we need to prove something about a σ-algebra $\mathcal{B} = \sigma(\mathcal{E})$ on a space X, generated by a class $\mathcal{E}$ of subsets of X. A common strategy for proving such a result is to identify a suitable class $\mathcal{C}$ containing $\mathcal{E}$ whose members have the desired property. If $\mathcal{C}$ can be shown to be a σ-field then $\mathcal{B} \subseteq \mathcal{C}$ follows and thereby all members of $\mathcal{B}$ have the desired property. Here are useful examples.

(a) Fix two points x and y of the underlying space. Suppose that for each $A \in \mathcal{E}$, $\{x, y\} \subseteq A$ or $\{x, y\} \subseteq A^c$. Show that the same property is true for all $A \in \mathcal{B}$. In other words, if the generating sets do not separate x and y, neither does the σ-field.

(b) Suppose Φ is a collection of functions from a space X into a measurable space $(Y, \mathcal{H})$. Let $\mathcal{B} = \sigma\{f : f \in \Phi\}$ be the smallest σ-algebra that makes all functions $f \in \Phi$ measurable, as defined in (1.1). Suppose g is a function from another space Ω into X. Let Ω have σ-algebra $\mathcal{F}$. Show that g is a measurable function from $(\Omega, \mathcal{F})$ into $(X, \mathcal{B})$ iff for each $f \in \Phi$, $f \circ g$ is a measurable function from $(\Omega, \mathcal{F})$ into $(Y, \mathcal{H})$.

(c) In the setting of part (b), suppose two points x and y of X satisfy $f(x) = f(y)$ for all $f \in \Phi$. Show that for each $B \in \mathcal{B}$, $\{x, y\} \subseteq B$ or $\{x, y\} \subseteq B^c$.

(d) Let $S \subseteq X$ such that $S \in \mathcal{B}$. Let

$$\mathcal{B}_1 = \{B \in \mathcal{B} : B \subseteq S\} = \{A \cap S : A \in \mathcal{B}\}$$

be the restricted σ-field on the subspace S. Show that $\mathcal{B}_1$ is the σ-field generated on the space S by the collection

$$\mathcal{E}_1 = \{E \cap S : E \in \mathcal{E}\}.$$

Show by example that $\mathcal{B}_1$ is not necessarily generated by

$$\mathcal{E}_2 = \{E \in \mathcal{E} : E \subseteq S\}.$$

Hint: Consider $\mathcal{C} = \{B \subseteq X : B \cap S \in \mathcal{B}_1\}$. For the example, note that $\mathcal{B}_{\mathbf{R}}$ is generated by the class $\{(-\infty, a] : a \in \mathbf{R}\}$ but none of these infinite intervals lie in a bounded interval such as $[0, 1]$.

(e) Let $(X, \mathcal{A}, \nu)$ be a measure space. Let $\mathcal{U}$ be a sub-σ-field of $\mathcal{A}$, and let $\mathcal{N} = \{A \in \mathcal{A} : \nu(A) = 0\}$ be the collection of sets of ν-measure zero (ν-null sets). Let $\mathcal{U}^* = \sigma(\mathcal{U} \cup \mathcal{N})$ be the σ-field generated by $\mathcal{U}$ and $\mathcal{N}$. Show that

$$\mathcal{U}^* = \{A \in \mathcal{A} : \text{ there exists } U \in \mathcal{U} \text{ such that } U \triangle A \in \mathcal{N}\}.$$

$\mathcal{U}^*$ is called the *augmentation* of $\mathcal{U}$.

Exercise 1.9. Suppose $\{A_i\}$ and $\{B_i\}$ are sequences of measurable sets in a measure space $(X, \mathcal{A}, \nu)$ such that $\nu(A_i \triangle B_i) = 0$. Then

$$\nu\big((\cup A_i) \triangle (\cup B_i)\big) = \nu\big((\cap A_i) \triangle (\cap B_i)\big) = 0.$$

Exercise 1.10. (Product σ-algebras.) Recall the setting of Example 1.3. For a subset $L \subseteq \mathcal{I}$ of indices, let $\mathcal{B}_L = \sigma\{f_i : i \in L\}$ denote the σ-algebra generated by the projections f_i for $i \in L$. So in particular, $\mathcal{B}_\mathcal{I} = \bigotimes_{i \in \mathcal{I}} \mathcal{A}_i$ is the full product σ-algebra.

(a) Show that for each $B \in \mathcal{B}_\mathcal{I}$ there exists a countable set $L \subseteq \mathcal{I}$ such that $B \in \mathcal{B}_L$. *Hint.* Do not try to reason starting from a particular set $B \in \mathcal{B}_\mathcal{I}$. Instead, try to say something useful about the class of sets for which a countable L exists.

(b) Let $\mathbf{R}^{[0,\infty)}$ be the space of all functions $x : [0, \infty) \to \mathbf{R}$, with the product σ-algebra generated by the projections $x \mapsto x(t)$, $t \in [0, \infty)$. Show that the set of continuous functions is not measurable.

Exercise 1.11. (a) Let $\mathcal{E}_1, \dots, \mathcal{E}_n$ be collections of measurable sets on $(\Omega, \mathcal{F}, P)$, each closed under intersections (if $A, B \in \mathcal{E}_i$ then $A \cap B \in \mathcal{E}_i$). Suppose

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n)$$

for all $A_1 \in \mathcal{E}_1, \dots, A_n \in \mathcal{E}_n$. Show that the σ-algebras $\sigma(\mathcal{E}_1), \dots, \sigma(\mathcal{E}_n)$ are independent. *Hint.* A straight-forward application of the π–λ theorem B.3.

(b) Let $\{\mathcal{A}_i : i \in \mathcal{I}\}$ be a collection of independent σ-algebras. Let I_1, $\dots$, I_n be pairwise disjoint subsets of $\mathcal{I}$, and let $\mathcal{B}_k = \sigma\{\mathcal{A}_i : i \in I_k\}$ for $1 \le k \le n$. Show that $\mathcal{B}_1, \dots, \mathcal{B}_n$ are independent.

(c) Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be sub-σ-algebras of $\mathcal{F}$. Assume $\sigma\{\mathcal{B}, \mathcal{C}\}$ is independent of $\mathcal{A}$, and $\mathcal{C}$ is independent of $\mathcal{B}$. Show that $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are independent, and so in particular $\mathcal{C}$ is independent of $\sigma\{\mathcal{A}, \mathcal{B}\}$.

(d) Show by example that the independence of $\mathcal{C}$ and $\sigma\{\mathcal{A}, \mathcal{B}\}$ does not necessarily follow from having $\mathcal{B}$ independent of $\mathcal{A}$, $\mathcal{C}$ independent of $\mathcal{A}$, and $\mathcal{C}$ independent of $\mathcal{B}$. This last assumption is called *pairwise independence* of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$. *Hint.* An example can be built from two independent fair coin tosses.

Exercise 1.12. If you have never done so, compute the moments of a centered Gaussian random variable: for $Z \sim \mathcal{N}(0, \sigma^2)$ and $k \in \mathbf{N}$, $E(Z^{2k+1}) = 0$ and $E(Z^{2k}) = (2k-1)(2k-3) \cdots 1 \cdot \sigma^{2k}$. This is useful to know.

Exercise 1.13. (Memoryless property.) Let $X \sim \text{Exp}(\lambda)$. Show that, conditional on $X > c$, $X - c \sim \text{Exp}(\lambda)$. In plain English: given that I have waited for c time units, the remaining waiting time is still Exp(λ)-distributed. This is the memoryless property of the exponential distribution.

Exercise 1.14. Independence allows us to average separately. Here is a special case that will be used in a later proof. Let $(\Omega, \mathcal{F}, P)$ be a probability space, U and V measurable spaces, $X : \Omega \to U$ and $Y : \Omega \to V$ measurable mappings, $f : U \times V \to \mathbf{R}$ a bounded measurable function (with respect to the product σ-algebra on $U \times V$). Assume that X and Y are independent. Let μ be the distribution of Y on the V-space, defined by $\mu(B) = P\{Y \in B\}$ for measurable sets $B \subseteq V$. Show that

$$E[f(X, Y)] = \int_V E[f(X, y)]\, \mu(dy).$$

Hints. Start with functions of the type $f(x, y) = g(x)h(y)$. Use Theorem B.4 from the appendix.

Exercise 1.15. Let $\{D_i : i \in \mathbf{N}\}$ be a countable partition of Ω, by which we mean that $\{D_i\}$ are pairwise disjoint and $\Omega = \bigcup D_i$. Let $\mathcal{D}$ be the σ-algebra generated by $\{D_i\}$.

(a) Show that $G \in \mathcal{D}$ iff $G = \bigcup_{i \in I} D_i$ for some $I \subseteq \mathbf{N}$.

(b) Let $X \in L^1(P)$. Let $U = \{i \in \mathbf{N} : P(D_i) > 0\}$. Show that

$$E(X|\mathcal{D})(\omega) = \sum_{i : i \in U} \frac{E(\mathbf{1}_{D_i} X)}{P(D_i)} \cdot \mathbf{1}_{D_i}(\omega).$$

Exercise 1.16. Complete the exponential case in Example 1.25 by finding $E(X|\mathcal{A})(\omega)$ for $\omega \in [0, c]$. Then verify the identity $E\big[E(X|\mathcal{A})\big] = EX$. Do you understand why $E(X|\mathcal{A})$ must be constant on $[0, c]$ and on (c, ∞)?

Exercise 1.17. Suppose $P(A) = 0$ or 1 for all $A \in \mathcal{A}$. Show that $E(X|\mathcal{A}) = EX$ for all $X \in L^1(P)$.

Exercise 1.18. Let (X, Y) be an $\mathbf{R}^2$-valued random vector with joint density $f(x, y)$. This means that for any bounded Borel function ϕ on $\mathbf{R}^2$,

$$E[\phi(X, Y)] = \iint_{\mathbf{R}^2} \phi(x, y) f(x, y)\, dx\, dy.$$

The marginal density of Y is defined by

$$f_Y(y) = \int_{\mathbf{R}} f(x, y)\, dx.$$

Let

$$f(x|y) = \begin{cases} \dfrac{f(x, y)}{f_Y(y)}, & \text{if } f_Y(y) > 0 \\ 0, & \text{if } f_Y(y) = 0. \end{cases}$$

(a) Show that $f(x|y)f_Y(y) = f(x,y)$ for almost every (x,y), with respect to Lebesgue measure on $\mathbf{R}^2$. *Hint:* Let

$$H = \{(x,y) : f(x|y)f_Y(y) \neq f(x,y)\}.$$

Show that $m(H_y) = 0$ for each y-section of H, and use Tonelli's theorem.

(b) Show that $f(x|y)$ functions as a conditional density of X, given that $Y = y$, in this sense: for a bounded Borel function h on $\mathbf{R}$,

$$E[h(X)|Y](\omega) = \int_{\mathbf{R}} h(x)f(x|Y(\omega))\,dx.$$

Exercise 1.19. (Gaussian processes.) Let $\mathcal{I}$ be an arbitrary index set, and $c : \mathcal{I}^2 \to \mathbf{R}$ a positive definite function, which means that for any $t_1, \dots, t_n \in \mathcal{I}$ and real $\alpha_1, \dots, \alpha_n$,

$$\sum_{1 \le i,j \le n} \alpha_i \alpha_j c(t_i, t_j) \ge 0. \tag{1.44}$$

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and on that space a stochastic process $\{X_t : t \in \mathcal{I}\}$ such that, for each choice of indices $t_1, \dots, t_n \in \mathcal{I}$, the n-vector $(X_{t_1}, \dots, X_{t_n})$ has $\mathcal{N}(0, \Gamma)$ distribution with covariance matrix $\Gamma_{t_i,t_j} = c(t_i, t_j)$. A process whose finite dimensional distributions are Gaussian is called a Gaussian process.

Exercise 1.20. It is not too hard to write down impossible requirements for a stochastic process. Suppose $\{X_t : 0 \le t \le 1\}$ is a real-valued stochastic process that satisfies

(i) X_s and X_t are independent whenever $s \neq t$.

(ii) Each X_t has the same distribution, and variance 1.

(iii) The path $t \mapsto X_t(\omega)$ is continuous for almost every ω.

Show that a process satisfying these conditions cannot exist.

Chapter 2

Stochastic Processes

This chapter first covers general matters in the theory of stochastic processes, and then discusses the two most important processes, Brownian motion and Poisson processes.

2.1. Filtrations and stopping times

The set of nonnegative reals is denoted by $\mathbf{R}_+ = [0, \infty)$. Similarly $\mathbf{Q}_+$ for nonnegative rationals and $\mathbf{Z}_+ = \{0, 1, 2, \dots\}$ for nonnegative integers. The set of natural numbers is $\mathbf{N} = \{1, 2, 3, \dots\}$.

The discussion always takes place on a fixed probability space $(\Omega, \mathcal{F}, P)$. We will routinely assume that this space is complete as a measure space. This means that if $D \in \mathcal{F}$ and $P(D) = 0$, then all subsets of D lie in $\mathcal{F}$ and have probability zero. This is not a restriction because every measure space can be completed. See Section 1.1.4.

A *filtration* on a probability space $(\Omega, \mathcal{F}, P)$ is a collection of σ-fields $\{\mathcal{F}_t : t \in \mathbf{R}_+\}$ that satisfy

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \quad \text{for all } 0 \le s < t < \infty.$$

Whenever the index t ranges over nonnegative reals, we write simply $\{\mathcal{F}_t\}$ for $\{\mathcal{F}_t : t \in \mathbf{R}_+\}$. Given a filtration $\{\mathcal{F}_t\}$ we can add a last member to it by defining

$$\mathcal{F}_\infty = \sigma\Big(\bigcup_{0 \le t < \infty} \mathcal{F}_t \Big). \tag{2.1}$$

$\mathcal{F}_\infty$ is contained in $\mathcal{F}$ but can be strictly smaller than $\mathcal{F}$.

Example 2.1. The natural interpretation for the index $t \in \mathbf{R}_+$ is time. Let X_t denote the price of some stock at time t and assume it makes sense to imagine that it is defined for all $t \in \mathbf{R}_+$. At time t we know the entire evolution $\{X_s : s \in [0,t]\}$ up to the present. In other words, we are in possession of the information in the σ-algebra $\mathcal{F}_t^X = \sigma\{X_s : s \in [0,t]\}$. $\{\mathcal{F}_t^X\}$ is a basic example of a filtration.

We will find it convenient to assume that each $\mathcal{F}_t$ contains all subsets of $\mathcal{F}$-measurable P-null events. This is more than just assuming that each $\mathcal{F}_t$ is complete, but it entails no loss of generality. To achieve this, first complete $(\Omega, \mathcal{F}, P)$, and then replace $\mathcal{F}_t$ with

$$\bar{\mathcal{F}}_t = \{B \in \mathcal{F} : \text{ there exist } A \in \mathcal{F}_t \text{ such that } P(A \triangle B) = 0 \}. \tag{2.2}$$

Exercise 1.8(e) verified that $\bar{\mathcal{F}}_t$ is a σ-algebra. The filtration $\{\bar{\mathcal{F}}_t\}$ is sometimes called *complete*, or the *augmented* filtration.

At the most general level, a *stochastic process* is a collection of random variables $\{X_i : i \in \mathcal{I}\}$ indexed by some arbitrary index set $\mathcal{I}$, and all defined on the same probability space. For us the index set is most often $\mathbf{R}_+$ or some subset of it. Let $X = \{X_t : t \in \mathbf{R}_+\}$ be a process on $(\Omega, \mathcal{F}, P)$. It is convenient to regard X as a function on $\mathbf{R}_+ \times \Omega$ through the formula $X(t,\omega) = X_t(\omega)$. Indeed, we shall use the notations $X(t,\omega)$ and $X_t(\omega)$ interchangeably. When a process X is discussed without explicit mention of an index set, then $\mathbf{R}_+$ is assumed.

If the random variables X_t take their values in a space S, we say $X = \{X_t : t \in \mathbf{R}_+\}$ is an *S-valued process*. To even talk about S-valued random variables, S needs to have a σ-algebra so that a notion of measurability exists. Often in general accounts of the theory S is assumed a metric space, and then the natural σ-algebra on S is the Borel σ-field $\mathcal{B}_S$. We have no cause to consider anything more general than $S = \mathbf{R}^d$, the d-dimensional Euclidean space. Unless otherwise specified, in this section a process is always $\mathbf{R}^d$-valued. Of course, most important is the real-valued case with state space $\mathbf{R}^1 = \mathbf{R}$.

A process $X = \{X_t : t \in \mathbf{R}_+\}$ is *adapted* to the filtration $\{\mathcal{F}_t\}$ if X_t is $\mathcal{F}_t$-measurable for each $0 \le t < \infty$. The smallest filtration to which X is adapted is the filtration that it generates, defined by

$$\mathcal{F}_t^X = \sigma\{X_s : 0 \le s \le t\}.$$

A process X is *measurable* if X is $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable as a function from $\mathbf{R}_+ \times \Omega$ into $\mathbf{R}^d$. Furthermore, X is *progressively measurable* if the restriction of the function X to $[0,T] \times \Omega$ is $\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T$-measurable for each T. More explicitly, the requirement is that for each $B \in \mathcal{B}_{\mathbf{R}^d}$, the event

$$\{(t,\omega) \in [0,T] \times \Omega : X(t,\omega) \in B\}$$

lies in the σ-algebra $\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T$. If X is progressively measurable then it is also adapted, but the reverse implication does not hold (Exercise 2.6).

Properties of random objects are often interpreted in such a way that they are not affected by events of probability zero. For example, let $X = \{X_t : t \in \mathbf{R}_+\}$ and $Y = \{Y_t : t \in \mathbf{R}_+\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, P)$. As functions on $\mathbf{R}_+ \times \Omega$, X and Y are equal if $X_t(\omega) = Y_t(\omega)$ for each $\omega \in \Omega$ and $t \in \mathbf{R}_+$. A useful relaxation of this strict notion of equality is called indistinguishability. We say X and Y are *indistinguishable* if there exists an event $\Omega_0 \subseteq \Omega$ such that $P(\Omega_0) = 1$ and for each $\omega \in \Omega_0$, $X_t(\omega) = Y_t(\omega)$ for all $t \in \mathbf{R}_+$. Since most statements about processes ignore events of probability zero, for all practical purposes indistinguishable processes can be regarded as equal.

Another, even weaker notion is modification: Y is a *modification* (also called *version*) of X if for each t, $P\{X_t = Y_t\} = 1$.

Equality in distribution is also of importance for processes X and Y: $X \overset{d}{=} Y$ means that $P\{X \in A\} = P\{Y \in A\}$ for all measurable sets for which this type of statement makes sense. In all reasonable situations equality in distribution between processes follows from the weaker equality of finite-dimensional distributions:

$$\begin{aligned} &P\{X_{t_1} \in B_1, X_{t_2} \in B_2, \dots, X_{t_m} \in B_m\} \\ &\qquad = P\{Y_{t_1} \in B_1, Y_{t_2} \in B_2, \dots, Y_{t_m} \in B_m\} \end{aligned} \tag{2.3}$$

for all finite subsets $\{t_1, t_2, \dots, t_m\}$ of indices and all choices of the measurable sets $B_1, B_2, \dots, B_m$ in the range space. This can be proved for example with Lemma B.5 from the appendix.

Assuming that the probability space $(\Omega, \mathcal{F}, P)$ and the filtration $\{\mathcal{F}_t\}$ are complete conveniently avoids certain measurability complications. For example, if X is adapted and $P\{X_t = Y_t\} = 1$ for each $t \in \mathbf{R}_+$, then Y is also adapted. To see the reason, let $B \in \mathcal{B}_{\mathbf{R}^d}$, and note that

$$\{Y_t \in B\} = \{X_t \in B\} \cup \{Y_t \in B,\, X_t \notin B\} \setminus \{Y_t \notin B,\, X_t \in B\}.$$

Since all subsets of zero probability events lie in $\mathcal{F}_t$, we conclude that there are events $D_1, D_2 \in \mathcal{F}_t$ such that $\{Y_t \in B\} = \{X_t \in B\} \cup D_1 \setminus D_2$, which shows that Y is adapted.

In particular, since the point of view is that indistinguishable processes should really be viewed as one and the same process, it is sensible that such processes cannot differ in adaptedness or measurability.

A *stopping time* is a random variable $\tau : \Omega \to [0, \infty]$ such that $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ for each $0 \le t < \infty$. Many operations applied to stopping times produce new stopping times. Often used ones include the minimum and the maximum. If σ and τ are stopping times (for the same filtration) then

$$\{\sigma \wedge \tau \le t\} = \{\sigma \le t\} \cup \{\tau \le t\} \in \mathcal{F}_t$$

so $\sigma \wedge \tau$ is a stopping time. Similarly $\sigma \vee \tau$ can be shown to be a stopping time.

Example 2.2. Here is an illustration of the notion of stopping time.

(a) If you instruct your stockbroker to sell all your shares in company ABC on May 1, you are specifying a deterministic time. The time of sale does not depend on the evolution of the stock price.

(b) If you instruct your stockbroker to sell all your shares in company ABC as soon as the price exceeds 20, you are specifying a stopping time. Whether the sale happened by May 5 can be determined by inspecting the stock price of ABC Co. until May 5.

(c) Suppose you instruct your stockbroker to sell all your shares in company ABC on May 1 if the price will be lower on June 1. Again the sale time depends on the evolution as in case (b). But now the sale time is not a stopping time because to determine whether the sale happened on May 1 we need to look into the future.

So the notion of a stopping time is eminently sensible because it makes precise the idea that today's decisions must be based on the information available today and not on future information.

If τ is a stopping time, the σ-field of events known at time τ is defined by

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t \text{ for all } 0 \le t < \infty\}. \tag{2.4}$$

A deterministic time is a special case of a stopping time. If $\tau(\omega) = u$ for all ω, then $\mathcal{F}_\tau = \mathcal{F}_u$.

If $\{X_t\}$ is a process and τ is a stopping time, X_τ denotes the value of the process at the random time τ, in other words $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$. The random variable X_τ is defined on the event $\{\tau < \infty\}$, so not necessarily on the whole space Ω unless τ is finite. Or at least almost surely finite so that X_τ is defined with probability one.

Here are some basic properties of these concepts. Infinities arise naturally, and we use the conventions that $\infty \le \infty$ and $\infty = \infty$ are true, but $\infty < \infty$ is not.

Lemma 2.3. *Let σ and τ be stopping times, and X a process.*

(i) *For $A \in \mathcal{F}_\sigma$, the events $A \cap \{\sigma \le \tau\}$ and $A \cap \{\sigma < \tau\}$ lie in $\mathcal{F}_\tau$. In particular, $\sigma \le \tau$ implies $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.*

(ii) *Both τ and $\sigma \wedge \tau$ are $\mathcal{F}_\tau$-measurable. The events $\{\sigma \le \tau\}$, $\{\sigma < \tau\}$, and $\{\sigma = \tau\}$ lie in both $\mathcal{F}_\sigma$ and $\mathcal{F}_\tau$.*

(iii) *If the process X is progressively measurable then $X(\tau)$ is $\mathcal{F}_\tau$-measurable on the event $\{\tau < \infty\}$.*

Proof. Part (i). Let $A \in \mathcal{F}_\sigma$. For the first statement, we need to show that $(A \cap \{\sigma \le \tau\}) \cap \{\tau \le t\} \in \mathcal{F}_t$. Write

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

All terms above lie in $\mathcal{F}_t$. (a) The first by the definition of $A \in \mathcal{F}_\sigma$. (b) The second because both $\sigma \wedge t$ and $\tau \wedge t$ are $\mathcal{F}_t$-measurable random variables: for any $u \in \mathbf{R}$, $\{\sigma \wedge t \le u\}$ equals Ω if $u \ge t$ and $\{\sigma \le u\}$ if $u < t$, a member of $\mathcal{F}_t$ in both cases. (c) $\{\tau \le t\} \in \mathcal{F}_t$ since τ is a stopping time.

In particular, if $\sigma \le \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

To show $A \cap \{\sigma < \tau\} \in \mathcal{F}_\tau$, write

$$A \cap \{\sigma < \tau\} = \bigcup_{n \ge 1} A \cap \{\sigma + \tfrac{1}{n} \le \tau\}.$$

All members of the union on the right lie in $\mathcal{F}_\tau$ by the first part of the proof, because $\sigma \le \sigma + \frac{1}{n}$ implies $A \in \mathcal{F}_{\sigma+1/n}$. (And you should observe that for a constant $u \ge 0$, $\sigma + u$ is also a stopping time.)

Part (ii). Since $\{\tau \le s\} \cap \{\tau \le t\} = \{\tau \le s \wedge t\} \in \mathcal{F}_t \; \forall s$, by the definition of a stopping time, τ is $\mathcal{F}_\tau$-measurable. By the same token, the stopping time $\sigma \wedge \tau$ is $\mathcal{F}_{\sigma\wedge\tau}$-measurable, hence also $\mathcal{F}_\tau$-measurable.

Taking $A = \Omega$ in part (a) gives $\{\sigma \le \tau\}$ and $\{\sigma < \tau\} \in \mathcal{F}_\tau$. Taking the difference gives $\{\sigma = \tau\} \in \mathcal{F}_\tau$, and taking complements gives $\{\sigma > \tau\}$ and $\{\sigma \ge \tau\} \in \mathcal{F}_\tau$. Now we can interchange σ and τ in the conclusions.

Part (iii). We claim first that $\omega \mapsto X(\tau(\omega) \wedge t, \omega)$ is $\mathcal{F}_t$-measurable. To see this, write it as the composition

$$\omega \mapsto (\tau(\omega) \wedge t, \omega) \mapsto X(\tau(\omega) \wedge t, \omega).$$

The first step $\omega \mapsto (\tau(\omega) \wedge t, \omega)$ is measurable as a map from $(\Omega, \mathcal{F}_t)$ into the product space $\big([0,t] \times \Omega, \mathcal{B}_{[0,t]} \otimes \mathcal{F}_t\big)$ if the components have the correct measurability (Exercise 1.8(b)). It was already argued above in part (i) that $\omega \mapsto \tau(\omega) \wedge t$ is measurable from $\mathcal{F}_t$ into $\mathcal{B}_{[0,t]}$. The other component is the identity map $\omega \mapsto \omega$.

The second step of the composition is the map $(s, \omega) \mapsto X(s, \omega)$. By the progressive measurability assumption for X, this step is measurable from $\big([0,t] \times \Omega, \mathcal{B}_{[0,t]} \otimes \mathcal{F}_t\big)$ into $(\mathbf{R}^d, \mathcal{B}_{\mathbf{R}^d})$.

We have shown that $\{X_{\tau\wedge t} \in B\} \in \mathcal{F}_t$ for $B \in \mathcal{B}_{\mathbf{R}^d}$, and so

$$\{X_\tau \in B,\, \tau < \infty\} \cap \{\tau \le t\} = \{X_{\tau\wedge t} \in B\} \cap \{\tau \le t\} \in \mathcal{F}_t.$$

This shows that $\{X_\tau \in B,\, \tau < \infty\} \in \mathcal{F}_\tau$ which was the claim. $\square$

All the stochastic processes we study will have some regularity properties as functions of t, when ω is fixed. These are regularity properties of paths. A stochastic process $X = \{X_t : t \in \mathbf{R}_+\}$ is *continuous* if for each $\omega \in \Omega$, the path $t \mapsto X_t(\omega)$ is continuous as a function of t. The properties *left-continuous* and *right-continuous* have the obvious analogous meaning. An $\mathbf{R}^d$-valued process X is *right continuous with left limits* (or *cadlag* as the French acronym for this property goes) if the following is true for all $\omega \in \Omega$:

$$X_t(\omega) = \lim_{s \searrow t} X_s(\omega) \text{ for all } t \in \mathbf{R}_+, \text{ and}$$

$$\text{the left limit } \; X_{t-}(\omega) = \lim_{s \nearrow t} X_s(\omega) \; \text{ exists in } \mathbf{R}^d \text{ for all } t > 0.$$

Above $s \searrow t$ means that s approaches t from above (from the right), and $s \nearrow t$ approach from below (from the left). Finally, we also need to consider the reverse situation, namely a process that is *left continuous with right limits*, and for that we use the term *caglad*. Regularity properties of these types of functions are collected in Appendix A.1.

X is a *finite variation process* (FV process) if for each $\omega \in \Omega$ the path $t \mapsto X_t(\omega)$ has bounded variation on each compact interval $[0, T]$. In other words, the total variation function $V_{X(\omega)}(T) < \infty$ for each ω and T. But $V_{X(\omega)}(T)$ does not have to be bounded uniformly in ω.

We shall use all these terms also of a process that has a particular path property for almost every ω. For example, if $t \mapsto X_t(\omega)$ is continuous for all ω in a set Ω_0 of probability 1, then we can define $\widetilde{X}_t(\omega) = X_t(\omega)$ for $\omega \in \Omega_0$ and $\widetilde{X}_t(\omega) = 0$ for $\omega \notin \Omega_0$. Then $\widetilde{X}$ has all paths continuous, and X and $\widetilde{X}$ are indistinguishable. Since we regard indistinguishable processes as equal, it makes sense to regard X itself as a continuous process. When we prove results under hypotheses of path regularity, we assume that the path condition holds for each ω. Typically the result will be the same for processes that are indistinguishable.

Note, however, that processes that are modifications of each other can have quite different path properties (Exercise 2.5).

The next two lemmas record some technical benefits of path regularity.

Lemma 2.4. *Let X be adapted to the filtration $\{\mathcal{F}_t\}$, and suppose X is either left- or right-continuous. Then X is progressively measurable.*

Proof. Suppose X is right-continuous. Fix $T < \infty$. Define on $[0, T] \times \Omega$ the function

$$X_n(t, \omega) = X(0, \omega) \cdot \mathbf{1}_{\{0\}}(t) + \sum_{k=0}^{2^n - 1} X\big(\tfrac{(k+1)T}{2^n}, \omega\big) \cdot \mathbf{1}_{(kT2^{-n}, (k+1)T2^{-n}]}(t).$$

X_n is a sum of products of $\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T$-measurable functions, hence itself $\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T$-measurable. By right-continuity $X_n(t,\omega) \to X(t,\omega)$ as $n \to \infty$, hence X is also $\mathcal{B}_{[0,T]} \otimes \mathcal{F}_T$-measurable when restricted to $[0,T] \times \Omega$.

We leave the case of left-continuity as an exercise. □

Checking indistinguishability between two processes with some path regularity reduces to an a.s. equality check at a fixed time.

Lemma 2.5. *Suppose X and Y are right-continuous processes defined on the same probability space. Suppose $P\{X_t = Y_t\} = 1$ for all t in some dense countable subset S of $\mathbf{R}_+$. Then X and Y are indistinguishable. The same conclusion holds under the assumption of left-continuity if $0 \in S$.*

Proof. Let $\Omega_0 = \bigcap_{s \in S} \{\omega : X_s(\omega) = Y_s(\omega)\}$. By assumption, $P(\Omega_0) = 1$. Fix $\omega \in \Omega_0$. Given $t \in \mathbf{R}_+$, there exists a sequence s_n in S such that $s_n \searrow t$. By right-continuity,

$$X_t(\omega) = \lim_{n\to\infty} X_{s_n}(\omega) = \lim_{n\to\infty} Y_{s_n}(\omega) = Y_t(\omega).$$

Hence $X_t(\omega) = Y_t(\omega)$ for all $t \in \mathbf{R}_+$ and $\omega \in \Omega_0$, and this says X and Y are indistinguishable.

For the left-continuous case the origin $t = 0$ needs a separate assumption because it cannot be approached from the left. □

Filtrations also have certain kinds of limits and continuity properties. Given a filtration $\{\mathcal{F}_t\}$, define the σ-fields

$$\mathcal{F}_{t+} = \bigcap_{s:s>t} \mathcal{F}_s. \tag{2.5}$$

$\{\mathcal{F}_{t+}\}$ is a new filtration, and $\mathcal{F}_{t+} \supseteq \mathcal{F}_t$. If $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t, we say $\{\mathcal{F}_t\}$ is *right-continuous*. Performing the same operation again does not lead to anything new: if $\mathcal{G}_t = \mathcal{F}_{t+}$ then $\mathcal{G}_{t+} = \mathcal{G}_t$, as you should check. So in particular $\{\mathcal{F}_{t+}\}$ is a right-continuous filtration.

Right-continuity is the important property, but we could also define

$$\mathcal{F}_{0-} = \mathcal{F}_0 \quad \text{and} \quad \mathcal{F}_{t-} = \sigma\Bigl(\bigcup_{s:s<t} \mathcal{F}_s\Bigr) \quad \text{for } t > 0. \tag{2.6}$$

Since a union of σ-fields is not necessarily a σ-field, $\mathcal{F}_{t-}$ needs to be defined as the σ-field generated by the union of $\mathcal{F}_s$ over $s < t$. The generation step was unnecessary in the definition of $\mathcal{F}_{t+}$ because any intersection of σ-fields is again a σ-field. If $\mathcal{F}_t = \mathcal{F}_{t-}$ for all t, we say $\{\mathcal{F}_t\}$ is *left-continuous*. .

It is convenient to note that, since $\mathcal{F}_s$ depends on s in a monotone fashion, the definitions above can be equivalently formulated through sequences. For example, if $s_j > t$ is a sequence such that $s_j \searrow t$, then $\mathcal{F}_{t+} = \bigcap_j \mathcal{F}_{s_j}$.

The assumption that $\{\mathcal{F}_t\}$ is both complete and right-continuous is sometimes expressed by saying that $\{\mathcal{F}_t\}$ satisfies the *usual conditions*. In many books these are standing assumptions. When we develop the stochastic integral we assume the completeness. We shall not assume right-continuity as a routine matter, and we alert the reader whenever that assumption is used.

Since $\mathcal{F}_t \subseteq \mathcal{F}_{t+}$, $\{\mathcal{F}_{t+}\}$ admits more stopping times than $\{\mathcal{F}_t\}$. So there are benefits to having a right-continuous filtration. Let us explore this a little.

Lemma 2.6. *A $[0,\infty]$-valued random variable τ is a stopping time with respect to $\{\mathcal{F}_{t+}\}$ iff $\{\tau < t\} \in \mathcal{F}_t$ for all $t \in \mathbf{R}_+$.*

Proof. Suppose τ is an $\{\mathcal{F}_{t+}\}$-stopping time. Then for each $n \in \mathbf{N}$,

$$\{\tau \le t - n^{-1}\} \in \mathcal{F}_{(t-n^{-1})+} \subseteq \mathcal{F}_t,$$

and so $\{\tau < t\} = \bigcup_n \{\tau \le t - n^{-1}\} \in \mathcal{F}_t$.

Conversely, if $\{\tau < t + n^{-1}\} \in \mathcal{F}_{t+n^{-1}}$ for all $n \in \mathbf{N}$, then for all $m \in \mathbf{N}$, $\{\tau \le t\} = \bigcap_{n:n\ge m} \{\tau < t + n^{-1}\} \in \mathcal{F}_{t+m^{-1}}$. And so $\{\tau \le t\} \in \bigcap_m \mathcal{F}_{t+m^{-1}} = \mathcal{F}_{t+}$. ☐

Given a set H, define

$$\tau_H(\omega) = \inf\{t \ge 0 : X_t(\omega) \in H\}. \tag{2.7}$$

This is called the *hitting time* of the set H. If the infimum is taken over $t > 0$ then the above time is called the *first entry time* into the set H. These are the most important random times we wish to deal with, so it is crucial to know whether they are stopping times.

Lemma 2.7. *Let X be a process adapted to a filtration $\{\mathcal{F}_t\}$ and assume X is left- or right-continuous. If G is an open set, then τ_G is a stopping time with respect to $\{\mathcal{F}_{t+}\}$. In particular, if $\{\mathcal{F}_t\}$ is right-continuous, τ_G is a stopping time with respect to $\{\mathcal{F}_t\}$.*

Proof. If the path $s \mapsto X_s(\omega)$ is left- or right-continuous, $\tau_G(\omega) < t$ iff $X_s(\omega) \in G$ for some $s \in [0,t)$ iff $X_q(\omega) \in G$ for some rational $q \in [0,t)$. (If X is right-continuous, every value X_s for $s \in [0,t)$ can be approached from the right along values X_q for rational q. If X is left-continuous this is true for all values except $s = 0$, but 0 is among the rationals so it gets taken care of.) Thus we have

$$\{\tau_G < t\} = \bigcup_{q \in \mathbf{Q}_+ \cap [0,t)} \{X_q \in G\} \in \sigma\{X_s : 0 \le s < t\} \subseteq \mathcal{F}_t. \qquad \square$$

Example 2.8. Assuming X continuous would not improve the conclusion to $\{\tau_G \le t\} \in \mathcal{F}_t$. To see this, let $G = (b, \infty)$ for some $b > 0$ and consider the two paths

$$X_s(\omega_0) = X_s(\omega_1) = bs \quad \text{for } 0 \le s \le 1$$

while

$$\left.\begin{aligned} X_s(\omega_0) &= bs \\ X_s(\omega_1) &= b(2-s) \end{aligned}\right\} \quad \text{for } s \ge 1.$$

Now $\tau_G(\omega_0) = 1$ while $\tau_G(\omega_1) = \infty$. Since $X_s(\omega_0)$ and $X_s(\omega_1)$ agree for $s \in [0,1]$, the points ω_0 and ω_1 must be together either inside or outside any event in $\mathcal{F}_1^X$ [Exercise 1.8(c)]. But clearly $\omega_0 \in \{\tau_G \le 1\}$ while $\omega_1 \notin \{\tau_G \le 1\}$. This shows that $\{\tau_G \le 1\} \notin \mathcal{F}_1^X$.

There is an alternative way to register arrival into a set, if we settle for getting infinitesimally close. For a process X, let $X[s,t] = \{X(u) : s \le u \le t\}$, with (topological) closure $\overline{X[s,t]}$. For a set H define

$$\sigma_H = \inf\{t \ge 0 : \overline{X[0,t]} \cap H \ne \emptyset\}. \tag{2.8}$$

Note that for a cadlag path,

$$\overline{X[0,t]} = \{X(u) : 0 \le u \le t\} \cup \{X(u-) : 0 < u \le t\}. \tag{2.9}$$

Lemma 2.9. *Suppose X is a cadlag process adapted to $\{\mathcal{F}_t\}$ and H is a closed set. Then σ_H is a stopping time.*

Proof. Fix $t \ge 0$. First we claim that

$$\{\sigma_H \le t\} = \{X(0) \in H\} \cup \{X(s) \in H \text{ or } X(s-) \in H \text{ for some } s \in (0,t]\}.$$

It is clear that the event on the right is contained in the event on the left. To prove the opposite containment, suppose $\sigma_H \le t$. If the event on the right does *not* happen, then by the definition of σ_H as an infimum, for each $k \in \mathbf{N}$ there exists $t < u_k \le t + 1/k$ such that either $X(u_k) \in H$ or $X(u_k-) \in H$. Then $u_k \to t$, and by the right-continuity of paths, both $X(u_k)$ and $X(u_k-)$ converge to $X(t)$, which thus must lie in H. The equality above is checked.

Let

$$H_n = \{y : \text{ there exists } x \in H \text{ such that } |x - y| < n^{-1}\}$$

be the n^{-1}-neighborhood of H. Let U contain all the rationals in $[0,t]$ and the point t itself. Next we claim

$$\begin{aligned} &\{X(0) \in H\} \cup \{X(s) \in H \text{ or } X(s-) \in H \text{ for some } s \in (0,t]\} \\ &\qquad = \bigcap_{n=1}^{\infty} \bigcup_{q \in U} \{X(q) \in H_n\}. \end{aligned}$$

To justify this, note first that if $X(s) = y \in H$ for some $s \in [0,t]$ or $X(s-) = y \in H$ for some $s \in (0,t]$, then we can find a sequence $q_j \in U$ such that $X(q_j) \to y$, and then $X(q_j) \in H_n$ for all large enough j. Conversely, suppose we have $q_n \in U$ such that $X(q_n) \in H_n$ for all n. Extract a convergent subsequence $q_n \to s$. By the cadlag property a further subsequence of $X(q_n)$ converges to either $X(s)$ or $X(s-)$. By the closedness of H, one of these lies in H.

Combining the set equalities proved shows that $\{\sigma_H \le t\} \in \mathcal{F}_t$. □

Lemma 2.9 fails for caglad processes, unless the filtration is assumed right-continuous (Exercise 2.15). For a continuous process X and a closed set H the random times defined by (2.7) and (2.8) coincide. So we get this corollary.

Corollary 2.10. *Assume X is continuous and H is closed. Then τ_H is a stopping time.*

Remark 2.11. (A look ahead.) The stopping times discussed above will play a role in the development of the stochastic integral in the following way. To integrate an unbounded real-valued process X we need stopping times $\zeta_k \nearrow \infty$ such that $X_t(\omega)$ stays bounded for $0 < t \le \zeta_k(\omega)$. Caglad processes will be an important class of integrands. For a caglad X Lemma 2.7 shows that

$$\zeta_k = \inf\{t \ge 0 : |X_t| > k\}$$

are stopping times, provided $\{\mathcal{F}_t\}$ is right-continuous. Left-continuity of X then guarantees that $|X_t| \le k$ for $0 < t \le \zeta_k$.

Of particular interest will be a caglad process X that satisfies $X_t = Y_{t-}$ for $t > 0$ for some adapted cadlag process Y. Then by Lemma 2.9 we get the required stopping times by

$$\zeta_k = \inf\{t > 0 : |Y_t| \ge k \text{ or } |Y_{t-}| \ge k\}$$

without having to assume that $\{\mathcal{F}_t\}$ is right-continuous.

Remark 2.12. (You can ignore all the above hitting time complications.) The following is a known fact.

Theorem 2.13. *Assume the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions, and X is a progressively measurable process with values in some metric space. Then τ_H defined by (2.7), or the same with infimum restricted to $t > 0$, are stopping times for every Borel set H.*

This is a deep theorem. A fairly accessible recent proof appears in [1]. The reader may prefer to use this theorem in the sequel, and always assume that filtrations satisfy the usual conditions. We shall not do so in the text to avoid proliferating the mysteries we have to accept without justification.

2.2. Quadratic variation

In stochastic analysis many processes turn out to have infinite total variation and it becomes necessary to use quadratic variation as a measure of path oscillation. For example, we shall see in the next chapter that if a continuous martingale M has finite variation, then $M_t = M_0$.

Let Y be a stochastic process. For a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{m(\pi)} = t\}$ of $[0, t]$ we can form the sum of squared increments

$$\sum_{i=0}^{m(\pi)-1} (Y_{t_{i+1}} - Y_{t_i})^2.$$

We say that these sums converge to the random variable $[Y]_t$ in probability as $\operatorname{mesh}(\pi) = \max_i(t_{i+1} - t_i) \to 0$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P\biggl\{ \biggl| \sum_{i=0}^{m(\pi)-1} (Y_{t_{i+1}} - Y_{t_i})^2 - [Y]_t \biggr| \geq \varepsilon \biggr\} \leq \varepsilon \tag{2.10}$$

for all partitions π with $\operatorname{mesh}(\pi) \leq \delta$. We express this limit as

$$\lim_{\operatorname{mesh}(\pi)\to 0} \sum_i (Y_{t_{i+1}} - Y_{t_i})^2 = [Y]_t \quad \text{in probability.} \tag{2.11}$$

Definition 2.14. The *quadratic variation process* $[Y] = \{[Y]_t : t \in \mathbf{R}_+\}$ of a stochastic process Y is a process such that $[Y]_0 = 0$, the paths $t \mapsto [Y]_t(\omega)$ are nondecreasing for all ω, and the limit (2.11) holds for all $t \geq 0$.

We will see in the case of Brownian motion that limit (2.11) cannot be required to hold almost surely, unless we pick the partitions carefully. Hence limits in probability are used.

Between two processes we define a quadratic covariation in terms of quadratic variations.

Definition 2.15. Let X and Y be two stochastic processes on the same probability space. The *(quadratic) covariation process* $[X, Y] = \{[X, Y]_t : t \in \mathbf{R}_+\}$ is defined by

$$[X, Y] = \bigl[\tfrac{1}{2}(X + Y)\bigr] - \bigl[\tfrac{1}{2}(X - Y)\bigr] \tag{2.12}$$

provided the quadratic variation processes on the right exist in the sense of Definition 2.14.

From the identity $ab = \frac{1}{4}(a + b)^2 - \frac{1}{4}(a - b)^2$ applied to $a = X_{t_{i+1}} - X_{t_i}$ and $b = Y_{t_{i+1}} - Y_{t_i}$ it follows that, for each $t \in \mathbf{R}_+$,

$$\lim_{\operatorname{mesh}(\pi)\to 0} \sum_i (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) = [X, Y]_t \quad \text{in probability.} \tag{2.13}$$

Utilizing these limits together with the identities

$$ab = \tfrac{1}{2}\big((a+b)^2 - a^2 - b^2\big) = \tfrac{1}{2}\big(a^2 + b^2 - (a-b)^2\big)$$

gives these almost sure equalities at fixed times:

$$[X,Y]_t = \tfrac{1}{2}\big([X+Y]_t - [X]_t - [Y]_t\big) \quad \text{a.s.} \tag{2.14}$$

and

$$[X,Y]_t = \tfrac{1}{2}\big([X]_t + [Y]_t - [X-Y]_t\big) \quad \text{a.s.} \tag{2.15}$$

provided all the processes in question exist.

We defined $[X,Y]$ by (2.12) instead of by the limits (2.13) so that the property $[Y,Y]=[Y]$ is immediately clear. Limits (2.13) characterize $[X,Y]_t$ only almost surely at each fixed t and imply nothing about path properties. This is why monotonicity was made part of Definition 2.14. Monotonicity ensures that the quadratic variation of the identically zero process is indistinguishable from the zero process. (Exercise 2.5 gives another process that satisfies limits (2.11) for $Y=0$.)

We give a preliminary discussion of quadratic variation and covariation in this section. That quadratic variation exists for Brownian motion and the Poisson process will be proved later in this chapter. The case of local martingales is discussed in Section 3.4 of Chapter 3. Existence of quadratic variation for FV processes follows purely analytically from Lemma A.10 in Appendix A.

In the next proposition we show that if X is cadlag, then we can take $[X]_t$ also cadlag. Technically, we show that for each t, $[X]_{t+} = [X]_t$ almost surely. Then the process $[X]_{t+}$ has cadlag paths (Exercise 2.12) and satisfies the definition of quadratic variation. From definition (2.12) it then follows that $[X,Y]$ can also be taken cadlag when X and Y are cadlag.

Furthermore, we show that the jumps of the covariation match the jumps of the processes. For any cadlag process Z, the *jump at t* is denoted by

$$\Delta Z(t) = Z(t) - Z(t-).$$

The statement below about jumps can be considered preliminary. After developing more tools we can strengthen the statement for semimartingales Y so that the equality $\Delta[Y]_t = (\Delta Y_t)^2$ is true for all $t \in \mathbf{R}_+$ outside a single exceptional event of probability zero. (See Lemma 5.40.)

Proposition 2.16. *Suppose X and Y are cadlag processes, and $[X,Y]$ exists in the sense of Definition 2.15. Then there exists a cadlag modification of $[X,Y]$. For any t, $\Delta[X,Y]_t = (\Delta X_t)(\Delta Y_t)$ almost surely.*

Proof. It suffices to treat the case $X = Y$. Pick $\delta, \varepsilon > 0$. Fix $t < u$. Pick $\eta > 0$ so that

$$P\biggl\{ \biggl| [X]_u - [X]_t - \sum_{i=0}^{m(\pi)-1} (X_{t_{i+1}} - X_{t_i})^2 \biggr| < \varepsilon \biggr\} > 1 - \delta \tag{2.16}$$

whenever $\pi = \{t = t_0 < t_1 < \cdots < t_m(\pi) = u\}$ is a partition of $[t, u]$ with $\mathrm{mesh}(\pi) < \eta$. (The little observation needed here is in Exercise 2.10.) Pick such a partition π. Keeping t_1 fixed, refine π further in $[t_1, u]$ so that

$$P\biggl\{ \biggl| [X]_u - [X]_{t_1} - \sum_{i=1}^{m(\pi)-1} (X_{t_{i+1}} - X_{t_i})^2 \biggr| < \varepsilon \biggr\} > 1 - \delta.$$

Taking the intersection of these events, we have that with probability at least $1 - 2\delta$,

$$\begin{aligned} [X]_u - [X]_t &\le \sum_{i=0}^{m(\pi)-1} (X_{t_{i+1}} - X_{t_i})^2 + \varepsilon \\ &= (X_{t_1} - X_t)^2 + \sum_{i=1}^{m(\pi)-1} (X_{t_{i+1}} - X_{t_i})^2 + \varepsilon \\ &\le (X_{t_1} - X_t)^2 + [X]_u - [X]_{t_1} + 2\varepsilon \end{aligned}$$

which rearranges to

$$[X]_{t_1} \le [X]_t + (X_{t_1} - X_t)^2 + 2\varepsilon.$$

Looking back, we see that this argument works for any $t_1 \in (t, t + \eta)$. By the monotonicity $[X]_{t+} \le [X]_{t_1}$, so for all these t_1,

$$P\bigl\{[X]_{t+} \le [X]_t + (X_{t_1} - X_t)^2 + 2\varepsilon\bigr\} > 1 - 2\delta.$$

Shrink $\eta > 0$ further so that for $t_1 \in (t, t + \eta)$ by right continuity

$$P\bigl\{(X_{t_1} - X_t)^2 < \varepsilon\bigr\} > 1 - \delta.$$

The final estimate is

$$P\bigl\{[X]_t \le [X]_{t+} \le [X]_t + 3\varepsilon\bigr\} > 1 - 3\delta.$$

Since $\varepsilon, \delta > 0$ were arbitrary, it follows that $[X]_{t+} = [X]_t$ almost surely. As explained before the statement of the proposition, this implies that we can choose a version of $[X]$ with cadlag paths.

To bound the jump at u, return to the partition π chosen for (2.16). Let $s = t_{m(\pi)-1}$. Keeping s fixed refine π sufficiently in $[t, s]$ so that, with probability at least $1-2\delta$,

$$\begin{aligned}[X]_u - [X]_t &\le \sum_{i=0}^{m(\pi)-1} (X_{t_{i+1}} - X_{t_i})^2 + \varepsilon \\ &= (X_u - X_s)^2 + \sum_{i=0}^{m(\pi)-2} (X_{t_{i+1}} - X_{t_i})^2 + \varepsilon \\ &\le (X_u - X_s)^2 + [X]_s - [X]_t + 2\varepsilon\end{aligned}$$

which rearranges, through $\Delta[X]_u \le [X]_u - [X]_s$, to give

$$P\big\{\Delta[X]_u \le (X_u - X_s)^2 + 2\varepsilon\big\} > 1 - 2\delta.$$

Here $s \in (u-\eta, u)$ was arbitrary. Again we can pick η small enough so that for all such s with probability at least $1-\delta$,

$$\big|(X_u - X_s)^2 - (\Delta X_u)^2\big| < \varepsilon. \tag{2.17}$$

Since ε, δ are arbitrary, we have $\Delta[X]_u \le (\Delta X_u)^2$ almost surely.

For the other direction, return to the calculation above with partition π and $s = t_{m(\pi)-1}$, but now derive opposite inequalities: with probability at least $1-2\delta$,

$$[X]_u - [X]_s \ge (X_u - X_s)^2 - \varepsilon \ge (\Delta X_u)^2 - 2\varepsilon.$$

We can take s close enough to u so that $P\{\Delta[X]_u \ge [X]_u - [X]_s - \varepsilon\} > 1-\delta$. Then we have $\Delta[X]_u \ge (\Delta X_u)^2 - 3\varepsilon$ with probability at least $1-3\delta$. Letting ε, δ to zero once more completes the proof. □

In particular, we can say that in the cadlag case $[Y]$ is an increasing process, according to this definition.

Definition 2.17. An *increasing process* $A = \{A_t : 0 \le t < \infty\}$ is an adapted process such that, for almost every ω, $A_0(\omega) = 0$ and $s \mapsto A_s(\omega)$ is nondecreasing and right-continuous. Monotonicity implies the existence of left limits A_{t-}, so it follows that an increasing process is cadlag.

Next two useful inequalities.

Lemma 2.18. *Suppose the processes below exist. Then at a fixed t,*

$$|\,[X,Y]_t| \le [X]_t^{1/2}[Y]_t^{1/2} \quad a.s. \tag{2.18}$$

and more generally for $0 \le s < t$

$$|\,[X,Y]_t - [X,Y]_s| \le \big([X]_t - [X]_s\big)^{1/2}\big([Y]_t - [Y]_s\big)^{1/2} \quad a.s. \tag{2.19}$$

Furthermore,

$$\big|\,[X]_t - [Y]_t\,\big| \le [X-Y]_t + 2[X-Y]_t^{1/2}[Y]_t^{1/2} \quad a.s. \tag{2.20}$$

In the cadlag case the inequalities are valid simultaneously at all $s < t \in \mathbf{R}_+$, with probability 1.

Proof. The last statement follows from the earlier ones because the inequalities are a.s. valid simultaneously at all rational times and then limits capture all time points.

Inequalities (2.18)–(2.19) follow from the Cauchy-Schwarz inequality

$$\Big|\sum x_i y_i\Big| \le \Big(\sum x_i^2\Big)^{1/2}\Big(\sum y_i^2\Big)^{1/2}. \tag{2.21}$$

For (2.19) use also the observation (Exercise 2.10) that the increments of the (co)variation processes are limits of sums over partitions, as in (2.16).

From the identity $a^2 - b^2 = (a-b)^2 + 2(a-b)b$ applied to increments of X and Y follows

$$[X] - [Y] = [X-Y] + 2[X-Y, Y].$$

Utilizing (2.18),

$$\begin{aligned} \big|\,[X] - [Y]\,\big| &\le [X-Y] + \big|\,2[X-Y,Y]\,\big| \\ &\le [X-Y] + 2[X-Y]^{1/2}[Y]^{1/2}. \end{aligned}$$

□

In all our applications $[X, Y]$ will be a cadlag process. As the difference of two increasing processes in Definition (2.12), $[X, Y]_t$ is BV on any compact time interval. Lebesgue-Stieltjes integrals over time intervals with respect to $[X, Y]$ have an important role in the development. These are integrals with respect to the Lebesgue-Stieltjes measure $\Lambda_{[X,Y]}$ defined by

$$\Lambda_{[X,Y]}(a, b] = [X, Y]_b - [X, Y]_a, \quad 0 \le a < b < \infty,$$

as explained in Section 1.1.9. Note that there is a hidden ω in all these quantities. This integration over time is done separately for each fixed ω. This kind of operation is called "path by path" because ω represents the path of the underlying process $[X, Y]$.

When the origin is included in the time interval, we assume $[X, Y]_{0-} = 0$, so the Lebesgue-Stieltjes measure $\Lambda_{[X,Y]}$ gives zero measure to the singleton $\{0\}$. The Lebesgue-Stieltjes integrals obey the following useful inequality.

Proposition 2.19 (Kunita-Watanabe inequality)**.** *Fix ω such that $[X]$, $[Y]$ and $[X, Y]$ exist and are right-continuous on the interval $[0, T]$. Then for any $\mathcal{B}_{[0,T]} \otimes \mathcal{F}$-measurable bounded functions G and H on $[0, T] \times \Omega$,*

$$\begin{aligned} &\left|\int_{[0,T]} G(t,\omega)H(t,\omega)\,d[X,Y]_t(\omega)\right| \\ &\qquad \le \left\{\int_{[0,T]} G(t,\omega)^2\,d[X]_t(\omega)\right\}^{1/2}\left\{\int_{[0,T]} H(t,\omega)^2\,d[Y]_t(\omega)\right\}^{1/2}. \end{aligned} \tag{2.22}$$

The integrals above are Lebesgue-Stieltjes integrals with respect to the t-variable, evaluated for the fixed ω.

Proof. Once ω is fixed, the result is an analytic lemma, and the dependence of G and H on ω is irrelevant. We included this dependence so that the statement better fits its later applications. It is a property of product-measurability that for a fixed ω, $G(t,\omega)$ and $H(t,\omega)$ are measurable functions of t.

Consider first step functions

$$g(t) = \alpha_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{m-1} \alpha_i \mathbf{1}_{(s_i, s_{i+1}]}(t)$$

and

$$h(t) = \beta_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{m-1} \beta_i \mathbf{1}_{(s_i, s_{i+1}]}(t)$$

where $0 = s_1 < \cdots < s_m = T$ is a partition of $[0,T]$. (Note that g and h can be two arbitrary step functions. If they come with distinct partitions, $\{s_i\}$ is the common refinement of these partitions.) Then

$$\begin{aligned}
\left| \int_{[0,T]} g(t)h(t)\, d[X,Y]_t \right| &= \left| \sum_i \alpha_i \beta_i \bigl([X,Y]_{s_{i+1}} - [X,Y]_{s_i} \bigr) \right| \\
&\le \sum_i |\alpha_i \beta_i| \bigl([X]_{s_{i+1}} - [X]_{s_i} \bigr)^{1/2} \bigl([Y]_{s_{i+1}} - [Y]_{s_i} \bigr)^{1/2} \\
&\le \left\{ \sum_i |\alpha_i|^2 \bigl([X]_{s_{i+1}} - [X]_{s_i} \bigr) \right\}^{1/2} \left\{ \sum_i |\beta_i|^2 \bigl([Y]_{s_{i+1}} - [Y]_{s_i} \bigr) \right\}^{1/2} \\
&= \left\{ \int_{[0,T]} g(t)^2\, d[X]_t \right\}^{1/2} \left\{ \int_{[0,T]} h(t)^2\, d[Y]_t \right\}^{1/2}
\end{aligned}$$

where we applied (2.19) and then Schwarz inequality (2.21).

Let g and h be two arbitrary bounded Borel functions on $[0,T]$, and pick $0 < C < \infty$ so that $|g| \le C$ and $|h| \le C$. Let $\varepsilon > 0$. Define the bounded Borel measure

$$\mu = \Lambda_{[X]} + \Lambda_{[Y]} + |\Lambda_{[X,Y]}|$$

on $[0,T]$. Above, $\Lambda_{[X]}$ is the positive Lebesgue-Stieltjes measure of the function $t \mapsto [X]_t$ (for the fixed ω under consideration), same for $\Lambda_{[Y]}$, and $|\Lambda_{[X,Y]}|$ is the positive total variation measure of the signed Lebesgue-Stieltjes measure $\Lambda_{[X,Y]}$. By Lemma A.17 we can choose step functions $\widetilde{g}$ and $\widetilde{h}$ so that $|\widetilde{g}| \le C$, $|\widetilde{h}| \le C$, and

$$\int \bigl(|g - \widetilde{g}| + |h - \widetilde{h}| \bigr)\, d\mu < \frac{\varepsilon}{2C}.$$

On the one hand

$$\left|\int_{[0,T]} gh\,d[X,Y]_t - \int_{[0,T]} \widetilde{g}\widetilde{h}\,d[X,Y]_t\right| \le \int_{[0,T]} \bigl|gh - \widetilde{g}\widetilde{h}\,\bigr|\,d|\Lambda_{[X,Y]}|$$
$$\le C\int_{[0,T]} |g-\widetilde{g}|\,d|\Lambda_{[X,Y]}| + C\int_{[0,T]} |h-\widetilde{h}|\,d|\Lambda_{[X,Y]}| \le \varepsilon.$$

On the other hand,

$$\left|\int_{[0,T]} g^2\,d[X]_t - \int_{[0,T]} \widetilde{g}^2\,d[X]_t\right| \le \int_{[0,T]} \bigl|g^2 - \widetilde{g}^2\,\bigr|\,d[X]_t$$
$$\le 2C\int_{[0,T]} |g-\widetilde{g}|\,d[X]_t \le \varepsilon,$$

with a similar bound for h. Putting these together with the inequality already proved for step functions gives

$$\left|\int_{[0,T]} gh\,d[X,Y]_t\right| \le \varepsilon + \left\{\varepsilon + \int_{[0,T]} g^2\,d[X]_t\right\}^{1/2}\left\{\varepsilon + \int_{[0,T]} h^2\,d[Y]_t\right\}^{1/2}.$$

Since $\varepsilon > 0$ was arbitrary, we can let $\varepsilon \to 0$. The inequality as stated in the proposition is obtained by choosing $g(t) = G(t,\omega)$ and $h(t) = H(t,\omega)$. □

Remark 2.20. Inequality (2.22) has the following corollary. As in the proof, let $|\Lambda_{[X,Y](\omega)}|$ be the total variation measure of the signed Lebesgue-Stieltjes measure $\Lambda_{[X,Y](\omega)}$ on $[0,T]$. For a fixed ω, (1.13) implies that $|\Lambda_{[X,Y](\omega)}| \ll \Lambda_{[X,Y](\omega)}$ and the Radon-Nikodym derivative

$$\phi(t) = \frac{d|\Lambda_{[X,Y](\omega)}|}{d\Lambda_{[X,Y](\omega)}}(t)$$

on $[0,T]$ satisfies $|\phi(t)| \le 1$. For an arbitrary bounded Borel function g on $[0,T]$

$$\int_{[0,T]} g(t)\,|\Lambda_{[X,Y](\omega)}|(dt) = \int_{[0,T]} g(t)\phi(t)\,d[X,Y]_t(\omega).$$

Combining this with (2.22) gives

$$\int_{[0,T]} \bigl|G(t,\omega)H(t,\omega)\bigr|\,|\Lambda_{[X,Y](\omega)}|(dt)$$
$$\le \left\{\int_{[0,T]} G(t,\omega)^2\,d[X]_t(\omega)\right\}^{1/2}\left\{\int_{[0,T]} H(t,\omega)^2\,d[Y]_t(\omega)\right\}^{1/2}. \tag{2.23}$$

2.3. Path spaces and Markov processes

So far we have thought of a stochastic process as a collection of random variables on a probability space. An extremely fruitful, more abstract view regards a process as a probability distribution on a *path space*. This is a natural generalization of the notion of probability distribution of a random variable or a random vector. If $Y = (Y_1, \dots, Y_n)$ is an $\mathbf{R}^n$-valued random vector on a probability space $(\Omega, \mathcal{F}, P)$, its distribution μ is the Borel probability measure on $\mathbf{R}^n$ defined by

$$\mu(B) = P\{\omega : Y(\omega) \in B\}, \quad B \in \mathcal{B}_{\mathbf{R}^n}.$$

One can even forget about the "abstract" probability space $(\Omega, \mathcal{F}, P)$, and redefine Y on the "concrete" space $(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \mu)$ as the identity random variable $Y(s) = s$ for $s \in \mathbf{R}^n$.

To generalize this notion for an $\mathbf{R}^d$-valued process $X = \{X_t : 0 \le t < \infty\}$, we have to choose a suitable measurable space U so that X can be thought of as a measurable map $X : \Omega \to U$. For a fixed ω the value $X(\omega)$ is the function $t \mapsto X_t(\omega)$, so the space U has to be a space of functions, or a "path space." The path regularity of X determines which space U will do. Here are the three most important choices.

(i) Without any further assumptions, X is a measurable map into the product space $(\mathbf{R}^d)^{[0,\infty)}$ with product σ-field $\mathcal{B}(\mathbf{R}^d)^{\otimes[0,\infty)}$.

(ii) If X is an $\mathbf{R}^d$-valued cadlag process, then a suitable path space is $D = D_{\mathbf{R}^d}[0,\infty)$, the space of $\mathbf{R}^d$-valued cadlag functions ξ on $[0,\infty)$, with the σ-algebra generated by the coordinate projections $\xi \mapsto \xi(t)$ from D into $\mathbf{R}^d$. It is possible to define a metric on D that makes it a complete, separable metric space, and under which the Borel σ-algebra is the one generated by the coordinate mappings. This is the so-called *Skorohod metric*, see for example [**2, 6**]. Thus we can justifiably denote this σ-algebra by $\mathcal{B}_D$.

(iii) If X is an $\mathbf{R}^d$-valued continuous process, then X maps into $C = C_{\mathbf{R}^d}[0,\infty)$, the space of $\mathbf{R}^d$-valued continuous functions on $[0,\infty)$. This space is naturally metrized by

$$r(\eta, \zeta) = \sum_{k=1}^{\infty} 2^{-k}\Bigl(1 \wedge \sup_{0\le t\le k} |\eta(t) - \zeta(t)|\Bigr), \quad \eta, \zeta \in C. \tag{2.24}$$

This is the metric of uniform convergence on compact sets. (C, r) is a complete, separable metric space, and its Borel σ-algebra $\mathcal{B}_C$ is generated by the coordinate mappings. C is a subspace of D, and indeed the notions of convergence and measurability in C coincide with the notions it inherits as a subspace of D.

Generating the σ-algebra of the path space with the coordinate functions guarantees that X is a measurable mapping from Ω into the path space (Exercise 1.8(b)). Then we can define the distribution μ of the process on the path space. For example, if X is cadlag, then define $\mu(B) = P\{X \in B\}$ for $B \in \mathcal{B}_D$. As in the case of the random vector, we can switch probability spaces. Take $(D, \mathcal{B}_D, \mu)$ as the new probability space, and define the process $\{Y_t\}$ on D via the coordinate mappings: $Y_t(\omega) = \omega(t)$ for $\omega \in D$. Then the old process X and the new process Y have the same distribution, because by definition

$$P\{X \in B\} = \mu(B) = \mu\{\omega \in D : \omega \in B\} = \mu\{\omega \in D : Y(\omega) \in B\}$$
$$= \mu\{Y \in B\}.$$

One benefit from this construction is that it leads naturally to a theory of weak convergence of processes, which is used in many applications. This comes from specializing the well-developed theory of weak convergence of probability measures on metric spaces to the case of a path space.

The two most important general classes of stochastic processes are martingales and Markov processes. Both classes are defined by the relationship of the process $X = \{X_t\}$ to a filtration $\{\mathcal{F}_t\}$. It is always first assumed that $\{X_t\}$ is adapted to $\{\mathcal{F}_t\}$.

Let X be a real-valued process. Then X is a *martingale* with respect to $\{\mathcal{F}_t\}$ if X_t is integrable for each t, and

$$E[X_t|\mathcal{F}_s] = X_s \text{ for all } s < t.$$

For the definition of a Markov process X can take its values in an abstract space, but $\mathbf{R}^d$ is sufficiently general for us. An $\mathbf{R}^d$-valued process X satisfies the *Markov property* with respect to $\{\mathcal{F}_t\}$ if

$$P[X_t \in B|\mathcal{F}_s] = P[X_t \in B|X_s] \text{ for all } s < t \text{ and } B \in \mathcal{B}_{\mathbf{R}^d}. \tag{2.25}$$

A martingale represents a fair gamble in the sense that, given all the information up to the present time s, the expectation of the future fortune X_t is the same as the current fortune X_s. Stochastic analysis relies heavily on martingale theory. The Markov property is a notion of causality. It says that, given the present state X_s, future events are independent of the past.

These notions are of course equally sensible in discrete time. Let us give the most basic example in discrete time, since that is simpler than continuous time. Later in this chapter we will have sophisticated continuous-time examples when we discuss Brownian motion and Poisson processes.

Example 2.21. (Random Walk) Let $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. random variables. Define the partial sums by $S_0 = 0$, and $S_n = X_1 + \cdots + X_n$ for $n \geq 1$. Then S_n is a *Markov chain* (the term for a Markov process in discrete time). If $EX_i = 0$ then S_n is a martingale. The natural filtration to use here is the one generated by the process: $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$.

Martingales are treated in Chapter 3. In the remainder of this section we discuss the Markov property and then the strong Markov property. These topics are not necessary for all that follows, but we do make use of the Markov property of Brownian motion for some calculations and exercises.

With Markov processes it is natural to consider the whole family of processes obtained by varying the initial state. In the previous example, to have the random walk start at x, we simply say $S_0 = x$ and $S_n = x + X_1 + \cdots + X_n$. The definition of such a family of processes is conveniently expressed in terms of probability distributions on a path space. Below we give the definition of a *time-homogeneous* cadlag Markov process. Time-homogeneity means that the transition mechanism does not change with time: given that the state of the process is x at time s, the chances of residing in set B at time $s+t$ depend on (x, t, B) but not on s.

On the path space D, the *coordinate variables* are defined by $X_t(\omega) = \omega(t)$ for $\omega \in D$, and the natural filtration is $\mathcal{F}_t = \sigma\{X_s : s \le t\}$. We write also $X(t)$ when subscripts are not convenient. The *shift maps* $\theta_s : D \to D$ are defined by $(\theta_s\omega)(t) = \omega(s+t)$. In other words, the path $\theta_s\omega$ has its time origin translated to s and the path before s is deleted. For an event $A \in \mathcal{B}_D$, the inverse image

$$\theta_s^{-1} A = \{\omega \in D : \theta_s\omega \in A\}$$

represents the event that "A happens starting at time s."

Definition 2.22. An $\mathbf{R}^d$-valued *Markov process* is a collection $\{P^x : x \in \mathbf{R}^d\}$ of probability measures on $D = D_{\mathbf{R}^d}[0, \infty)$ with these properties:

(a) $P^x\{\omega \in D : \omega(0) = x\} = 1$.

(b) For each $A \in \mathcal{B}_D$, the function $x \mapsto P^x(A)$ is measurable on $\mathbf{R}^d$.

(c) $P^x[\theta_t^{-1} A | \mathcal{F}_t](\omega) = P^{X_t(\omega)}(A)$ for P^x-almost every $\omega \in D$, for every $x \in \mathbf{R}^d$ and $A \in \mathcal{B}_D$.

Requirement (a) in the definition says that x is the initial state under the measure P^x. You should think of P^x as the probability distribution of the entire process given that the initial state is x. Requirement (b) is for technical purposes. If we wish to start the process in a random initial state X_0 with distribution μ, then we use the path measure $P^\mu(A) = \int P^x(A)\,\mu(dx)$. P^x itself is the special case $\mu = \delta_x$.

Requirement (c) is the Markov property. It appears qualitatively different from (2.25) because the event A can depend on the entire process, but in fact (c) would be just as powerful if it were stated with an event of the type $A = \{X_s \in B\}$. The general case can then be derived by first going inductively to finite-dimensional events, and then by a π-λ argument to all $A \in \mathcal{B}_D$. In the context of Markov processes E^x stands for expectation under the measure P^x. Parts (b) and (c) together imply (2.25), with the help of property (viii) of Theorem 1.26.

What (c) says is that, if we know that the process is in state y at time t, then regardless of the past, the future of the process behaves exactly as a new process started from y. Technically, conditional on $X_t = y$ and the entire information in $\mathcal{F}_t$, the probabilities of the future process $\{X_{t+u} : u \geq 0\}$ obey the measure P^y. Informally we say that at each time point the Markov process *restarts itself* from its current state, forgetting its past. (Exercise 2.16 practices this in a simple calculation.)

The *transition probability* of the Markov process is $p(t, x, B) = P^x(X_t \in B)$. For each fixed $t \in \mathbf{R}_+$, $p(t, x, B)$ is a measurable function of $x \in \mathbf{R}^d$ and a Borel probability measure in the set argument $B \in \mathcal{B}_{\mathbf{R}^d}$. It gives the conditional probability in (2.25), regardless of the initial distribution:

$$P^\mu(X_{s+t} \in B|\mathcal{F}_s)(\omega) = p(t, X_s(\omega), B). \tag{2.26}$$

The next calculation justifies. Let $A \in \mathcal{F}_s$ and $B \in \mathcal{B}_{\mathbf{R}^d}$.

$$\begin{aligned}
E^\mu[\mathbf{1}_A \mathbf{1}\{X_{s+t} \in B\}] &= \int E^x[\mathbf{1}_A \mathbf{1}\{X_{s+t} \in B\}]\,\mu(dx) \\
&= \int E^x[\mathbf{1}_A(\omega) P^x(X_{s+t} \in B \,|\, \mathcal{F}_s)(\omega)]\,\mu(dx) \\
&= \int E^x\big[\mathbf{1}_A(\omega) P^x(\theta_s^{-1}\{X_t \in B\} \,|\, \mathcal{F}_s)(\omega)\big]\,\mu(dx) \\
&= \int E^x\big[\mathbf{1}_A(\omega) P^{X_s(\omega)}\{X_t \in B\}\big]\,\mu(dx) \\
&= \int E^x[\mathbf{1}_A(\omega) p(t, X_s(\omega), B)]\,\mu(dx) \\
&= E^\mu[\mathbf{1}_A\, p(t, X_s, B)].
\end{aligned}$$

The fourth equality above used property (c) of Definition 2.22. The integration variable ω was introduced temporarily to indicate those quantities that are integrated by the outside expectation E^x.

Finite-dimensional distributions of the Markov process are iterated integrals of the transition probabilities. For $0 = s_0 < s_1 < \cdots < s_n$ and a bounded Borel function f on $\mathbf{R}^{d(n+1)}$,

$$\begin{aligned}
E^\mu[f(X_{s_0}, X_{s_1}, \dots, X_{s_n})] &= \int \cdots \int \mu(dx_0) p(s_1, x_0, dx_1) \\
&\quad p(s_2 - s_1, x_1, dx_2) \cdots p(s_n - s_{n-1}, x_{n-1}, dx_n)\, f(x_0, x_1, \dots, x_n).
\end{aligned} \tag{2.27}$$

(Exercise 2.17.)

There is also a related semigroup property for a family of operators. On bounded measurable functions g on the state space, define operators $S(t)$ by

$$S(t)g(x) = E^x[g(X_t)]. \tag{2.28}$$

(The way to think about this is that $S(t)$ maps g into a new function $S(t)g$, and the formula above tells us how to compute the values $S(t)g(x)$.) $S(0)$

is the identity operator: $S(0)g = g$. The *semigroup property* is $S(s+t) = S(s)S(t)$ where the multiplication of $S(s)$ and $S(t)$ means composition. This can be checked with the Markov property:

$$\begin{aligned} S(s+t)g(x) &= E^x[g(X_{s+t})] = E^x\big[E^x\{g(X_{s+t}) \,|\, \mathcal{F}_s\}\big] \\ &= E^x\big[E^{X(s)}\{g(X_t)\}\big] = E^x\big[\big(S(t)g\big)(X_s)\big] \\ &= S(s)\big(S(t)g\big)(x). \end{aligned}$$

Another question suggests itself. Statement (c) in Definition 2.22 restarts the process at a particular time t from the state y that the process is at. But since the Markov process is supposed to forget its past and transitions are time-homogeneous, should not the same restarting take place for example at the first visit to state y? This scenario is not covered by statement (c) because this first visit happens at a random time. So we ask whether we can replace time t in part (c) with a stopping time τ. With some additional regularity the answer is affirmative. The property we are led to formulate is called the *strong Markov property*.

The additional regularity needed is that the probability distribution $P^x(X_t \in \cdot\,)$ of the state at a fixed time t is a continuous function of the initial state x. Recall the notion of weak convergence of probability measures for spaces more general than $\mathbf{R}$ introduced below Definition 1.19. A Markov process $\{P^x\}$ is called a *Feller process* if for all $t \ge 0$, all $x_j, x \in \mathbf{R}^d$, and all $g \in C_b(\mathbf{R}^d)$ (the space of bounded continuous functions on $\mathbf{R}^d$)

$$x_j \to x \quad \text{implies} \quad E^{x_j}[g(X_t)] \to E^x[g(X_t)]. \tag{2.29}$$

Equation (2.30) below is the strong Markov property, and a Markov process that satisfies this property is a strong Markov process. It is formulated for a function Y of both a time point and a path, hence more generally than the Markov property in Definition 2.22(c). This is advantageous for some applications. Again, the state space could really be any metric space but for concreteness we think of $\mathbf{R}^d$.

Theorem 2.23. *Let $\{P^x\}$ be a Feller process with state space $\mathbf{R}^d$. Let $Y(s,\omega)$ be a bounded, jointly measurable function of $(s,\omega) \in \mathbf{R}_+ \times D$ and τ a stopping time on D. Let the initial state x be arbitrary. Then on the event $\{\tau < \infty\}$ the equality*

$$E^x[Y(\tau, X \circ \theta_\tau) \,|\, \mathcal{F}_\tau](\omega) = E^{\omega(\tau)}[Y(\tau(\omega), X)] \tag{2.30}$$

holds for P^x-almost every path ω.

Before turning to the proof, let us sort out the meaning of statement (2.30). First, we introduced a random shift θ_τ defined by $(\theta_\tau\omega)(s) = \omega(\tau(\omega) + s)$ for those paths ω for which $\tau(\omega) < \infty$. On both sides of the identity inside the expectations X denotes the identity random variable on D: $X(\omega) = \omega$. The purpose of writing $Y(\tau, X \circ \theta_\tau)$ is to indicate that the path argument in Y has been translated by τ, so that the integrand $Y(\tau, X \circ \theta_\tau)$ on the left is a function of the future process after time τ. On the right the expectation should be read like this:

$$E^{\omega(\tau)}[Y(\tau(\omega), X)] = \int_D Y(\tau(\omega), \widetilde{\omega})\, P^{\omega(\tau)}(d\widetilde{\omega}).$$

In other words, the first argument of Y on the right inherits the value $\tau(\omega)$ which (note!) has been fixed by the conditioning on $\mathcal{F}_\tau$ on the left side of (2.30). The path argument $\widetilde{\omega}$ obeys the probability measure $P^{\omega(\tau)}$, in other words it is the process restarted from the current state $\omega(\tau)$.

Example 2.24. Let us perform a simple calculation to illustrate the mechanics. Let $\tau = \inf\{t \ge 0 : X_t = z \text{ or } X_{t-} = z\}$ be the first hitting time of point z, a stopping time by Lemma 2.9. Suppose the process is continuous which means that $P^x(C) = 1$ for all x. Suppose further that $P^x(\tau < \infty) = 1$. From this and path continuity $P^x(X_\tau = z) = 1$. Let us use (2.30) to check that, for a fixed $t > 0$, $P^x(X_{\tau+t} \in A) = P^z(X_t \in A)$. We take $Y(\omega) = \mathbf{1}\{\omega(t) \in A\}$ so now the Y-function does not need a time argument.

$$\begin{aligned} P^x(X_{\tau+t} \in A) &= E^x[Y \circ \theta_\tau] = E^x\big[E^x(Y \circ \theta_\tau \,|\, \mathcal{F}_\tau)\big] \\ &= E^x[E^{X(\tau)}(Y)] = E^x[E^z(Y)] = E^z(Y) = P^z(X_t \in A). \end{aligned}$$

The E^x-expectation goes away because the integrand $E^z(Y)$ is no longer random, it is merely a constant.

Remark 2.25. Up to now in our discussion we have used the filtration $\{\mathcal{F}_t\}$ on D or C generated by the coordinates. Our important examples, such as Brownian motion and the Poisson process, actually satisfy the Markov property (part (c) of Definition 2.22) under the larger filtration $\{\mathcal{F}_{t+}\}$. The strong Markov property holds also for the larger filtration because the proof below goes through as long as the Markov property is true.

Proof of Theorem 2.23. Let $A \in \mathcal{F}_\tau$. What needs to be shown is that

$$E^x\big[\mathbf{1}_A \mathbf{1}_{\{\tau<\infty\}} Y(\tau, X \circ \theta_\tau)\big] = \int_{A\cap\{\tau<\infty\}} E^{\omega(\tau)}[Y(\tau(\omega), X)]\, P^x(d\omega). \tag{2.31}$$

As so often with these proofs, we do a calculation for a special case and then appeal to a general principle to complete the proof.

First we assume that all the possible finite values of τ can be arranged in an increasing sequence $t_1 < t_2 < t_3 < \cdots$. Then

$$\begin{aligned}
E^x\big[\mathbf{1}_A\mathbf{1}_{\{\tau<\infty\}}Y(\tau, X\circ\theta_\tau)\big] &= \sum_n E^x\big[\mathbf{1}_A\mathbf{1}_{\{\tau=t_n\}}Y(\tau, X\circ\theta_\tau)\big]\\
&= \sum_n E^x\big[\mathbf{1}_A\mathbf{1}_{\{\tau=t_n\}}Y(t_n, X\circ\theta_{t_n})\big]\\
&= \sum_n E^x\big[\mathbf{1}_A(\omega)\mathbf{1}_{\{\tau(\omega)=t_n\}}E^x\{Y(t_n, X\circ\theta_{t_n})\,|\,\mathcal{F}_{t_n}\}(\omega)\big]\\
&= \sum_n E^x\big[\mathbf{1}_A(\omega)\mathbf{1}_{\{\tau(\omega)=t_n\}}E^{\omega(t_n)}\{Y(t_n, X)\}\big]\\
&= E^x\big[\mathbf{1}_A(\omega)\mathbf{1}_{\{\tau(\omega)<\infty\}}E^{\omega(\tau(\tau))}\{Y(\tau(\omega), X)\}\big]
\end{aligned}$$

where we used the basic Markov property and wrote ω for the integration variable in the last two E^x-expectations in order to not mix this up with the variable X inside the $E^{\omega(\tau)}$-expectation. The expression $\omega(\tau(\omega))$ does make sense since $\tau(\omega) \in \mathbf{R}_+$!

Now let τ be a general stopping time. The next stage is to define $\tau_n = 2^{-n}([2^n\tau]+1)$. Check that these are stopping times that satisfy $\{\tau_n < \infty\} = \{\tau < \infty\}$ and $\tau_n \searrow \tau$ as $n \nearrow \infty$. Since $\tau_n > \tau$, $A \in \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$. The possible finite values of τ_n are $\{2^{-n}k : k \in \mathbf{N}\}$ and so we already know the result for τ_n:

$$\begin{aligned}
&E^x\big[\mathbf{1}_A\mathbf{1}_{\{\tau<\infty\}}Y(\tau_n, X\circ\theta_{\tau_n})\big]\\
&\qquad = \int_{A\cap\{\tau<\infty\}} E^{\omega(\tau_n)}[Y(\tau_n(\omega), X)]\,P^x(d\omega).
\end{aligned} \tag{2.32}$$

The idea is to let $n \nearrow \infty$ above and argue that in the limit we recover (2.31). For this we need some continuity. We take Y of the following type:

$$Y(s,\omega) = f_0(s)\cdot\prod_{i=1}^m f_i(\omega(s_i)) \tag{2.33}$$

where $0 \le s_1 < \ldots < s_n$ are time points and $f_0, f_1, \ldots, f_m$ are bounded continuous functions. Then on the left of (2.32) we have inside the expectation

$$\begin{aligned}
Y(\tau_n, \theta_{\tau_n}\omega) &= f_0(\tau_n)\cdot\prod_{i=1}^m f_i(\omega(\tau_n + s_i))\\
&\underset{n\to\infty}{\longrightarrow} f_0(\tau)\cdot\prod_{i=1}^m f_i(\omega(\tau + s_i)) = Y(\tau, \theta_\tau\omega).
\end{aligned}$$

The limit above used the *right* continuity of the path ω. Dominated convergence will now take the left side of (2.32) to the left side of (2.31).

On the right of (2.32) we have inside the integral

$$\begin{aligned}E^{\eta(\tau_n)}[Y(\tau_n(\eta), X)] &= \int Y(\tau_n(\eta), \omega)\, P^{\eta(\tau_n)}(d\omega)\\ &= f_0(\tau_n(\eta)) \int \prod_{i=1}^m f_i(\omega(s_i))\, P^{\eta(\tau_n)}(d\omega).\end{aligned}$$

In order to assert that the above converges to the integrand on the right side of (2.31), we need to prove the continuity of

$$f_0(s) \int \prod_{i=1}^m f_i(\omega(s_i))\, P^x(d\omega)$$

as a function of $(s, x) \in \mathbf{R}_+ \times \mathbf{R}^d$. Since products preserve continuity, what really is needed is the extension of the definition (2.29) of the Feller property to expectations of more complicated continuous functions of the path. We leave this to the reader (Exercise 2.18). Now dominated convergence again takes the right side of (2.32) to the right side of (2.31).

We have verified (2.31) for Y of type (2.33), and it remains to argue that this extends to all bounded measurable Y. This will follow from Theorem B.4. The class $\mathcal{R}$ in that Theorem are sets of the type $B = \{(s, \omega) : s \in A_0, \omega(s_1) \in A_1, \ldots, \omega(s_m) \in A_m\}$ where $A_0 \subseteq \mathbf{R}_+$ and $A_i \subseteq \mathbf{R}^d$ are closed sets. (2.33) holds for $\mathbf{1}_B$ because we can find continuous $0 \le f_{i,n} \le 1$ such that $f_{i,n} \to \mathbf{1}_{A_i}$ as $n \to \infty$, and then the corresponding $Y_n \to \mathbf{1}_B$ and (2.33) extends by dominated convergence. Sets of type B generate $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{B}_D$ because this σ-algebra is generated by coordinate functions. This completes the proof of the strong Markov property for Feller processes. ☐

Next we discuss the two most important processes, Brownian motion and the Poisson process.

2.4. Brownian motion

Informally Brownian motion could be characterized as random walk that takes infinitesimal steps infinitely fast. Its paths are continuous but highly oscillatory.

Definition 2.26. One some probability space $(\Omega, \mathcal{F}, P)$, let $\{\mathcal{F}_t\}$ be a filtration and $B = \{B_t : 0 \le t < \infty\}$ an adapted real-valued stochastic process. Then B is a one-dimensional *Brownian motion* with respect to $\{\mathcal{F}_t\}$ if it has these two properties.

(i) For almost every ω, the path $t \mapsto B_t(\omega)$ is continuous.

(ii) For $0 \le s < t$, $B_t - B_s$ is independent of $\mathcal{F}_s$ and has normal distribution with mean zero and variance $t - s$.

If furthermore

(iii) $B_0 = 0$ almost surely

then B is a *standard Brownian motion*. Since the definition involves both the process and the filtration, sometimes one calls this B an $\{\mathcal{F}_t\}$*–Browian motion*, or the pair $\{B_t, \mathcal{F}_t : 0 \le t < \infty\}$ is called the Brownian motion.

To be explicit, point (ii) of the definition can be expressed by the requirement

$$E\big[Z \cdot h(B_t - B_s)\big] = E(Z) \cdot \frac{1}{\sqrt{2\pi(t-s)}} \int_{\mathbf{R}} h(x) \exp\Big\{-\frac{x^2}{2(t-s)}\Big\}\, dx$$

for all bounded $\mathcal{F}_s$-measurable random variables Z and all bounded Borel functions h on $\mathbf{R}$. By an inductive argument, it follows that for any $0 \le s_0 < s_1 < \cdots < s_n$, the increments

$$B_{s_1} - B_{s_0}, B_{s_2} - B_{s_1}, \dots, B_{s_n} - B_{s_{n-1}}$$

are independent random variables and independent of $\mathcal{F}_{s_0}$ (Exercise 2.20). Furthermore, the joint distribution of the increments is not changed by a shift in time: namely, the joint distribution of the increments above is the same as the joint distribution of the increments

$$B_{t+s_1} - B_{t+s_0}, B_{t+s_2} - B_{t+s_1}, \dots, B_{t+s_n} - B_{t+s_{n-1}}$$

for any $t \ge 0$. These two points are summarized by saying that Brownian motion has *stationary, independent increments.*

A *d-dimensional standard Brownian motion* is an $\mathbf{R}^d$-valued process $B_t = (B_t^1, \dots, B_t^d)$ with the property that each component B_t^i is a one-dimensional standard Brownian motion (relative to the underlying filtration $\{\mathcal{F}_t\}$), and the coordinates B^1, B^2, …, B^d are independent. This is equivalent to requiring that

(i) $B_0 = 0$ almost surely.

(ii) For almost every ω, the path $t \mapsto B_t(\omega)$ is continuous.

(iii) For $0 \le s < t$, $B_t - B_s$ is independent of $\mathcal{F}_s$, and has multivariate normal distribution with mean zero and covariance matrix $(t-s)\underset{d\times d}{I}$.

Above, $\underset{d\times d}{I}$ is the $d \times d$ identity matrix.

To create a Brownian motion B_t with a more general initial distribution μ (the probability distribution of B_0), take a standard Brownian motion $(\widetilde{B}_t, \widetilde{\mathcal{F}}_t)$ and a μ-distributed random variable X independent of $\widetilde{\mathcal{F}}_\infty$, and define $B_t = X + \widetilde{B}_t$. The filtration is now $\mathcal{F}_t = \sigma\{X, \widetilde{\mathcal{F}}_t\}$. Since $B_t - B_s = \widetilde{B}_t - \widetilde{B}_s$, $\widetilde{\mathcal{F}}_\infty$ is independent of X, and $\widetilde{B}_t - \widetilde{B}_s$ is independent of $\widetilde{\mathcal{F}}_s$, Exercise

1.11(c) implies that $B_t - B_s$ is independent of $\mathcal{F}_s$. Conversely, if a process B_t satisfies parts (i) and (ii) of Definition 2.26, then $\widetilde{B}_t = B_t - B_0$ is a standard Brownian motion, independent of B_0.

The construction (proof of existence) of Brownian motion is rather technical, and hence relegated to Section B.2 in the Appendix. For the underlying probability space the construction uses the "canonical" path space $C = C_{\mathbf{R}}[0,\infty)$. Let $B_t(\omega) = \omega(t)$ be the coordinate projections on C, and $\mathcal{F}_t^B = \sigma\{B_s : 0 \le s \le t\}$ the filtration generated by the coordinate process.

Theorem 2.27. *There exists a Borel probability measure P^0 on the path space $C = C_{\mathbf{R}}[0,\infty)$ such that the process $B = \{B_t : 0 \le t < \infty\}$ on the probability space $(C, \mathcal{B}_C, P^0)$ is a standard one-dimensional Brownian motion with respect to the filtration $\{\mathcal{F}_t^B\}$.*

The proof of this existence theorem relies on the Kolmogorov Extension Theorem 1.28. The probability measure P^0 on C constructed in the theorem is called *Wiener measure* to recognize that Norbert Wiener was the first to give a rigorous construction of Brownian motion. Brownian motion itself is sometimes also called the *Wiener process*. Once we know that Brownian motion starting at the origin exists, we can construct Brownian motion with an arbitrary initial point (random or deterministic) following the description after Definition 2.26.

The construction gives us the following regularity property of paths. Fix $0 < \gamma < \frac{1}{2}$. For P^0-almost every $\omega \in C$,

$$\sup_{0 \le s < t \le T} \frac{|B_t(\omega) - B_s(\omega)|}{|t-s|^\gamma} < \infty \quad \text{for all } T < \infty. \tag{2.34}$$

This property is expressed by saying that Brownian paths are *Hölder continuous* with exponent γ. We shall show later in this section that this property is not true for $\gamma > \frac{1}{2}$.

2.4.1. Brownian motion as a martingale and a strong Markov process. We discuss mainly properties of the one-dimensional case. The multidimensional versions of the statements follow naturally from the one-dimensional case. We shall use the term Brownian motion to denote a process that satisfies (i) and (ii) of Definition 2.26, and call it standard if also $B_0 = 0$.

A fundamental property of Brownian motion is that it is both a martingale and a Markov process.

Proposition 2.28. *Suppose $B = \{B_t\}$ is a Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$ on $(\Omega, \mathcal{F}, P)$. Then B_t and $B_t^2 - t$ are martingales with respect to $\{\mathcal{F}_t\}$.*

Proof. Follows from the properties of Brownian increments and basic properties of conditional expectations. Let $s < t$.

$$E[B_t|\mathcal{F}_s] = E[B_t - B_s|\mathcal{F}_s] + E[B_s|\mathcal{F}_s] = B_s,$$

and

$$\begin{aligned} E[B_t^2|\mathcal{F}_s] &= E[(B_t - B_s + B_s)^2|\mathcal{F}_s] \\ &= E[(B_t - B_s)^2|\mathcal{F}_s] + 2B_s E[B_t - B_s|\mathcal{F}_s] + B_s^2 \\ &= (t - s) + B_s^2. \end{aligned}$$

□

Next we show that Brownian motion restarts itself independently of the past. This is the heart of the Markov property. Also, it is useful to know that the filtration of a Brownian motion can always be both augmented with the null events and made right-continuous.

Proposition 2.29. *Suppose $B = \{B_t\}$ is a Brownian motion with respect to a filtration $\{\mathcal{F}_t\}$ on $(\Omega, \mathcal{F}, P)$.*

(a) *We can assume that $\mathcal{F}_t$ contains every set A for which there exists an event $N \in \mathcal{F}$ such that $A \subseteq N$ and $P(N) = 0$. (This is the notion of a complete or augmented filtration introduced earlier.) Furthermore, $B = \{B_t\}$ is also a Brownian motion with respect to the right-continuous filtration $\{\mathcal{F}_{t+}\}$.*

(b) *Fix $s \in \mathbf{R}_+$ and define $Y_t = B_{s+t} - B_s$. Then the process $Y = \{Y_t : 0 \le t < \infty\}$ is independent of $\mathcal{F}_{s+}$ and it is a standard Brownian motion with respect to the filtration $\{\mathcal{G}_t\}$ defined by $\mathcal{G}_t = \mathcal{F}_{(s+t)+}$.*

Proof. Definition (2.2) shows how to complete the filtration. Of course, the adaptedness of B to the filtration is not harmed by enlarging the filtration, the issue is the independence of $\bar{\mathcal{F}}_s$ and $B_t - B_s$. If $G \in \mathcal{F}$ has $A \in \mathcal{F}_s$ such that $P(A \triangle G) = 0$, then $P(G \cap H) = P(A \cap H)$ for any event H. In particular, the independence of $\bar{\mathcal{F}}_s$ from $B_t - B_s$ follows.

The rest follows from a single calculation. Fix $s \ge 0$ and $0 = t_0 < t_1 < t_2 < \cdots < t_n$, and for $h \ge 0$ abbreviate

$$\xi(h) = (B_{s+h+t_1} - B_{s+h}, B_{s+h+t_2} - B_{s+h+t_1}, \ldots, B_{s+h+t_n} - B_{s+h+t_{n-1}})$$

for a vector of Brownian increments. Let Z be a bounded $\mathcal{F}_{s+}$-measurable random variable. For each $h > 0$, Z is $\mathcal{F}_{s+h}$-measurable. Since Brownian increments are independent of the past (see the strengthening of property (ii) of Definition 2.26 given in Exercise 2.20), $\xi(h)$ is independent of Z. Let f be a bounded continuous function on $\mathbf{R}^n$. By path continuity, independence, and finally the stationarity of Brownian increments,

$$
\begin{aligned}
&E\big[\,Z\cdot f(\xi(0))\big] = \lim_{h\searrow 0} E\big[\,Z\cdot f(\xi(h))\big] \\
&= \lim_{h\searrow 0} EZ\cdot E\big[\,f(\xi(h))\big] = EZ\cdot E\big[\,f(\xi(0))\big] \\
&= EZ\cdot E\big[\,f(B_{s+t_1}-B_s, B_{s+t_2}-B_{s+t_1},\dots,B_{s+t_n}-B_{s+t_{n-1}})\big] \\
&= EZ\cdot E\big[\,f(B_{t_1}-B_0, B_{t_2}-B_{t_1},\dots,B_{t_n}-B_{t_{n-1}})\big].
\end{aligned} \tag{2.35}
$$

The equality of the first and last members of the above calculation extends by Lemma B.6 from continuous f to bounded Borel f.

Now we can harvest the conclusions. The independence of $\mathcal{F}_{s+}$ and $B_t - B_s$ is contained in (2.35) so the fact that B is a Brownian motion with respect to $\{\mathcal{F}_{t+}\}$ has been proved. Secondly, since $\xi(0) = (Y_{t_1}, Y_{t_2} - Y_{t_1}, \dots, Y_{t_n} - Y_{t_{n-1}})$ and the vector $\eta = (Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$ is a function of $\xi(0)$, we conclude that η is independent of $\mathcal{F}_{s+}$. This being true for all choices of time points $0 < t_1 < t_2 < \cdots < t_n$ implies that the entire process Y is independent of $\mathcal{F}_{s+}$, and the last member of (2.35) shows that given $\mathcal{F}_{s+}$, Y has the distribution of standard Brownian motion.

Finally, the independence of $Y_{t_2} - Y_{t_1}$ and $\mathcal{G}_{t_1}$ is the same as the independence of $B_{s+t_2} - B_{s+t_1}$ of $\mathcal{F}_{(s+t_1)+}$ which was already argued. □

Parts (a) and (b) of the lemma together assert that Y is a standard Brownian motion, independent of $\bar{\mathcal{F}}_{t+}$, the filtration obtained by replacing $\{\mathcal{F}_t\}$ with the augmented right-continuous version. (The order of the two operations on the filtration is immaterial, in other words the σ-algebra $\bigcap_{s:s>t} \bar{\mathcal{F}}_s$ agrees with the augmentation of $\bigcap_{s:s>t} \mathcal{F}_s$, see Exercise 2.4.)

The key calculation (2.35) of the previous proof used only right-continuity. Thus the same argument gives us this lemma that we can apply to other processes.

Lemma 2.30. *Suppose $X = (X_t : t \in \mathbf{R}_+)$ is a right-continuous process adapted to a filtration $\{\mathcal{F}_t\}$ and for all $s < t$ the increment $X_t - X_s$ is independent of $\mathcal{F}_s$. Then $X_t - X_s$ is independent of $\bar{\mathcal{F}}_{s+}$.*

Next we develop some properties of Brownian motion by concentrating on the "canonical setting". The underlying probability space is the path space $C = C_{\mathbf{R}}[0,\infty)$ with the coordinate process $B_t(\omega) = \omega(t)$ and the filtration $\mathcal{F}^B_t = \sigma\{B_s : 0 \le s \le t\}$ generated by the coordinates. For each $x \in \mathbf{R}$ there is a probability measure P^x on C under which $B = \{B_t\}$ is Brownian motion started at x. Expectation under P^x is denoted by E^x and satisfies

$$E^x[H] = E^0[H(x + B)]$$

for any bounded $\mathcal{B}_C$-measurable function H. On the right $x + B$ is a sum of a point and a process, interpreted as the process whose value at time t is $x + B_t$. (In Theorem 2.27 we constructed P^0, and the equality above can be taken as the definition of P^x.)

On C we have the shift maps $\{\theta_s : 0 \le s < \infty\}$ defined by $(\theta_s\omega)(t) = \omega(s+t)$ that move the time origin to s. The shift acts on the process B by $\theta_s B = \{B_{s+t} : t \ge 0\}$.

A consequence of Lemma 2.29(a) is that the coordinate process B is a Brownian motion also relative to the larger filtration $\mathcal{F}^B_{t+} = \bigcap_{s:s>t} \mathcal{F}^B_s$. We shall show that members of $\mathcal{F}^B_t$ and $\mathcal{F}^B_{t+}$ differ only by null sets. (These σ-algebras are different, see Exercise 2.13.) This will have interesting consequences when we take $t = 0$. The next proposition establishes the Markov property with respect to the larger filtration $\{\mathcal{F}^B_{t+}\}$.

Proposition 2.31. *Let H be a bounded $\mathcal{B}_C$-measurable function on C.*

(a) *$E^x[H]$ is a Borel measurable function of x.*

(b) *For each $x \in \mathbf{R}$*

$$E^x[H \circ \theta_s | \mathcal{F}^B_{s+}](\omega) = E^{B_s(\omega)}[H] \quad \text{for } P^x\text{-almost every } \omega. \tag{2.36}$$

In particular, the family $\{P^x\}$ on C satisfies Definition 2.22 of a Markov process with respect to the filtration $\{\mathcal{F}^B_{t+}\}$.

Proof. Part (a). Suppose we knew that $x \mapsto P^x(F)$ is measurable for each closed set $F \subseteq C$. Then the π-λ Theorem B.3 implies that $x \mapsto P^x(A)$ is measurable for each $A \in \mathcal{B}_C$ (fill in the details for this claim as an exercise). Since linear combinations and limits preserve measurability, it follows that $x \mapsto E^x[H]$ is measurable for any bounded $\mathcal{B}_C$-measurable function H.

To show that $x \mapsto P^x(F)$ is measurable for each closed set F, consider first a bounded continuous function H on C. (Recall that C is metrized by the metric (2.24) of uniform continuity on compact intervals.) If $x_j \to x$ in $\mathbf{R}$, then by continuity and dominated convergence,

$$E^{x_j}[H] = E^0[H(x_j + B)] \longrightarrow E^0[H(x + B)] = E^x[H]$$

so $E^x[H]$ is continuous in x, which makes it Borel measurable. The indicator function $\mathbf{1}_F$ of a closed set can be written as a bounded pointwise limit of continuous functions H_n (see (B.2) in the appendix). So it follows that

$$P^x(F) = \lim_{n\to\infty} E^x[H_n]$$

is also Borel measurable in x.

Part (b). We can write the shifted process as $\theta_s B = B_s + Y$ where $Y_t = B_{s+t} - B_s$. Let Z be a bounded $\mathcal{F}^B_{s+}$-measurable random variable. By

Lemma 2.29(b), Y is a standard Brownian motion, independent of (Z, B_s) because the latter pair is $\mathcal{F}^B_{s+}$-measurable. Consequently

$$\begin{aligned} E^x\big[Z \cdot H(\theta_s B)\big] &= E^x[Z \cdot H(B_s + Y)] \\ &= \int_C E^x\big[Z \cdot H(B_s + \zeta)\big]\, P^0(d\zeta). \end{aligned}$$

By independence the expectation over Y can be separated from the expectation over (Z, B_s). (This was justified in Exercise 1.14.) P^0 is the distribution of Y because Y is a standard Brownian motion. Next move the $P^0(d\zeta)$ integral back inside, and observe that

$$\int_C H(y + \zeta)\, P^0(d\zeta) = E^y[H]$$

for any point y, including $y = B_s(\omega)$. This gives

$$E^x\big[Z \cdot H(\theta_s B)\big] = E^x\big[Z \cdot E^{B_s}(H)\big].$$

The proof is complete. □

Proposition 2.32. *Let H be a bounded $\mathcal{B}_C$-measurable function on C. Then for any $x \in \mathbf{R}$ and $0 \le s < \infty$,*

(2.37) $$E^x[H|\mathcal{F}^B_{s+}] = E^x[H|\mathcal{F}^B_s] \quad P^x\text{-almost surely.}$$

Proof. Suppose first H is of the type

$$H(\omega) = \prod_{i=1}^n \mathbf{1}_{A_i}(\omega(t_i))$$

for some $0 \le t_1 < t_2 < \cdots < t_n$ and $A_i \in \mathcal{B}_{\mathbf{R}}$. By separating those factors where $t_i \le s$, we can write $H = H_1 \cdot (H_2 \circ \theta_s)$ where H_1 is $\mathcal{F}^B_s$-measurable. Then

$$E^x[H|\mathcal{F}^B_{s+}] = H_1 \cdot E^x[H_2 \circ \theta_s|\mathcal{F}^B_{s+}] = H_1 \cdot E^{B_s}[H_2]$$

which is $\mathcal{F}^B_s$-measurable. Since $\mathcal{F}^B_{s+}$ contains $\mathcal{F}^B_s$, (2.37) follows from property (viii) of conditional expectations given in Theorem 1.26.

Let $\mathcal{H}$ be the collection of bounded functions H for which (2.37) holds. By the linearity and the monotone convergence theorem for conditional expectations (Theorem B.14), $\mathcal{H}$ satisfies the hypotheses of Theorem B.4. For the π-system $\mathcal{S}$ needed for Theorem B.4 take the class of events of the form

$$\{\omega : \omega(t_1) \in A_1, \ldots, \omega(t_n) \in A_n\}$$

for $0 \le t_1 < t_2 < \cdots < t_n$ and $A_i \in \mathcal{B}_{\mathbf{R}}$. We checked above that indicator functions of these sets lie in $\mathcal{H}$. Furthermore, these sets generate $\mathcal{B}_C$ because $\mathcal{B}_C$ is generated by coordinate projections. By Theorem B.4 $\mathcal{H}$ contains all bounded $\mathcal{B}_C$-measurable functions. □

Corollary 2.33. *If $A \in \mathcal{F}^B_{t+}$ then there exists $B \in \mathcal{F}^B_t$ such that $P^x(A \triangle B) = 0$.*

Proof. Let $Y = E^x(\mathbf{1}_A | \mathcal{F}^B_t)$, and $B = \{Y = 1\} \in \mathcal{F}^B_t$. The event $A \triangle B$ is contained in $\{\mathbf{1}_A \neq Y\}$, because $\omega \in A \setminus B$ implies $\mathbf{1}_A(\omega) = 1 \neq Y(\omega)$, while $\omega \in B \setminus A$ implies $Y(\omega) = 1 \neq 0 = \mathbf{1}_A(\omega)$. By (2.37), $\mathbf{1}_A = Y$ P^x-almost surely. Hence $\{\mathbf{1}_A \neq Y\}$, and thereby $A \triangle B$, is a P^x-null event. □

Corollary 2.34. (Blumenthal's 0–1 Law) *Let $x \in \mathbf{R}$. Then for $A \in \mathcal{F}^B_{0+}$, $P^x(A)$ is 0 or 1.*

Proof. The σ-algebra $\mathcal{F}^B_0$ satisfies the 0–1 law under P^x, because $P^x\{B_0 \in G\} = \mathbf{1}_G(x)$. Then every P^x-conditional expectation with respect to $\mathcal{F}^B_0$ equals the expectation (Exercise 1.17). The following equalities are valid P^x-almost surely for $A \in \mathcal{F}^B_{0+}$:

$$\mathbf{1}_A = E^x(\mathbf{1}_A | \mathcal{F}^B_{0+}) = E^x(\mathbf{1}_A | \mathcal{F}^B_0) = P^x(A).$$

Thus there must exist points $\omega \in C$ such that $\mathbf{1}_A(\omega) = P^x(A)$, and so the only possible values for $P^x(A)$ are 0 and 1. □

From the 0–1 law we get a fact that suggests something about the fast oscillation of Brownian motion: if it starts at the origin, then in any nontrivial time interval $(0, \varepsilon)$ the process is both positive and negative, and hence by continuity also zero. To make this precise, define

$$\begin{aligned} \sigma = \inf\{t > 0 : B_t > 0\}, \ \tau = \inf\{t > 0 : B_t < 0\}, \\ \text{and } \ T_0 = \inf\{t > 0 : B_t = 0\}. \end{aligned} \tag{2.38}$$

Corollary 2.35. *P^0-almost surely $\sigma = \tau = T_0 = 0$.*

Proof. To see that the event $\{\sigma = 0\}$ lies in $\mathcal{F}^B_{0+}$, write

$$\{\sigma = 0\} = \bigcap_{m=n}^{\infty} \{B_q > 0 \text{ for some rational } q \in (0, \tfrac{1}{m})\} \in \mathcal{F}^B_{n^{-1}}.$$

Since this is true for every $n \in \mathbf{N}$, $\{\sigma = 0\} \in \bigcap_{n \in \mathbf{N}} \mathcal{F}^B_{n^{-1}} = \mathcal{F}^B_{0+}$. Same argument shows $\{\tau = 0\} \in \mathcal{F}^B_{0+}$.

Since each variable B_t is a centered Gaussian,

$$P^0\{\sigma \le \tfrac{1}{m}\} \ge P^0\{B_{1/m} > 0\} = \tfrac{1}{2}$$

and so

$$P^0\{\sigma = 0\} = \lim_{m \to \infty} P^0\{\sigma \le \tfrac{1}{m}\} \ge \tfrac{1}{2}.$$

The convergence of the probability happens because the events $\{\sigma \le \frac{1}{m}\}$ shrink down to $\{\sigma = 0\}$ as $m \to \infty$. By Blumenthal's 0–1 law, $P^0\{\sigma = 0\} = 0$ or 1, so this quantity has to be 1. Again, the same argument for τ.

Finally, fix ω so that $\sigma(\omega) = \tau(\omega) = 0$. Then there exist $s, t \in (0, \varepsilon)$ such that $B_s(\omega) < 0 < B_t(\omega)$. By continuity, $B_u(\omega) = 0$ for some u between s and t. Hence $T_0(\omega) < \varepsilon$. Since $\varepsilon > 0$ can be taken arbitrarily small, $T_0(\omega) = 0$. □

Identity (2.36) verified that Brownian motion is a Markov process, also under the larger filtration $\mathcal{G}_t = \mathcal{F}^B_{t+}$. The transition probability of Brownian motion is a normal distribution and consequently has a density function. Namely,

$$p(t, x, A) = \int_A p(t, x, y)\, dy \qquad \text{for } A \text{ in } \mathcal{B}_{\mathbf{R}}$$

with

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\Big\{-\frac{(x-y)^2}{2t}\Big\}. \tag{2.39}$$

This transition probability density of Brownian motion is called the *Gaussian kernel*. An important analytic fact about the Gaussian kernel is that it gives the fundamental solution of the *heat equation* $\rho_t = \frac{1}{2}\rho_{xx}$. (See Section 9.2 for an explanation of this.)

Next we strengthen the Markov property of Brownian motion to the strong Markov property. Recall the definition from (2.30).

Proposition 2.36. *Brownian motion is a Feller process, and consequently a strong Markov process under the filtration* $\{\mathcal{G}_t\}$.

Proof. It only remains to observe the Feller property: for $g \in C_b(\mathbf{R})$

$$E^x[g(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} g(x+y) \exp\Big\{-\frac{y^2}{2t}\Big\}\, dx$$

and the continuity as a function of x is clear by dominated convergence. □

A natural way to understand the strong Markov property of Brownian motion is that, on the event $\{\tau < \infty\}$, the process $\widetilde{B}_t = B_{\tau+t} - B_\tau$ is a standard Brownian motion, independent of $\mathcal{G}_\tau$. Formally we can extract this point from the statement of the strong Markov property as follows. Given a bounded measurable function g on the path space C, define h by $h(\omega) = g(\omega - \omega(0))$. Then

$$E^x[g(\widetilde{B}) \,|\, \mathcal{G}_\tau](\omega) = E^x[h \circ \theta_\tau \,|\, \mathcal{G}_\tau](\omega) = E^{\omega(\tau)}[h] = E^0[g].$$

The last equality comes from the definition of the measures P^x:

$$E^x[h] = E^0[h(x+B)] = E^0[g(x+B-x)] = E^0[g].$$

The Markov and strong Markov properties are valid also for d-dimensional Brownian motion. The definitions and proofs are straightforward extensions.

Continuing still in 1 dimension, a favorite application of the strong Markov property of Brownian motion is the *reflection principle*, which is useful for calculations. Define the *running maximum* of Brownian motion by

$$M_t = \sup_{0 \le s \le t} B_s \tag{2.40}$$

Proposition 2.37. (Reflection principle) *Let $a \le b$ and $b > 0$ be real numbers. Then*

$$P^0(B_t \le a, M_t \ge b) = P^0(B_t \ge 2b - a). \tag{2.41}$$

Each inequality in the statement above can be either strict or weak, by taking limits (Exercise 2.23) and because B_t has a continuous distribution.

Proof. Let

$$\tau_b = \inf\{t \ge 0 : B_t = b\}$$

be the hitting time of point b. By path continuity $M_t \ge b$ is equivalent to $\tau_b \le t$. In the next calculation ω indicates quantities that are random for the outer expectation but constant for the inner probability.

$$\begin{aligned} P^0(B_t(\omega) \le a, M_t(\omega) \ge b) &= P^0(\tau_b(\omega) \le t, B_t(\omega) \le a) \\ &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} P^0(B_t \le a \,|\, \mathcal{F}_{\tau_b})(\omega)\big] \\ &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} P^b(B_{t-\tau_b(\omega)} \le a)\big] \\ &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} P^b(B_{t-\tau_b(\omega)} \ge 2b - a)\big] \\ &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} P^0(B_t \ge 2b - a \,|\, \mathcal{F}_{\tau_b})(\omega)\big] \\ &= P^0(M_t(\omega) \ge b, B_t(\omega) \ge 2b - a) = P^0(B_t(\omega) \ge 2b - a). \end{aligned}$$

On the third and the fourth line $\tau_b(\omega)$ appears in two places, and it is a constant in the inner probability. Then comes the reflection: by symmetry, Brownian motion started at b is equally likely to reside below a as above $b + (b - a) = 2b - a$. The last equality drops the condition on M_t that is now superfluous because $2b - a \ge b$.

The reader may feel a little uncomfortable about the cavalier way of handling the strong Markov property. Let us firm it up by introducing explicitly $Y(s, \omega)$ that allows us to mechanically apply (2.30):

$$Y(s, \omega) = \mathbf{1}\{s \le t, \omega(t - s) \ge 2b - a\} - \mathbf{1}\{s \le t, \omega(t - s) \le a\}.$$

By symmetry of Brownian motion, for any $s \le t$,

$$\begin{aligned} E^b[Y(s,B)] &= P^b\{B_{t-s} \ge 2b-a\} - P^b\{B_{t-s} \le a\} \\ &= P^0\{B_{t-s} \ge b-a\} - P^0\{B_{t-s} \le a-b\} = 0. \end{aligned}$$

In the next calculation, begin with this equality, note that $b = B_{\tau_b}(\omega)$ and then apply the strong Markov property. Use X as the identity mapping on C in the first innermost expectation.

$$\begin{aligned} 0 &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} E^b\{Y(\tau_b(\omega), X)\}\big] \\ &= E^0\big[\mathbf{1}\{\tau_b(\omega) \le t\} E^0\{Y(\tau_b, \theta_{\tau_b} B) \,|\, \mathcal{F}_{\tau_b}\}(\omega)\big] \\ &= E^0\big[\mathbf{1}\{\tau_b \le t\} Y(\tau_b, \theta_{\tau_b} B)\big] \\ &= P^0\{\tau_b \le t, B_t \ge 2b-a\} - P^0\{\tau_b \le t, B_t \le a\} \\ &= P^0\{B_t \ge 2b-a\} - P^0\{\tau_b \le t, B_t \le a\}. \end{aligned}$$

The second last step used

$$Y(\tau_b, \theta_{\tau_b} B) = \mathbf{1}\{\tau_b \le t, B_t \ge 2b-a\} - \mathbf{1}\{\tau_b \le t, B_t \le a\}$$

which is a direct consequence of the definition of Y. $\square$

An immediate corollary of (2.41) is that M_t has the same distribution as $|B_t|$: taking $a = b > 0$

$$\begin{aligned} P^0(M_t \ge b) &= P^0(B_t \le b, M_t \ge b) + P^0(B_t > b, M_t \ge b) \\ &= P^0(B_t \ge b) + P^0(B_t > b) = 2P^0(B_t \ge b) = P^0(|B_t| \ge b). \end{aligned} \tag{2.42}$$

With a little more effort one can write down the joint density of (B_t, M_t) (Exercise 2.24).

2.4.2. Path regularity of Brownian motion.

As a byproduct of the construction of Brownian motion in Section B.2 we obtained Hölder continuity of paths with any exponent strictly less than $\frac{1}{2}$.

Theorem 2.38. *Fix $0 < \gamma < \frac{1}{2}$. The following is true almost surely for Brownian motion: for every $T < \infty$ there exists a finite constant $C(\omega)$ such that*

$$|B_t(\omega) - B_s(\omega)| \le C(\omega)|t-s|^\gamma \quad \textit{for all } 0 \le s,t \le T. \tag{2.43}$$

Next we prove a result from the opposite direction. Namely, for an exponent strictly larger than $\frac{1}{2}$ there is not even local Hölder continuity. ("Local" here means that the property holds in a small enough interval around a given point.)

Theorem 2.39. *Let B be a Brownian motion. For finite positive reals γ, C, and ε define the event*

$$G(\gamma, C, \varepsilon) = \{\text{there exists } s \in \mathbf{R}_+ \text{ such that } |B_t - B_s| \le C|t-s|^\gamma \\ \text{for all } t \in [s-\varepsilon, s+\varepsilon]\}.$$

Then if $\gamma > \frac{1}{2}$, $P\big(G(\gamma, C, \varepsilon)\big) = 0$ for all positive C and ε.

Proof. Fix $\gamma > \frac{1}{2}$. Since only increments of Brownian motion are involved, we can assume that the process in question is a standard Brownian motion. (B_t and $\widetilde{B}_t = B_t - B_0$ have the same increments.) In the proof we want to deal only with a bounded time interval. So define

$$H_k(C, \varepsilon) = \{\text{there exists } s \in [k, k+1] \text{ such that } |B_t - B_s| \le C|t-s|^\gamma \\ \text{for all } t \in [s-\varepsilon, s+\varepsilon] \cap [k, k+1]\}.$$

$G(\gamma, C, \varepsilon)$ is contained in $\bigcup_k H_k(C, \varepsilon)$, so it suffices to show $P\big(H_k(C, \varepsilon)\big) = 0$ for all k. Since $Y_t = B_{k+t} - B_k$ is a standard Brownian motion, $P\big(H_k(C, \varepsilon)\big) = P\big(H_0(C, \varepsilon)\big)$ for each k. Finally, what we show is $P\big(H_0(C, \varepsilon)\big) = 0$.

Fix $m \in \mathbf{N}$ such that $m(\gamma - \frac{1}{2}) > 1$. Let $\omega \in H_0(C, \varepsilon)$, and pick $s \in [0, 1]$ so that the condition of the event is satisfied. Consider n large enough so that $m/n < \varepsilon$. Imagine partitioning $[0, 1]$ into intervals of length $\frac{1}{n}$. Let

$$X_{n,k} = \max\{|B_{(j+1)/n} - B_{j/n}| : k \le j \le k+m-1\} \quad \text{for } 0 \le k \le n-m.$$

The point s has to lie in one of the intervals $[\frac{k}{n}, \frac{k+m}{n}]$, for some $0 \le k \le n-m$. For this particular k,

$$\begin{aligned} |B_{(j+1)/n} - B_{j/n}| &\le |B_{(j+1)/n} - B_s| + |B_s - B_{j/n}| \\ &\le C\big(|\tfrac{j+1}{n} - s|^\gamma + |s - \tfrac{j}{n}|^\gamma\big) \le 2C(\tfrac{m}{n})^\gamma \end{aligned}$$

for all the j-values in the range $k \le j \le k+m-1$. (Simply because the points $\frac{j}{n}$ and $\frac{j+1}{n}$ are within ε of s. Draw a picture.) In other words, $X_{n,k} \le 2C(\frac{m}{n})^\gamma$ for this k-value.

Now consider all the possible k-values, recall that Brownian increments are stationary and independent, and note that by basic Gaussian properties,

B_t has the same distribution as $t^{1/2}B_1$.

$$
\begin{aligned}
P\big(H_0(C,\varepsilon)\big) &\le \sum_{k=0}^{n-m} P\{X_{n,k} \le 2C(\tfrac{m}{n})^\gamma\} \le nP\{X_{n,0} \le 2C(\tfrac{m}{n})^\gamma\} \\
&= n\prod_{j=0}^{m-1} P\big\{\, |B_{(j+1)/n} - B_{j/n}| \le 2C(\tfrac{m}{n})^\gamma \big\} = nP\big\{\, |B_{1/n}| \le 2C(\tfrac{m}{n})^\gamma \big\}^m \\
&= nP\big\{\, |B_1| \le 2Cn^{1/2-\gamma}m^\gamma \big\}^m \\
&= n\bigg(\frac{1}{\sqrt{2\pi}} \int_{-2Cn^{1/2-\gamma}m^\gamma}^{2Cn^{1/2-\gamma}m^\gamma} e^{-x^2/2}\,dx \bigg)^m \le n\bigg(\frac{1}{\sqrt{2\pi}} 4Cn^{1/2-\gamma}m^\gamma \bigg)^m \\
&\le K(m)n^{1-m(\gamma-1/2)}.
\end{aligned}
$$

In the last stages above we bounded $e^{-x^2/2}$ above by 1, and then collected some of the constants and m-dependent quantities into the function $K(m)$. The bound is valid for all large n, while m is fixed. Thus we may let $n \to \infty$, and obtain $P\big(H_0(C,\varepsilon)\big) = 0$. This proves the theorem. □

Corollary 2.40. *The following is true almost surely for Brownian motion: the path $t \mapsto B_t(\omega)$ is not differentiable at any time point.*

Proof. Suppose $t \mapsto B_t(\omega)$ is differentiable at some point s. This means that there is a real-valued limit

$$\xi = \lim_{t\to s} \frac{B_t(\omega) - B_s(\omega)}{t-s}.$$

Thus if the integer M satisfies $M > |\xi| + 1$, we can find another integer k such that for all $t \in [s-k^{-1}, s+k^{-1}]$,

$$-M \le \frac{B_t(\omega) - B_s(\omega)}{t-s} \le M \quad \text{which implies} \quad |B_t(\omega) - B_s(\omega)| \le M|t-s|.$$

Consequently $\omega \in G(1, M, k^{-1})$.

This reasoning shows that if $t \mapsto B_t(\omega)$ is differentiable at even a single time point, then ω lies in the union $\bigcup_M \bigcup_k G(1, M, k^{-1})$. This union has probability zero by the previous theorem. □

Functions of bounded variation are differences of nondecreasing functions, and monotone functions can be differentiated at least Lebesgue–almost everywhere. Hence the above theorem implies that Brownian motion paths are of unbounded variation on every interval.

To recapitulate, the previous results show that a Brownian path is Hölder continuous with any exponent $\gamma < \frac{1}{2}$ but not for any $\gamma > \frac{1}{2}$. For the sake of completeness, here is the precise result. Proofs can be found for example in [**9**, **14**].

Theorem 2.41. (Lévy's modulus of continuity.) *Almost surely,*

$$\varlimsup_{\delta\searrow 0} \sup_{\substack{0\le s<t\le 1\\ t-s\le\delta}} \frac{B_t - B_s}{\sqrt{2\delta\log\delta^{-1}}} = 1.$$

Next we show that Brownian motion has finite *quadratic variation* $[B]_t = t$. Quadratic variation was introduced in Section 2.2. This notion occupies an important role in stochastic analysis and will be discussed more in the martingale chapter. As a corollary we get another proof of the unbounded variation of Brownian paths, a proof that does not require knowledge about the differentiation properties of BV functions.

Recall again the notion of the mesh $\text{mesh}(\pi) = \max_i(t_{i+1} - t_i)$ of a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{m(\pi)} = t\}$ of $[0, t]$.

Proposition 2.42. *Let B be a Brownian motion. For any partition π of $[0, t]$,*

$$E\biggl[\Bigl(\sum_{i=0}^{m(\pi)-1}(B_{t_{i+1}} - B_{t_i})^2 - t\Bigr)^2\biggr] \le 2t\,\text{mesh}(\pi). \tag{2.44}$$

In particular

$$\lim_{\text{mesh}(\pi)\to 0} \sum_{i=0}^{m(\pi)-1}(B_{t_{i+1}} - B_{t_i})^2 = t \quad \text{in } L^2(P). \tag{2.45}$$

If we have a sequence of partitions π^n such that $\sum_n \text{mesh}(\pi^n) < \infty$, then the convergence above holds almost surely along this sequence.

Proof. Straightforward computation, utilizing the facts that Brownian increments are independent, $B_s - B_r$ has mean zero normal distribution with variance $s - r$, and so its fourth moment is $E[(B_s - B_r)^4] = 3(s - r)^2$. Let $\Delta t_i = t_{i+1} - t_i$.

$$\begin{aligned}
E\biggl[\Bigl(\sum_{i=0}^{m(\pi)-1}(B_{t_{i+1}} - B_{t_i})^2 - t\Bigr)^2\biggr] &= \sum_i E\bigl[(B_{t_{i+1}} - B_{t_i})^4\bigr] \\
&\quad + \sum_{i\ne j} E\bigl[(B_{t_{i+1}} - B_{t_i})^2(B_{t_{j+1}} - B_{t_j})^2\bigr] - 2t\sum_i E\bigl[(B_{t_{i+1}} - B_{t_i})^2\bigr] + t^2 \\
&= 3\sum_i(\Delta t_i)^2 + \sum_{i\ne j}\Delta t_i\cdot\Delta t_j - 2t^2 + t^2 = 2\sum_i(\Delta t_i)^2 + \sum_{i,j}\Delta t_i\cdot\Delta t_j - t^2 \\
&= 2\sum_i(\Delta t_i)^2 \le 2\,\text{mesh}(\pi)\sum_i\Delta t_i = 2\,\text{mesh}(\pi)t.
\end{aligned}$$

By Chebychev's inequality,

$$\begin{aligned}P\bigg\{\,\bigg|\sum_{i=0}^{m(\pi^n)-1}(B_{t^n_{i+1}}-B_{t^n_i})^2-t\bigg|\ge\varepsilon\bigg\}\\ &\le\varepsilon^{-2}E\bigg[\bigg(\sum_{i=0}^{m(\pi^n)-1}(B_{t^n_{i+1}}-B_{t^n_i})^2-t\bigg)^2\bigg]\\ &\le 2t\varepsilon^{-2}\,\mathrm{mesh}(\pi^n).\end{aligned}$$

If $\sum_n \mathrm{mesh}(\pi^n)<\infty$, these numbers have a finite sum over n (in short, they are *summable*). Hence the asserted convergence follows from the Borel-Cantelli Lemma. □

Corollary 2.43. *The following is true almost surely for a Brownian motion B: the path $t\mapsto B_t(\omega)$ is not a member of $BV[0,T]$ for any $0<T<\infty$.*

Proof. Pick an ω such that

$$\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\big(B_{(i+1)T/2^n}(\omega)-B_{iT/2^n}(\omega)\big)^2=T$$

for each $T=k^{-1}$ for $k\in\mathbf{N}$. Such ω's form a set of probability 1 by the previous proposition, because the partitions $\{iT2^{-n}:0\le i\le 2^n\}$ have meshes 2^{-n} that form a summable sequence. Furthermore, by almost sure continuity, we can assume that

$$\lim_{n\to\infty}\max_{0\le i\le 2^n-1}\big|B_{(i+1)T/2^n}(\omega)-B_{iT/2^n}(\omega)\big|=0$$

for each $T=k^{-1}$. (Recall that a continuous function is uniformly continuous on a closed, bounded interval.) And now for each such T,

$$\begin{aligned}T&=\lim_{n\to\infty}\sum_{i=0}^{2^n-1}\big(B_{(i+1)T/2^n}(\omega)-B_{iT/2^n}(\omega)\big)^2\\ &\le\lim_{n\to\infty}\bigg\{\max_{0\le i\le 2^n-1}\big|B_{(i+1)T/2^n}(\omega)-B_{iT/2^n}(\omega)\big|\bigg\}\\ &\qquad\times\sum_{i=0}^{2^n-1}\big|B_{(i+1)T/2^n}(\omega)-B_{iT/2^n}(\omega)\big|.\end{aligned}$$

Since the maximum in braces vanishes as $n\to\infty$, the last sum must converge to ∞. Consequently the path $t\mapsto B_t(\omega)$ is not BV in any interval $[0,k^{-1}]$. Any other nontrivial inteval $[0,T]$ contains an interval $[0,k^{-1}]$ for some k, and so this path cannot have bounded variation on any interval $[0,T]$. □

2.5. Poisson processes

Poisson processes describe configurations of random points. The definition and construction for a general state space is no more complex than for the real line, so we define and construct the Poisson point process on an abstract measure space first.

Let $0 < \alpha < \infty$. A nonnegative integer-valued random variable X has *Poisson distribution with parameter* α (*Poisson*(α)*–distribution*) if

$$P\{X = k\} = e^{-\alpha}\frac{\alpha^k}{k!} \quad \text{for } k \in \mathbf{Z}_+.$$

To describe point processes we also need the extreme cases: a Poisson variable with parameter $\alpha = 0$ is identically zero, or $P\{X = 0\} = 1$, while a Poisson variable with parameter $\alpha = \infty$ is identically infinite: $P\{X = \infty\} = 1$. A Poisson(α) variable has mean and variance α. A sum of independent Poisson variables (including a sum of countably infinitely many terms) is again Poisson distributed. These properties make the next definition possible.

Definition 2.44. Let $(S, \mathcal{A}, \mu)$ be a σ-finite measure space. A process $\{N(A) : A \in \mathcal{A}\}$ indexed by the measurable sets is a *Poisson point process with mean measure* μ if

- (i) Almost surely, $N(\cdot)$ is a $\mathbf{Z} \cup \{\infty\}$-valued measure on $(S, \mathcal{A})$.
- (ii) $N(A)$ is Poisson distributed with parameter $\mu(A)$.
- (iii) For any pairwise disjoint $A_1, A_2, \dots, A_n \in \mathcal{A}$, the random variables $N(A_1)$, $N(A_2)$, $\dots$, $N(A_n)$ are independent.

The interpretation is that $N(A)$ is the number of points in the set A. N is also called a *Poisson random measure.*

Observe that items (i) and (ii) give a complete description of all the finite-dimensional distributions of $\{N(A)\}$. For arbitrary $B_1, B_2, \dots, B_m \in \mathcal{A}$, we can find disjoint $A_1, A_2, \dots, A_n \in \mathcal{A}$ so that each B_j is a union of some of the A_i's. Then each $N(B_j)$ is a certain sum of $N(A_i)$'s, and we see that the joint distribution of $N(B_1)$, $N(B_2)$, $\dots$, $N(B_m)$ is determined by the joint distribution of $N(A_1)$, $N(A_2)$, $\dots$, $N(A_n)$.

Proposition 2.45. *Let $(S, \mathcal{A}, \mu)$ be a σ-finite measure space. Then a Poisson point process $\{N(A) : A \in \mathcal{A}\}$ with mean measure μ exists.*

Proof. Let $S_1, S_2, S_3, \dots$ be disjoint measurable sets such that $S = \bigcup S_i$ and $\mu(S_i) < \infty$. We shall first define a Poisson point process N_i supported on the subset S_i (this means that N_i has no points outside S_i). If $\mu(S_i) = 0$, define $N_i(A) = 0$ for every measurable set $A \in \mathcal{A}$.

With this trivial case out of the way, we may assume $0 < \mu(S_i) < \infty$. Let $\{X^i_j : j \in \mathbf{N}\}$ be i.i.d. S_i-valued random variables with common probability distribution

$$P\{X^i_j \in B\} = \frac{\mu(B \cap S_i)}{\mu(S_i)} \quad \text{for measurable sets } B \in \mathcal{A}.$$

Independently of the $\{X^i_j : j \in \mathbf{N}\}$, let K_i be a Poisson($\mu(S_i)$) random variable. Define

$$N_i(A) = \sum_{j=1}^{K_i} \mathbf{1}_A(X^i_j) \quad \text{for measurable sets } A \in \mathcal{A}.$$

As the formula reveals, K_i decides how many points to place in S_i, and the $\{X^i_j\}$ give the locations of the points in S_i. We leave it as an exercise to check that N_i is a Poisson point process whose mean measure is μ restricted to S_i, defined by $\mu_i(B) = \mu(B \cap S_i)$.

We can repeat this construction for each S_i, and take the resulting random processes N_i mutually independent by a suitable product space construction. Finally, define

$$N(A) = \sum_i N_i(A).$$

Again, we leave checking the properties as an exercise. □

The most important Poisson processes are those on Euclidean spaces whose mean measure is a constant multiple of Lebesgue measure. These are called homogeneous Poisson point processes. When the points lie on the positive real line, they naturally acquire a temporal interpretation. For this case we make the next definition.

Definition 2.46. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{\mathcal{F}_t\}$ a filtration on it, and $\alpha > 0$. A *(homogeneous) Poisson process with rate* α is an adapted stochastic process $N = \{N_t : 0 \le t < \infty\}$ with these properties.

(i) $N_0 = 0$ almost surely.

(ii) For almost every ω, the path $t \mapsto N_t(\omega)$ is cadlag.

(iii) For $0 \le s < t$, $N_t - N_s$ is independent of $\mathcal{F}_s$, and has Poisson distribution with parameter $\alpha(t - s)$.

Proposition 2.47. *Homogeneous Poisson processes on* $[0, \infty)$ *exist.*

Proof. Let $\{N(A) : A \in \mathcal{B}_{(0,\infty)}\}$ be a Poisson point process on $(0, \infty)$ with mean measure αm. Define $N_0 = 0$, $N_t = N(0, t]$ for $t > 0$, and $\mathcal{F}^N_t = \sigma\{N_s : 0 \le s \le t\}$. Let $0 = s_0 < s_1 < \cdots < s_n \le s < t$. Then

$$N(s_0, s_1], N(s_1, s_2], \ldots, N(s_{n-1}, s_n], N(s, t]$$

are independent random variables, from which follows that the vector $(N_{s_1}, \ldots, N_{s_n})$ is independent of $N(s,t]$. Considering all such n-tuples (for various n) while keeping $s < t$ fixed shows that $\mathcal{F}_s$ is independent of $N(s,t] = N_t - N_s$.

The cadlag path property of N_t follows from properties of the Poisson process. Almost every ω has the property that $N(0,T] < \infty$ for all $T < \infty$. Given such an ω and t, there exist $t_0 < t < t_1$ such that $N(t_0, t) = N(t, t_1) = 0$. (There may be a point at t, but there cannot be sequences of points converging to t from either left or right.) Consequently N_s is constant for $t_0 < s < t$ and so the left limit N_{t-} exists. Also $N_s = N_t$ for $t \le s < t_1$ which gives the right continuity at t. □

One can show that the jumps of N_t are all of size 1 (Exercise 2.27). This is because the Lebesgue mean measure does not let two Poisson points sit on top of each other. The next lemma is proved just like its counterpart for Brownian motion, so we omit its proof.

Proposition 2.48. *Suppose $N = \{N_t\}$ is a homogeneous Poisson process with respect to a filtration $\{\mathcal{F}_t\}$ on $(\Omega, \mathcal{F}, P)$.*

(a) *N is a Poisson process also with respect to the augmented right-continuous filtration $\{\bar{\mathcal{F}}_{t+}\}$.*

(b) *Define $Y_t = N_{s+t} - N_s$ and $\mathcal{G}_t = \mathcal{F}_{(s+t)+}$. Then $Y = \{Y_t : 0 \le t < \infty\}$ is a homogeneous Poisson process with respect to the filtration $\{\mathcal{G}_t\}$, and independent of $\bar{\mathcal{F}}_{s+}$.*

Since the Poisson process is monotone nondecreasing it cannot be a martingale. We need to compensate by subtracting off the mean, and so we define the *compensated Poisson process* as

$$M_t = N_t - \alpha t.$$

Proposition 2.49. *M is a martingale.*

Proof. Follows from the independence of increments.

$$E[N_t|\mathcal{F}_s] = E[N_t - N_s|\mathcal{F}_s] + E[N_s|\mathcal{F}_s] = \alpha(t-s) + N_s. \qquad \square$$

The Markov property of the Poisson process follows next just like for Brownian motion. However, we should not take $\mathbf{R}$ as the state space, but instead $\mathbf{Z}$ (or alternatively $\mathbf{Z}_+$). A Poisson process with initial state $x \in \mathbf{Z}$ would be defined as $x + N_t$, where N is the process defined in Definition 2.46, and P^x would be the distribution of $\{x + N_t\}_{t \in \mathbf{R}_+}$ on the space $D_{\mathbf{Z}}[0,\infty)$ of $\mathbf{Z}$-valued cadlag paths. Because the state space is discrete the Feller property is automatically satisfied. (A *discrete space* is one where singleton sets $\{x\}$ are open. Every function on a discrete space is continuous.) Consequently homogeneous Poisson processes also satisfy the strong Markov property. The semigroup for a rate α Poisson process is

$$E^x[g(X_t)] = E^0[g(x+X_t)] = \sum_{k=0}^{\infty} g(x+k)e^{-\alpha}\frac{\alpha^k}{k!} \qquad \text{for } x \in \mathbf{Z}.$$

The reader should also be aware of another natural construction of N in terms of waiting times. We refer to [13, Section 4.8] for a proof.

Proposition 2.50. *Let $\{T_k : 1 \le k < \infty\}$ be i.i.d. rate α exponential random variables. Let $S_n = T_1 + \cdots + T_n$ for $n \ge 1$. Then $N(A) = \sum_n \mathbf{1}_A(S_n)$ defines a homogeneous rate α Poisson point process on $\mathbf{R}_+$, and $N_t = \max\{n : S_n \le t\}$ defines a rate α Poisson process with respect to its own filtration $\{\mathcal{F}_t^N\}$ with initial point $N_0 = 0$ a.s.*

Exercises

Exercise 2.1. To see that even increasing unions of σ-algebras can fail to be σ-algebras (so that the generation is necessary in (2.1)), look at this example. Let $\Omega = (0, 1]$, and for $n \in \mathbf{N}$ let $\mathcal{F}_n$ be the σ-algebra generated by the intervals $\{(k2^{-n}, (k+1)2^{-n}] : k \in \{0, 1, 2, \dots, 2^n - 1\}\}$. How about the intersection of the sets $(1 - 2^{-n}, 1]$?

Exercise 2.2. To ward off yet another possible pitfall: uncountable unions must be avoided. Even if $A = \bigcup_{t \in \mathbf{R}_+} A_t$ and each $A_t \in \mathcal{F}_t$, A may fail to be a member of $\mathcal{F}_\infty$. Here is an example. Let $\Omega = \mathbf{R}_+$ and let $\mathcal{F}_t = \mathcal{B}_{\mathbf{R}_+}$ for each $t \in \mathbf{R}_+$. Then also $\mathcal{F}_\infty = \mathcal{B}_{\mathbf{R}_+}$. Pick a subset A of $\mathbf{R}_+$ that is not a Borel subset. (The proof that such sets exist needs to be looked up from an analysis text.) Then take $A_t = \{t\}$ if $t \in A$ and $A_t = \emptyset$ otherwise.

Exercise 2.3. Let $\{\mathcal{F}_t\}$ be a filtration, and let $\mathcal{G}_t = \mathcal{F}_{t+}$. Show that $\mathcal{G}_{t-} = \mathcal{F}_{t-}$ for $t > 0$.

Exercise 2.4. Assume the probability space $(\Omega, \mathcal{F}, P)$ is complete. Let $\{\mathcal{F}_t\}$ be a filtration, $\mathcal{G}_t = \mathcal{F}_{t+}$ its right-continuous version, and $\mathcal{H}_t = \bar{\mathcal{F}}_t$ its augmentation. Augment $\{\mathcal{G}_t\}$ to get the filtration $\{\bar{\mathcal{G}}_t\}$, and define also $\mathcal{H}_{t+} = \bigcap_{s:s>t} \mathcal{H}_s$. Show that $\bar{\mathcal{G}}_t = \mathcal{H}_{t+}$. In other words, it is immaterial whether we augment before or after making the filtration right-continuous.

Hints. $\bar{\mathcal{G}}_t \subseteq \mathcal{H}_{t+}$ should be easy. For the other direction, if $C \in \mathcal{H}_{t+}$, then for each $s > t$ there exists $C_s \in \mathcal{F}_s$ such that $P(C \triangle C_s) = 0$. For any sequence $s_i \searrow t$, the set $\widetilde{C} = \bigcap_{m \ge 1} \bigcup_{i \ge m} C_{s_i}$ lies in $\mathcal{F}_{t+}$. Use Exercise 1.9.

Exercise 2.5. Let the underlying probability space be $\Omega = [0, 1]$ with P given by Lebesgue measure. Define two processes

$$X_t(\omega) = 0 \quad \text{and} \quad Y_t(\omega) = \mathbf{1}_{\{t = \omega\}}.$$

X is obviously continuous. Show that Y does not have a single continuous path, but X and Y are modifications of each other.

Exercise 2.6. (Example of an adapted but not progressively measurable process.) Let $\Omega = [0,1]$, and for each t let $\mathcal{F}_t$ be the σ-field generated by singletons on Ω. (Equivalently, $\mathcal{F}_t$ consists of all countable sets and their complements.) Let $X_t(\omega) = \mathbf{1}\{\omega = t\}$. Then $\{X_t : 0 \le t \le 1\}$ is adapted. But X on $[0,1] \times \Omega$ is not $\mathcal{B}_{[0,1]} \otimes \mathcal{F}_1$-measurable.

Hint. Show that elements of $\mathcal{B}_{[0,1]} \otimes \mathcal{F}_1$ are of the type

$$\Big(\bigcup_{s \in I} B_s \times \{s\} \Big) \cup \big(H \times I^c \big)$$

where I is a countable subset of Ω, each $B_t \in \mathcal{B}_{[0,1]}$, and H is either empty or $[0,1]$. Consequently the diagonal $\{(t,\omega) : X_t(\omega) = 1\}$ is not an element of $\mathcal{B}_{[0,1]} \otimes \mathcal{F}_1$.

Exercise 2.7. Let τ be a stopping time and fix $t \in \mathbf{R}_+$. Let $A \in \mathcal{F}_t$ satisfy $A \subseteq \{\tau \ge t\}$. Show that then $A \in \mathcal{F}_\tau$.

You might see this type of property expressed as $\mathcal{F}_t \cap \{\tau \ge t\} \subseteq \mathcal{F}_\tau$, even though strictly speaking intersecting $\mathcal{F}_t$ with $\{\tau \ge t\}$ is not legitimate. The intersection is used to express the idea that the σ-algebra $\mathcal{F}_t$ is restricted to the set $\{\tau \ge t\}$. Questionable but convenient usage is called *abuse of notation* among mathematicians. It is the kind of license that seasoned professionals can take but beginners should exercise caution!

Exercise 2.8. Show that if $A \in \mathcal{F}_\tau$, then $A \cap \{\tau < \infty\} \in \mathcal{F}_\infty$. Show that any measurable subset of $\{\tau = \infty\}$ is a member of $\mathcal{F}_\tau$.

The following rather contrived example illustrates that $\mathcal{F}_\tau$ does not have to lie inside $\mathcal{F}_\infty$. Take $\Omega = \{0,1,2\}$, $\mathcal{F} = 2^\Omega$, $\mathcal{F}_t = \{\{0\}, \{1,2\}, \emptyset, \Omega\}$ for all $t \in \mathbf{R}_+$, and $\tau(0) = 0$, $\tau(1) = \tau(2) = \infty$. Show that τ is a stopping time and find $\mathcal{F}_\tau$.

Exercise 2.9. Let σ be a stopping time and Z an $\mathcal{F}_\sigma$-measurable random variable. Show that for any $A \in \mathcal{B}_{[0,t]}$, $\mathbf{1}_{\{\sigma \in A\}} Z$ is $\mathcal{F}_t$-measurable.

Hint. Start with $A = [0,s]$. Use the π-λ theorem.

Exercise 2.10. Let $t < u$. Given that $[X,Y]_t$ and $[X,Y]_u$ satisfy the limit (2.13), show that

$$[X,Y]_u - [X,Y]_t = \lim_{\text{mesh}(\pi) \to 0} \sum_i (X_{s_{i+1}} - X_{s_i})(Y_{s_{i+1}} - Y_{s_i}) \tag{2.46}$$

where the limit is in probability and taken over partitions $\{s_i\}$ of $[t,u]$ as the mesh tends to 0.

Exercise 2.11. Suppose Y is a cadlag process. Show that while in the definition (1.14) of total variation $V_Y(t)$ the supremum can be replaced by the limit as the mesh tends to zero (Exercise 1.4), in the definition of the quadratic variation $[Y]_t$ the limit cannot in general be replaced by the supremum.

Exercise 2.12. Let $f \in BV[0,T]$ and define $g(t) = f(t+)$ for $t \in [0,T)$ and $g(T) = f(T)$. Show that g is a cadlag function.

Hint. A BV function is the difference of two nondecreasing functions (page 14).

Exercise 2.13. Let $\mathcal{F}_t = \sigma\{\omega(s) : 0 \le s \le t\}$ be the filtration generated by coordinates on the space $C = C_{\mathbf{R}}[0,\infty)$ of continuous functions. Let

$$H = \{\omega \in C : \; t \text{ is a local maximum for } \omega \;\}.$$

Show that $H \in \mathcal{F}_{t+} \setminus \mathcal{F}_t$.

Hints. To show $H \in \mathcal{F}_{t+}$, note that $\omega \in H$ iff for all large enough $n \in \mathbf{N}$, $\omega(t) \ge \omega(q)$ for all rational $q \in (t - n^{-1}, t + n^{-1})$. To show $H \notin \mathcal{F}_t$, use Exercise 1.8(b). For any $\omega \in H$ one can construct $\widetilde{\omega} \notin H$ such that $\omega(s) = \widetilde{\omega}(s)$ for $s \le t$.

Exercise 2.14. Let τ be a stopping time. Define

$$\mathcal{F}_{\tau-} = \sigma\Big\{\mathcal{F}_0 \cup \bigcup_{t>0} \big(\mathcal{F}_t \cap \{\tau > t\}\big)\Big\}. \tag{2.47}$$

See the remark in Exercise 2.7 that explains the notation.

(a) Show that for any stopping time σ, $\mathcal{F}_\sigma \cap \{\sigma < \tau\} \subseteq \mathcal{F}_{\tau-}$.

(b) Let σ_n be a nondecreasing sequence of stopping times such that $\sigma_n \nearrow \tau$ and $\sigma_n < \tau$ for each n. Show that $\mathcal{F}_{\sigma_n} \nearrow \mathcal{F}_{\tau-}$. This last convergence statement means that $\mathcal{F}_{\tau-} = \sigma(\cup_n \mathcal{F}_{\sigma_n})$.

Exercise 2.15. (a) Suppose X is a caglad process adapted to $\{\mathcal{F}_t\}$. Define $Z(t) = X(t+)$ for $0 \le t < \infty$. Show that Z is a cadlag process adapted to $\{\mathcal{F}_{t+}\}$.

(b) Show that Lemma 2.9 is valid for a caglad process under the additional assumption that $\{\mathcal{F}_t\}$ is right-continuous.

(c) Let Ω be the space of real-valued caglad paths, X the coordinate process $X_t(\omega) = \omega(t)$ on Ω, and $\{\mathcal{F}_t\}$ the filtration generated by coordinates. Show that Lemma 2.9 fails for this setting.

Exercise 2.16. Check that you are able to use the Markov property with this simple exercise. Let $\{P^x\}$ satisfy Definition 2.22. Let $r < s < t$ be time points and $A, B \in \mathcal{B}_{\mathbf{R}^d}$. Assuming that $P^x(X_r \in B, X_s = y) > 0$, show that

$$P^x(X_t \in A \,|\, X_r \in B, X_s = y) = P^y(X_{t-s} \in A)$$

with the understanding that the conditional probability on the left is defined in the elementary fashion by (1.28).

Exercise 2.17. Prove (2.27). *Hint.* Consider first functions of the type $f(x_0, x_1, \dots, x_n) = f_0(x_0)f_1(x_1)\cdots f_n(x_n)$. Condition on $\mathcal{F}_{s_{n-1}}$, use (2.26), and do induction on n. Then find a theorem in the appendix that allows you to extend this to all bounded Borel functions $f : \mathbf{R}^{d(n+1)} \to \mathbf{R}$.

Exercise 2.18. Let $\{P^x\}$ be a Feller process. Use the Markov property and the Feller property inductively to show that

$$\int_D \prod_{i=1}^m f_i(\omega(s_i))\, P^x(d\omega)$$

is a continuous function of $x \in \mathbf{R}^d$ for $f_1, \dots, f_m \in C_b(\mathbf{R}^d)$.

Exercise 2.19. Consider the coordinate Markov process on D space under probability measure P^x. Assume Feller continuity. Let τ be a finite stopping time, $A \in \mathcal{F}_\tau$ and $B \in \mathcal{B}_D$. Show that

$$P^x(A \cap \theta_\tau^{-1} B \mid X_\tau) = P^x(A \mid X_\tau)\, P^x(\,\theta_\tau^{-1} B \mid X_\tau).$$

If τ marks the present moment, this says that, *given the present, the past and the future are independent. Hint.* Use the strong Markov property and properties (vii) and (viii) from Theorem 1.26.

Exercise 2.20. Using property (ii) in Definition 2.26 show these two properties of Brownian motion, for any $0 \le s_0 < s_1 < \cdots < s_n$.

(a) The σ-algebras $\mathcal{F}_{s_0}, \sigma(B_{s_1} - B_{s_0}), \sigma(B_{s_2} - B_{s_1}), \dots, \sigma(B_{s_n} - B_{s_{n-1}})$ are independent.

(b) The distribution of the vector

$$(B_{t+s_1} - B_{t+s_0}, B_{t+s_2} - B_{t+s_1}, \dots, B_{t+s_n} - B_{t+s_{n-1}})$$

is the same for all $t \ge 0$.

Exercise 2.21. (Brownian motion as a Gaussian process.) A process $\{X_t\}$ is *Gaussian* if for all finite sets of indices $\{t_1, t_2, \dots, t_n\}$ the vector $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ has multivariate normal distribution as in Example 1.18(iv). It is a consequence of a π-λ argument or Kolmogorov's extension theorem that the distribution of a Gaussian process is entirely determined by two functions: the mean $m(t) = EX_t$ and the covariance $c(s,t) = \operatorname{Cov}(X_s, X_t) = E(X_sX_t) - m(s)m(t)$.

(a) Show that having Gaussian marginals does not imply that the joint distribution is Gaussian.

Hint. Consider this example: X is a standard normal, ξ is independent of X with distribution $P(\xi = \pm 1) = 1/2$, and $Y = \xi X$.

(b) Starting with points (ii) and (iii) of Definition 2.26 show that standard Brownian motion is a Gaussian process with $m(t) = 0$ and $c(s,t) = s \wedge t$.

Hint. A helpful observation might be that a linear transformation of a jointly Gaussian vector is also jointly Gaussian.

Exercise 2.22. Let B_t be a Brownian motion with respect to filtration $\{\mathcal{F}_t\}$ on probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{A}$ be another sub-σ-algebra of $\mathcal{F}$. Assume that $\mathcal{A}$ and $\mathcal{F}_\infty$ are independent. Let $\mathcal{G}_t = \sigma(\mathcal{F}_t, \mathcal{A})$. That is, $\mathcal{G}_t$ is the smallest σ-algebra that contains both $\mathcal{F}_t$ and $\mathcal{A}$, sometimes also denoted by $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{A}$. Show that B_t is a Brownian motion with respect to filtration $\{\mathcal{G}_t\}$.

Hint. $\mathcal{G}_t$ is generated by intersections $F \cap A$ with $F \in \mathcal{F}_t$ and $A \in \mathcal{A}$. Use the π-λ theorem.

Exercise 2.23. Use (2.41), limits, and the continuous distribution of Brownian motion to show

$$P^0(B_t \le a, M_t > b) = P^0(B_t > 2b - a)$$

and

$$P^0(B_t \le a, M_t > b) = P^0(B_t \ge 2b - a).$$

Exercise 2.24. Let B_t be standard Brownian motion and M_t its running maximum. Show that the joint density $f(x, y)$ of (B_t, M_t) is

$$f(x,y) = \frac{2(2y-x)}{\sqrt{2\pi}t^{3/2}} \exp\Bigl(-\frac{1}{2t}(2y-x)^2\Bigr) \tag{2.48}$$

on the domain $0 < y < \infty$, $-\infty < x < y$. *Hint:* Use (2.41) and convert $P^0(B_t \ge 2b - a)$ into a double integral of the form $\int_b^\infty dy \int_{-\infty}^a dx\, f(x, y)$.

Can you give an argument for why it is enough to consider the events $\{B_t \le a, M_t \ge b\}$ for $a < b$ and $b > 0$?

Exercise 2.25. Consider Brownian motion started at $x \ge 0$ and let $\tau_0 = \inf\{t \ge 0 : B_t = 0\}$ be the first hitting time of the origin. The process

$$X_t = \begin{cases} B_t, & t < \tau_0 \\ 0, & t \ge \tau_0 \end{cases}$$

is Brownian motion killed (or absorbed) at the origin. Show that on the positive half-line $(0, \infty)$, X_t started at $x > 0$ has density $q(t, x, y) = p(t, x, y) - p(t, x, -y)$ where $p(t, x, y)$ is the Gaussian kernel (2.39). In other words, for $x, z > 0$ derive

$$P^x(B_t > z, \tau_0 > t) = \int_z^\infty [p(t,x,y) - p(t,x,-y)]\, dy.$$

Hint. Start from $P^x(B_t > z, \tau_0 > t) = P^0(x + B_t > z, \min_{s \le t}(x + B_s) > 0)$, use symmetry and the reflection principle.

Exercise 2.26. Let B and X be two independent one-dimensional Brownian motions. Following the proof of Proposition 2.42, show that

$$\lim_{\text{mesh}(\pi)\to 0} \sum_{i=0}^{m(\pi)-1} (B_{t_{i+1}} - B_{t_i})(X_{t_{i+1}} - X_{t_i}) = 0 \quad \text{in } L^2(P). \tag{2.49}$$

As a consequence of your calculation, find the covariation $[B, X]$.

Can you also find $[B, X]$ from the definition in equation (2.12)?

Exercise 2.27. Let N be a homogeneous Poisson process. Show that for almost every ω, $N_t(\omega) - N_{t-}(\omega) = 0$ or 1 for all t.

Hint. If $N_t - N_{t-} \ge 2$ for some $t \in [0, T]$, then for any partition $0 = t_0 < t_1 < \cdots < t_n = T$, $N(t_i, t_{i+1}] \ge 2$ for some i. A crude bound shows that this probability can be made arbitrarily small by shrinking the mesh of the partition.

Exercise 2.28. Let N be a homogeneous rate α Poisson process on $\mathbf{R}_+$ with respect to a filtration $\{\mathcal{F}_t\}$, and $M_t = N_t - \alpha t$. Show that $M_t^2 - \alpha t$ and $M_t^2 - N_t$ are martingales.

Exercise 2.29. As in the previous execise, let N be a homogeneous rate α Poisson process and $M_t = N_t - \alpha t$. Use Corollary A.11 from the appendix to find the quadratic variation processes $[M]$ and $[N]$.

Chapter 3

Martingales

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $\{\mathcal{F}_t\}$. We assume that $\{\mathcal{F}_t\}$ is complete but not necessarily right-continuous, unless so specified. As defined in the previous chapter, a *martingale* with respect to $\{\mathcal{F}_t\}$ is a real-valued stochastic process $M = \{M_t : t \in \mathbf{R}_+\}$ adapted to $\{\mathcal{F}_t\}$ such that M_t is integrable for each t, and

$$E[M_t|\mathcal{F}_s] = M_s \text{ for all } s < t.$$

If the equality above is relaxed to

$$E[M_t|\mathcal{F}_s] \geq M_s \text{ for all } s < t$$

then M is a *submartingale*. M is a *supermartingale* if $-M$ is a submartingale. M is *square-integrable* if $E[M_t^2] < \infty$ for all t.

These properties are preserved by certain classes of functions.

Proposition 3.1. (a) *If M is a martingale and φ a convex function such that $\varphi(M_t)$ is integrable for all $t \geq 0$, then $\varphi(M_t)$ is a submartingale.*

(b) *If M is a submartingale and φ a nondecreasing convex function such that $\varphi(M_t)$ is integrable for all $t \geq 0$, then $\varphi(M_t)$ is a submartingale.*

Proof. Part (a) follows from Jensen's inequality. For $s < t$,

$$E[\varphi(M_t)|\mathcal{F}_s] \geq \varphi\big(E[M_t|\mathcal{F}_s]\big) = \varphi(M_s).$$

Part (b) follows from the same calculation, but now the last equality becomes the inequality $\geq$ due to the submartingale property $E[M_t|\mathcal{F}_s] \geq M_s$ and the monotonicity of φ. □

The martingales we work with have always right-continuous paths. Then it is sometimes convenient to enlarge the filtration to $\{\mathcal{F}_{t+}\}$ if the filtration is not right-continuous to begin with. The next proposition permits this move. An example of its use appears in the proof of Doob's inequality, Theorem 3.12.

Proposition 3.2. *Suppose M is a right-continuous submartingale with respect to a filtration $\{\mathcal{F}_t\}$. Then M is a submartingale also with respect to $\{\mathcal{F}_{t+}\}$.*

Proof. Let $s < t$ and consider $n > (t-s)^{-1}$. $M_t \vee c$ is a submartingale, so

$$E[M_t \vee c | \mathcal{F}_{s+n^{-1}}] \geq M_{s+n^{-1}} \vee c.$$

Since $\mathcal{F}_{s+} \subseteq \mathcal{F}_{s+n^{-1}}$,

$$E[M_t \vee c | \mathcal{F}_{s+}] \geq E[M_{s+n^{-1}} \vee c | \mathcal{F}_{s+}].$$

By the bounds

$$c \leq M_{s+n^{-1}} \vee c \leq E[M_t \vee c | \mathcal{F}_{s+n^{-1}}]$$

and Lemma B.16 from the Appendix, for a fixed c the random variables $\{M_{s+n^{-1}} \vee c\}$ are uniformly integrable. Let $n \to \infty$. Right-continuity of paths implies $M_{s+n^{-1}} \vee c \to M_s \vee c$. Uniform integrability then gives convergence in L^1. By Lemma B.17 there exists a subsequence $\{n_j\}$ such that conditional expectations converge almost surely:

$$E[M_{s+n_j^{-1}} \vee c | \mathcal{F}_{s+}] \to E[M_s \vee c | \mathcal{F}_{s+}].$$

Consequently

$$E[M_t \vee c | \mathcal{F}_{s+}] \geq E[M_s \vee c | \mathcal{F}_{s+}] = M_s \vee c \geq M_s.$$

As $c \to -\infty$, the dominated convergence theorem for conditional expectations (Theorem B.14) makes the conditional expectation on the left converge, and in the limit $E[M_t | \mathcal{F}_{s+}] \geq M_s$. □

The connection between the right-continuity of the (sub)martingale and the filtration goes the other way too. The statement below is Theorem 1.3.13 in [11].

Proposition 3.3. *Suppose the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions, in other words $(\Omega, \mathcal{F}, P)$ is complete, $\mathcal{F}_0$ contains all null events, and $\mathcal{F}_t = \mathcal{F}_{t+}$. Let M be a submartingale such that $t \mapsto EM_t$ is right-continuous. Then there exists a cadlag modification of M that is an $\{\mathcal{F}_t\}$-submartingale.*

3.1. Optional stopping

Optional stopping has to do with extending the submartingale property $E[M_t|\mathcal{F}_s] \geq M_s$ from deterministic times $s < t$ to stopping times. We begin with discrete stopping times.

Lemma 3.4. *Let M be a submartingale. Let σ and τ be two stopping times whose values lie in an ordered countable set $\{s_1 < s_2 < s_3 < \cdots\} \cup \{\infty\} \subseteq [0, \infty]$ where $s_j \nearrow \infty$. Then for any $T < \infty$,*

$$E[M_{\tau\wedge T}|\mathcal{F}_\sigma] \geq M_{\sigma\wedge\tau\wedge T}. \tag{3.1}$$

Proof. Fix n so that $s_n \leq T < s_{n+1}$. First observe that $M_{\tau\wedge T}$ is integrable, because

$$\begin{aligned} |M_{\tau\wedge T}| &= \sum_{i=1}^{n} \mathbf{1}\{\tau = s_i\}|M_{s_i}| + \mathbf{1}\{\tau > s_n\}|M_T| \\ &\leq \sum_{i=1}^{n} |M_{s_i}| + |M_T|. \end{aligned}$$

Next check that $M_{\sigma\wedge\tau\wedge T}$ is $\mathcal{F}_\sigma$-measurable. For discrete stopping times this is simple. We need to show that $\{M_{\sigma\wedge\tau\wedge T} \in B\} \cap \{\sigma \leq t\} \in \mathcal{F}_t$ for all $B \in \mathcal{B}_{\mathbf{R}}$ and t. Let s_j be the highest value not exceeding t. (If there is no such s_i, then $t < s_1$, the event above is empty and lies in $\mathcal{F}_t$.) Then

$$\{M_{\sigma\wedge\tau\wedge T} \in B\} \cap \{\sigma \leq t\} = \bigcup_{i=1}^{j} \Big(\{\sigma \wedge \tau = s_i\} \cap \{M_{s_i\wedge T} \in B\} \cap \{\sigma \leq t\} \Big).$$

This is a union of events in $\mathcal{F}_t$ because $s_i \leq t$ and $\sigma \wedge \tau$ is a stopping time.

Since both $E[M_{\tau\wedge T}|\mathcal{F}_\sigma]$ and $M_{\sigma\wedge\tau\wedge T}$ are $\mathcal{F}_\sigma$-measurable, (3.1) follows from checking that

$$E\{\mathbf{1}_A\, E[M_{\tau\wedge T}|\mathcal{F}_\sigma]\} \geq E\{\mathbf{1}_A M_{\sigma\wedge\tau\wedge T}\} \quad \text{for all } A \in \mathcal{F}_\sigma.$$

By the definition of conditional expectation, this reduces to showing

$$E[\mathbf{1}_A M_{\tau\wedge T}] \geq E[\mathbf{1}_A M_{\sigma\wedge\tau\wedge T}].$$

Decompose A according to whether $\sigma \leq T$ or $\sigma > T$. If $\sigma > T$, then $\tau \wedge T = \sigma \wedge \tau \wedge T$ and then

$$E[\mathbf{1}_{A\cap\{\sigma>T\}} M_{\tau\wedge T}] = E[\mathbf{1}_{A\cap\{\sigma>T\}} M_{\sigma\wedge\tau\wedge T}].$$

To handle the case $\sigma \leq T$ we decompose it into subcases

$$\begin{aligned} E[\mathbf{1}_{A\cap\{\sigma=s_i\}} M_{\tau\wedge T}] &\geq E[\mathbf{1}_{A\cap\{\sigma=s_i\}} M_{\sigma\wedge\tau\wedge T}] \\ &= E[\mathbf{1}_{A\cap\{\sigma=s_i\}} M_{s_i\wedge\tau\wedge T}] \quad \text{for } 1 \leq i \leq n. \end{aligned}$$

Since $A \cap \{\sigma = s_i\} \in \mathcal{F}_{s_i}$, by conditioning again on the left, this last conclusion will follow from

$$E[M_{\tau\wedge T}|\mathcal{F}_{s_i}] \geq M_{s_i\wedge\tau\wedge T} \quad \text{for } 1 \leq i \leq n. \tag{3.2}$$

We check (3.2) by an iterative argument. First an auxiliary inequality: for any j,

$$\begin{aligned} E[M_{s_{j+1}\wedge\tau\wedge T}|\mathcal{F}_{s_j}] &= E\big[M_{s_{j+1}\wedge T}\mathbf{1}\{\tau > s_j\} + M_{s_j\wedge\tau\wedge T}\mathbf{1}\{\tau \leq s_j\} \bigm| \mathcal{F}_{s_j}\big] \\ &= E[M_{s_{j+1}\wedge T}|\mathcal{F}_{s_j}] \cdot \mathbf{1}\{\tau > s_j\} + M_{s_j\wedge\tau\wedge T}\mathbf{1}\{\tau \leq s_j\} \\ &\geq M_{s_j\wedge T}\mathbf{1}\{\tau > s_j\} + M_{s_j\wedge\tau\wedge T}\mathbf{1}\{\tau \leq s_j\} \\ &= M_{s_j\wedge\tau\wedge T}. \end{aligned}$$

Above we used the fact that $M_{s_j\wedge\tau\wedge T}$ is $\mathcal{F}_{s_j}$-measurable (which was checked above) and then the submartingale property. Since $\tau \wedge T = s_{n+1} \wedge \tau \wedge T$ (recall that $s_n \leq T < s_{n+1}$), applying the above inequality to $j = n$ gives

$$E[M_{\tau\wedge T}|\mathcal{F}_{s_n}] \geq M_{s_n\wedge\tau\wedge T}$$

which is case $i = n$ of (3.2). Now do induction: assuming (3.2) has been checked for i and applying the auxiliary inequality again gives

$$\begin{aligned} E[M_{\tau\wedge T}|\mathcal{F}_{s_{i-1}}] &= E\big\{E[M_{\tau\wedge T}|\mathcal{F}_{s_i}] \bigm| \mathcal{F}_{s_{i-1}}\big\} \\ &\geq E\big\{M_{s_i\wedge\tau\wedge T} \bigm| \mathcal{F}_{s_{i-1}}\big\} \geq M_{s_{i-1}\wedge\tau\wedge T} \end{aligned}$$

which is (3.2) for $i - 1$. Repeat this until (3.2) has been proved down to $i = 1$. ☐

To extend this result to general stopping times, we assume some regularity on the paths of M. First we derive a moment bound.

Lemma 3.5. *Let M be a submartingale with right-continuous paths and $T < \infty$. Then for any stopping time ρ that satisfies $P\{\rho \leq T\} = 1$,*

$$E\,|M_\rho| \leq 2E[M_T^+] - E[M_0].$$

Proof. Define a discrete approximation of ρ by $\rho_n = T$ if $\rho = T$, and $\rho_n = 2^{-n}T(\lfloor 2^n\rho/T\rfloor + 1)$ if $\rho < T$. Then ρ_n is a stopping time with finitely many values in $[0, T]$, and $\rho_n \searrow \rho$ as $n \to \infty$.

Averaging over (3.1) gives $E[M_{\tau\wedge T}] \geq E[M_{\sigma\wedge\tau\wedge T}]$. Apply this to $\tau = \rho_n$ and $\sigma = 0$ to get

$$EM_{\rho_n} \geq EM_0.$$

Next, apply (3.1) to the submartingale $M_t^+ = M_t \vee 0$, with $\tau = T$ and $\sigma = \rho_n$ to get

$$EM_T^+ \geq EM_{\rho_n}^+.$$

From the above

$$EM^-_{\rho_n} = EM^+_{\rho_n} - EM_{\rho_n} \le EM^+_T - EM_0.$$

Combining these,

$$E|M_{\rho_n}| = EM^+_{\rho_n} + EM^-_{\rho_n} \le 2EM^+_T - EM_0.$$

Let $n \to \infty$, use right-continuity and apply Fatou's Lemma to get

$$E|M_\rho| \le \varliminf_{n\to\infty} E|M_{\rho_n}| \le 2EM^+_T - EM_0. \qquad \square$$

The conclusion from the previous lemma needed next is that for any stopping time τ and $T \in \mathbf{R}_+$, the stopped variable $M_{\tau\wedge t}$ is integrable. Here is the extension of Lemma 3.4 from discrete to general stopping times.

Theorem 3.6. *Let M be a submartingale with right-continuous paths, and let σ and τ be two stopping times. Then for $T < \infty$,*

$$E[M_{\tau\wedge T}|\mathcal{F}_\sigma] \ge M_{\sigma\wedge\tau\wedge T}. \tag{3.3}$$

Remark 3.7. See Exercise 3.2 for a standard example that shows that (3.3) is not true in general without the truncation at a finite time T.

Proof of Theorem 3.6. As pointed out before the theorem, $M_{\tau\wedge T}$ and $M_{\sigma\wedge\tau\wedge T}$ are integrable random variables. In particular, the conditional expectation is well-defined.

Define approximating discrete stopping times by $\sigma_n = 2^{-n}(\lfloor 2^n\sigma\rfloor + 1)$ and $\tau_n = 2^{-n}(\lfloor 2^n\tau\rfloor + 1)$. The interpretation for infinite values is that $\sigma_n = \infty$ if $\sigma = \infty$, and similarly for τ_n and τ.

Let $c \in \mathbf{R}$. The function $x \mapsto x \vee c$ is convex and nondecreasing, hence $M_t \vee c$ is also a submartingale. Applying Lemma 3.4 to this submartingale and the stopping times σ_n and τ_n gives

$$E[M_{\tau_n\wedge T} \vee c|\mathcal{F}_{\sigma_n}] \ge M_{\sigma_n\wedge\tau_n\wedge T} \vee c.$$

Since $\sigma \le \sigma_n$, $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n}$, and if we condition both sides of the above inequality on $\mathcal{F}_\sigma$, we get

$$E[M_{\tau_n\wedge T} \vee c|\mathcal{F}_\sigma] \ge E[M_{\sigma_n\wedge\tau_n\wedge T} \vee c|\mathcal{F}_\sigma]. \tag{3.4}$$

The purpose is now to let $n \to \infty$ in (3.4) and obtain the conclusion (3.3) for the truncated process $M_t \vee c$, and then let $c \searrow -\infty$ and get the conclusion. The time arguments converge from the right: $\tau_n \wedge T \searrow \tau \wedge T$ and $\sigma_n \wedge \tau_n \wedge T \searrow \sigma \wedge \tau \wedge T$. Then by the right-continuity of M,

$$M_{\tau_n\wedge T} \to M_{\tau\wedge T} \quad \text{and} \quad M_{\sigma_n\wedge\tau_n\wedge T} \to M_{\sigma\wedge\tau\wedge T}.$$

Next we justify convergence of the conditional expectations along a subsequence. By Lemma 3.4

$$c \le M_{\tau_n \wedge T} \vee c \le E[M_T \vee c | \mathcal{F}_{\tau_n}]$$

and

$$c \le M_{\sigma_n \wedge \tau_n \wedge T} \vee c \le E[M_T \vee c | \mathcal{F}_{\sigma_n \wedge \tau_n}].$$

Together with Lemma B.16 from the Appendix, these bounds imply that the sequences $\{M_{\tau_n \wedge T} \vee c : n \in \mathbf{N}\}$ and $\{M_{\sigma_n \wedge \tau_n \wedge T} \vee c : n \in \mathbf{N}\}$ are uniformly integrable. Since these sequences converge almost surely (as argued above), uniform integrability implies that they converge in L^1. By Lemma B.17 there exists a subsequence $\{n_j\}$ along which the conditional expectations converge almost surely:

$$E[M_{\tau_{n_j} \wedge T} \vee c | \mathcal{F}_\sigma] \to E[M_{\tau \wedge T} \vee c | \mathcal{F}_\sigma]$$

and

$$E[M_{\sigma_{n_j} \wedge \tau_{n_j} \wedge T} \vee c | \mathcal{F}_\sigma] \to E[M_{\sigma \wedge \tau \wedge T} \vee c | \mathcal{F}_\sigma].$$

(To get a subsequence that works for both limits, extract a subsequence for the first limit by Lemma B.17, and then apply Lemma B.17 again to extract a further subsubsequence for the second limit.) Taking these limits in (3.4) gives

$$E[M_{\tau \wedge T} \vee c | \mathcal{F}_\sigma] \ge E[M_{\sigma \wedge \tau \wedge T} \vee c | \mathcal{F}_\sigma].$$

M is right-continuous by assumption, hence progressively measurable, and so $M_{\sigma \wedge \tau \wedge T}$ is $\mathcal{F}_{\sigma \wedge \tau \wedge T}$-measurable. This is a sub-σ-field of $\mathcal{F}_\sigma$, and so

$$E[M_{\tau \wedge T} \vee c | \mathcal{F}_\sigma] \ge E[M_{\sigma \wedge \tau \wedge T} \vee c | \mathcal{F}_\sigma] = M_{\sigma \wedge \tau \wedge T} \vee c \ge M_{\sigma \wedge \tau \wedge T}.$$

As $c \searrow -\infty$, $M_{\tau \wedge T} \vee c \to M_{\tau \wedge T}$ pointwise, and for $c \le 0$ we have the integrable bound $|M_{\tau \wedge T} \vee c| \le |M_{\tau \wedge T}|$. Thus by the dominated convergence theorem for conditional expectations, almost surely

$$\lim_{c \to -\infty} E[M_{\tau \wedge T} \vee c | \mathcal{F}_\sigma] = E[M_{\tau \wedge T} | \mathcal{F}_\sigma].$$

This completes the proof. □

Corollary 3.8. *Suppose M is a right-continuous submartingale and τ is a stopping time. Then the stopped process $M^\tau = \{M_{\tau \wedge t} : t \in \mathbf{R}_+\}$ is a submartingale with respect to the original filtration $\{\mathcal{F}_t\}$.*

If M is a also a martingale, then M^τ is a martingale. And finally, if M is an L^2-martingale, then so is M^τ.

Proof. In (3.3), take $T = t$, $\sigma = s < t$. Then it becomes the submartingale property for M^τ:

$$E[M_{\tau \wedge t} | \mathcal{F}_s] \ge M_{\tau \wedge s}. \tag{3.5}$$

If M is a martingale, we can apply this to both M and $-M$. And if M is an L^2-martingale, Lemma 3.5 implies that so is M^τ. □

Corollary 3.9. *Suppose M is a right-continuous submartingale. Let $\{\sigma(u) : u \ge 0\}$ be a nondecreasing, $[0,\infty)$-valued process such that $\sigma(u)$ is a bounded stopping time for each u. Then $\{M_{\sigma(u)} : u \ge 0\}$ is a submartingale with respect to the filtration $\{\mathcal{F}_{\sigma(u)} : u \ge 0\}$.*

If M is a martingale or an L^2-martingale to begin with, then so is $M_{\sigma(u)}$.

Proof. For $u < v$ and $T \ge \sigma(v)$, (3.3) gives $E[M_{\sigma(v)}|\mathcal{F}_{\sigma(u)}] \ge M_{\sigma(u)}$. If M is a martingale, we can apply this to both M and $-M$. And if M is an L^2-martingale, Lemma 3.5 applied to the submartingale M^2 implies that $E[M^2_{\sigma(u)}] \le 2E[M_T^2] + E[M_0^2]$. □

The last corollary has the following implications:

(i) M^τ is a submartingale not only with respect to $\{\mathcal{F}_t\}$ but also with respect to $\{\mathcal{F}_{\tau\wedge t}\}$.

(ii) Let M be an L^2-martingale and τ a bounded stopping time. Then $\bar{M}_t = M_{\tau+t} - M_\tau$ is an L^2-martingale with respect to $\bar{\mathcal{F}}_t = \mathcal{F}_{\tau+t}$.

3.2. Inequalities and limits

Lemma 3.10. *Let M be a submartingale, $0 < T < \infty$, and H a finite subset of $[0,T]$. Then for $r > 0$,*

$$P\Big\{\max_{t\in H} M_t \ge r\Big\} \le r^{-1}E[M_T^+] \tag{3.6}$$

and

$$P\Big\{\min_{t\in H} M_t \le -r\Big\} \le r^{-1}\big(E[M_T^+] - E[M_0]\big). \tag{3.7}$$

Proof. Let $\sigma = \min\{t \in H : M_t \ge r\}$, with the interpretation that $\sigma = \infty$ if $M_t < r$ for all $t \in H$. (3.3) with $\tau = T$ gives

$$E[M_T] \ge E[M_{\sigma\wedge T}] = E[M_\sigma \mathbf{1}_{\{\sigma<\infty\}}] + E[M_T\mathbf{1}_{\{\sigma=\infty\}}],$$

from which

$$\begin{aligned} rP\Big\{\max_{t\in H} M_t \ge r\Big\} &= rP\{\sigma < \infty\} \le E[M_\sigma\mathbf{1}_{\{\sigma<\infty\}}] \le E[M_T\mathbf{1}_{\{\sigma<\infty\}}] \\ &\le E[M_T^+\mathbf{1}_{\{\sigma<\infty\}}] \le E[M_T^+]. \end{aligned}$$

This proves (3.6).

To prove (3.7), let $\tau = \min\{t \in H : M_t \le -r\}$. (3.4) with $\sigma = 0$ gives

$$E[M_0] \le E[M_{\tau\wedge T}] = E[M_\tau\mathbf{1}_{\{\tau<\infty\}}] + E[M_T\mathbf{1}_{\{\tau=\infty\}}],$$

from which

$$-rP\Big\{\min_{t\in H} M_t \le -r\Big\} = -rP\{\tau<\infty\} \ge E[M_\tau \mathbf{1}_{\{\tau<\infty\}}]$$
$$\ge E[M_0] - E[M_T\mathbf{1}_{\{\tau=\infty\}}] \ge E[M_0] - E[M_T^+]. \quad \square$$

Next we generalize this to uncountable suprema and infima.

Theorem 3.11. *Let M be a right-continuous submartingale and $0 < T < \infty$. Then for $r > 0$,*

$$P\Big\{\sup_{0\le t\le T} M_t \ge r\Big\} \le r^{-1}E[M_T^+] \tag{3.8}$$

and

$$P\Big\{\inf_{0\le t\le T} M_t \le -r\Big\} \le r^{-1}\big(E[M_T^+] - E[M_0]\big). \tag{3.9}$$

Proof. Let H be a countable dense subset of $[0,T]$ that contains 0 and T, and let $H_1 \subseteq H_2 \subseteq H_3 \subseteq \cdots$ be finite sets such that $H = \bigcup H_n$. Lemma 3.10 applies to the sets H_n. Let $b < r$. By right-continuity,

$$P\Big\{\sup_{0\le t\le T} M_t > b\Big\} = P\Big\{\sup_{t\in H} M_t > b\Big\} = \lim_{n\to\infty} P\Big\{\sup_{t\in H_n} M_t > b\Big\}$$
$$\le b^{-1}E[M_T^+].$$

Let $b \nearrow r$. This proves (3.8). (3.9) is proved by a similar argument. $\square$

When X has either left- or right-continuous paths with probability 1, we define

$$X_T^*(\omega) = \sup_{0\le t\le T} |X_t(\omega)|. \tag{3.10}$$

The measurability of X_T^* is checked as follows. First define $U = \sup_{s\in R}|X_s|$ where R contains T and all rationals in $[0,T]$. U is $\mathcal{F}_T$-measurable as a supremum of countably many $\mathcal{F}_T$-measurable random variables. On every left- or right-continuous path U coincides with X_T^*. Thus $U = X_T^*$ at least almost surely. By the completeness assumption on the filtration, all events of probability zero and their subsets lie in $\mathcal{F}_T$, and so X_T^* is also $\mathcal{F}_T$-measurable.

Theorem 3.12. (Doob's Inequality) *Let M be a nonnegative right-continuous submartingale and $0 < T < \infty$. Then for $1 < p < \infty$*

$$E\Big[\sup_{0\le t\le T} M_t^p\Big] \le \Big(\frac{p}{p-1}\Big)^p E\big[M_T^p\big]. \tag{3.11}$$

Proof. Since M is nonnegative, $M^*_T = \sup_{0\le t\le T} M_t$. The first part of the proof is to justify the inequality

$$P\{M^*_T > r\} \le r^{-1} E\big[M_T \mathbf{1}\{M^*_T \ge r\}\big] \tag{3.12}$$

for $r > 0$. Let

$$\tau = \inf\{t > 0 : M_t > r\}.$$

This is an $\{\mathcal{F}_{t+}\}$-stopping time by Lemma 2.7. By right-continuity $M_\tau \ge r$ when $\tau < \infty$. Quite obviously $M^*_T > r$ implies $\tau \le T$, and so

$$rP\{M^*_T > r\} \le E\big[M_\tau \mathbf{1}\{M^*_T > r\}\big] \le E\big[M_\tau \mathbf{1}\{\tau \le T\}\big].$$

Since M is a submartingale with respect to $\{\mathcal{F}_{t+}\}$ by Proposition 3.2, Theorem 3.6 gives

$$\begin{aligned} E[M_\tau \mathbf{1}_{\{\tau\le T\}}] &= E[M_{\tau\wedge T}] - E[M_T \mathbf{1}_{\{\tau>T\}}] \le E[M_T] - E[M_T \mathbf{1}_{\{\tau>T\}}] \\ &= E[M_T \mathbf{1}_{\{\tau\le T\}}] \le E\big[M_T \mathbf{1}\{M^*_T \ge r\}\big]. \end{aligned}$$

(3.12) has been verified.

Let $0 < b < \infty$. By (1.43) and Hölder's inequality,

$$\begin{aligned} E\big[(M^*_T \wedge b)^p\big] &= \int_0^b pr^{p-1} P[M^*_T > r]\, dr \le \int_0^b pr^{p-2} E\big[M_T \mathbf{1}\{M^*_T \ge r\}\big]\, dr \\ &= E\Big[M_T \cdot \int_0^{b\wedge M^*_T} pr^{p-2}\, dr\Big] = \frac{p}{p-1} \cdot\ E\big[M_T (b \wedge M^*_T)^{p-1}\big] \\ &\le \frac{p}{p-1} \cdot E\big[M^p_T\big]^{\frac{1}{p}} E\big[(b\wedge M^*_T)^p\big]^{\frac{p-1}{p}}. \end{aligned}$$

The truncation at b guarantees that the last expectation is finite so we can divide by it through the inequality to get

$$E\big[(M^*_T \wedge b)^p\big]^{\frac{1}{p}} \le \frac{p}{p-1} \cdot E\big[M^p_T\big]^{\frac{1}{p}}.$$

Raise both sides of this last inequality to power p and then let $b \nearrow \infty$. Monotone convergence theorem gives the conclusion. □

The obvious application would be to $M = |X|$ for a martingale X. By applying the previous inequalities to the stopped process $M_{t\wedge\tau}$, we can replace T with a bounded stopping time τ. We illustrate the idea with Doob's inequality.

Corollary 3.13. *Let M be a nonnegative right-continuous submartingale and τ a bounded stopping time. The for $1 < p < \infty$*

$$E\Big[\Big(\sup_{0\le t\le \tau} M_t\Big)^p\Big] \le \Big(\frac{p}{p-1}\Big)^p E\big[M^p_\tau\big]. \tag{3.13}$$

Proof. Pick T so that $\tau \le T$. Since $\sup_{0\le t\le T} M_{t\wedge\tau} = \sup_{0\le t\le \tau} M_t$ and $M_{T\wedge\tau} = M_\tau$, the result follows immediately by applying (3.11) to $M_{t\wedge\tau}$. □

Our treatment of martingales would not be complete without mentioning martingale limit theorems, although strictly speaking we do not need them for developing stochastic integration. Here is the basic martingale convergence theorem that gives almost sure convergence.

Theorem 3.14. *Let M be a right-continuous submartingale such that*

$$\sup_{t\in\mathbf{R}_+} E(M_t^+) < \infty.$$

Then there exists a random variable M_∞ such that $E|M_\infty| < \infty$ and $M_t(\omega) \to M_\infty(\omega)$ as $t\to\infty$ for almost every ω.

Now suppose $M = \{M_t : t \in \mathbf{R}_+\}$ is a martingale. When can we take the limit M_∞ and adjoin it to the process in the sense that $\{M_t : t \in [0,\infty]\}$ is a martingale? The integrability of M_∞ is already part of the conclusion of Theorem 3.14. However, to also have $E(M_\infty|\mathcal{F}_t) = M_t$ we need an additional hypotheses of uniform integrability (Definition B.15 in the appendix).

Theorem 3.15. *Let $M = \{M_t : t \in \mathbf{R}_+\}$ be a right-continuous martingale. Then the following four conditions are equivalent.*

(i) *The collection $\{M_t : t \in \mathbf{R}_+\}$ is uniformly integrable.*

(ii) *There exists an integrable random variable M_∞ such that*

$$\lim_{t\to\infty} E|M_t - M_\infty| = 0 \qquad (L^1 \text{ convergence}).$$

(iii) *There exists an integrable random variable M_∞ such that $M_t(\omega) \to M_\infty(\omega)$ almost surely and $E(M_\infty|\mathcal{F}_t) = M_t$ for all $t \in \mathbf{R}_+$.*

(iv) *There exists an integrable random variable Z such that $M_t = E(Z|\mathcal{F}_t)$ for all $t \in \mathbf{R}_+$.*

As quick corollaries we get for example the following statements.

Corollary 3.16. (a) *For $Z \in L^1(P)$, $E(Z|\mathcal{F}_t) \to E(Z|\mathcal{F}_\infty)$ as $t\to\infty$ both almost surely and in L^1.*

(b) (Lévy's 0-1 law) *For $A \in \mathcal{F}_\infty$, $E(\mathbf{1}_A|\mathcal{F}_t) \to \mathbf{1}_A$ as $t\to\infty$ both almost surely and in L^1.*

Proof. Part (b) follows from (a). To see (a), start by defining $M_t = E(Z|\mathcal{F}_t)$ so that statement (iv) of Theorem 3.15 is valid. By (ii) and (iii) we have an a.s. and L^1 limit M_∞ (explain why there cannot be two different limits). By construction M_∞ is $\mathcal{F}_\infty$-measurable. For $A \in \mathcal{F}_s$ we have for $t > s$ and by L^1 convergence

$$E[\mathbf{1}_A Z] = E[\mathbf{1}_A M_t] \to E[\mathbf{1}_A M_\infty].$$

A π-λ argument extends $E[\mathbf{1}_A Z] = E[\mathbf{1}_A M_\infty]$ to all $A \in \mathcal{F}_\infty$. $\square$

3.3. Local martingales and semimartingales

For a stopping time τ and a process $X = \{X_t : t \in \mathbf{R}_+\}$, the *stopped process* X^τ is defined by $X^\tau_t = X_{\tau\wedge t}$.

Definition 3.17. Let $M = \{M_t : t \in \mathbf{R}_+\}$ be a process adapted to a filtration $\{\mathcal{F}_t\}$. M is a *local martingale* if there exists a sequence of stopping times $\tau_1 \le \tau_2 \le \tau_3 \le \cdots$ such that $P\{\tau_k \nearrow \infty\} = 1$ and for each k M^{τ_k} is a martingale with respect to $\{\mathcal{F}_t\}$. M is a *local square-integrable martingale* if M^{τ_k} is a square-integrable martingale for each k. In both cases we say $\{\tau_k\}$ is a *localizing sequence* for M.

Remark 3.18. (a) Since $M_0^{\tau_k} = M_0$ the definition above requires that M_0 is integrable. This extra restriction can be avoided by phrasing the definition so that $M_{t\wedge\tau_n}\mathbf{1}_{\{\tau_n>0\}}$ is a martingale [**12, 14**]. Another way is to require that $M_{t\wedge\tau_n} - M_0$ is a martingale [**3, 10**]. We use the simple definition since we have no need for the extra generality of nonintegrable M_0.

(b) In some texts the definition of local martingale also requires that M^{τ_k} is uniformly integrable. This can be easily arranged (Exercise 3.10).

(c) Further localization gains nothing. That is, if M is an adapted process and $\rho_n \nearrow \infty$ (a.s.) are stopping times such that M^{ρ_n} is a local martingale for each n, then M itself is a local martingale (Exercise 3.5).

(d) Sometimes we consider a local martingale $\{M_t : t \in [0, T]\}$ restricted to a bounded time interval. Then it seems pointless to require that $\tau_n \nearrow \infty$. Indeed it is equivalent to require a nondecreasing sequence of stopping times σ_n such that $\{M_{t\wedge\sigma_n} : t \in [0, T]\}$ is a martingale for each n and, almost surely, $\sigma_n \ge T$ for large enough n. Given such a sequence σ_n one can check that the original definition can be recovered by taking $\tau_n = \sigma_n \cdot \mathbf{1}\{\sigma_n < T\} + \infty \cdot \mathbf{1}\{\sigma_n \ge T\}$.

We shall also use the shorter term *local L^2-martingale* for a local square-integrable martingale.

Lemma 3.19. *Suppose M is a local martingale and σ is an arbitrary stopping time. Then M^σ is also a local martingale. Similarly, if M is a local L^2-martingale, then so is M^σ. In both cases, if $\{\tau_k\}$ is a localizing sequence for M, then it is also a localizing sequence for M^σ.*

Proof. Let $\{\tau_k\}$ be a sequence of stopping times such that $\tau_k \nearrow \infty$ and M^{τ_k} is a martingale. By Corollary 3.8 the process $M^{\tau_k}_{\sigma\wedge t} = (M^\sigma)^{\tau_k}_t$ is a martingale. Thus the stopping times τ_k work also for M^σ.

If M^{τ_k} is an L^2-martingale, then so is $M^{\tau_k}_{\sigma\wedge t} = (M^\sigma)^{\tau_k}_t$, because by applying Lemma 3.5 to the submartingale $(M^{\tau_k})^2$,

$$E[M^2_{\sigma\wedge\tau_k\wedge t}] \le 2E[M^2_{\tau_k\wedge t}] + E[M_0^2]. \qquad \square$$

Only large jumps can prevent a cadlag local martingale from being a local L^2-martingale. See Exercise 3.12.

Lemma 3.20. *Suppose M is a cadlag local martingale, and there is a constant c such that $|M_t(\omega) - M_{t-}(\omega)| \le c$ for all $t \in \mathbf{R}_+$ and $\omega \in \Omega$. Then M is a local L^2-martingale.*

Proof. Let $\tau_k \nearrow \infty$ be stopping times such that M^{τ_k} is a martingale. Let

$$\rho_k = \inf\{t \ge 0 : |M_t| \text{ or } |M_{t-}| \ge k\}$$

be the stopping times defined by (2.8) and (2.9). By the cadlag assumption, each path $t \mapsto M_t(\omega)$ is locally bounded (means: bounded in any bounded time interval), and consequently $\rho_k(\omega) \nearrow \infty$ as $k \nearrow \infty$. Let $\sigma_k = \tau_k \wedge \rho_k$. Then $\sigma_k \nearrow \infty$, and M^{σ_k} is a martingale for each k. Furthermore,

$$|M^{\sigma_k}_t| = |M_{\tau_k\wedge\rho_k\wedge t}| \le \sup_{0\le s<\rho_k} |M_s| + |M_{\rho_k} - M_{\rho_k -}| \le k + c.$$

So M^{σ_k} is a bounded process, and in particular M^{σ_k} is an L^2-process. $\square$

Recall that the *usual conditions* on the filtration $\{\mathcal{F}_t\}$ meant that the filtration is complete (each $\mathcal{F}_t$ contains every subset of a P-null event in $\mathcal{F}$) and right-continuous ($\mathcal{F}_t = \mathcal{F}_{t+}$).

Theorem 3.21 (Fundamental Theorem of Local Martingales)**.** *Assume $\{\mathcal{F}_t\}$ is complete and right-continuous. Suppose M is a cadlag local martingale and $c > 0$. Then there exist cadlag local martingales $\widetilde{M}$ and A such that the jumps of $\widetilde{M}$ are bounded by c, A is an FV process, and $M = \widetilde{M} + A$.*

A proof of the fundamental theorem of local martingales can be found in Section III.6 of [12]. Combining this theorem with the previous lemma gives the following corollary, which we will find useful because L^2-martingales are the starting point for developing stochastic integration.

Corollary 3.22. *Assume $\{\mathcal{F}_t\}$ is complete and right-continuous. Then a cadlag local martingale M can be written as a sum $M = \widetilde{M} + A$ of a cadlag local L^2-martingale $\widetilde{M}$ and a local martingale A that is an FV process.*

Definition 3.23. A cadlag process Y is a *semimartingale* if it can be written as $Y_t = Y_0 + M_t + V_t$ where M is a cadlag local martingale, V is a cadlag FV process, and $M_0 = V_0 = 0$.

By the previous corollary, we can always select the local martingale part of a semimartingale to be a local L^2-martingale. The normalization $M_0 = V_0 = 0$ does not exclude anything since we can always replace M with $M - M_0$ without losing either the local martingale or local L^2-martingale property, and also $V - V_0$ has the same total variation as V does.

Remark 3.24. (A look ahead.) The definition of a semimartingale appears ad hoc. But stochastic analysis will show us that the class of semimartingales has several attractive properties. (i) For a suitably bounded predictable process X and a semimartingale Y, the stochastic integral $\int X\,dY$ is again a semimartingale. (ii) $f(Y)$ is a semimartingale for a C^2 function f. The class of local L^2-martingales has property (i) but not (ii). In order to develop a sensible stochastic calculus that involves functions of processes, it is necessary to extend the class of processes considered from local martingales to semimartingales.

3.4. Quadratic variation for semimartingales

Quadratic variation and covariation were discussed in general in Section 2.2, and here we look at these notions for martingales, local martingales and semimartingales. We begin with the examples we already know.

Example 3.25 (Brownian motion)**.** From Proposition 2.42 and Exercise 2.26, for two independent Brownian motions B and Y, $[B]_t = t$ and $[B, Y]_t = 0$.

Example 3.26 (Poisson process)**.** Let N be a homogeneous rate α Poisson process, and $M_t = N_t - \alpha t$ the compensated Poisson process. Then $[M] = [N] = N$ by Corollary A.11. If $\widetilde{N}$ is an independent rate $\widetilde{\alpha}$ Poisson process with $\widetilde{M_t} = \widetilde{N_t} - \widetilde{\alpha}t$, $[M, \widetilde{M}] = 0$ by Lemma A.10 because with probability one M and $\widetilde{M}$ have no jumps in common.

Next a general existence theorem. A proof can be found in Section 2.3 of [6].

Theorem 3.27. *Let M be a right-continuous local martingale with respect to a filtration $\{\mathcal{F}_t\}$. Then the quadratic variation process $[M]$ exists in the sense of Definition 2.14. There is a version of $[M]$ with the following properties. $[M]$ is a real-valued, right-continuous, nondecreasing adapted process such that $[M]_0 = 0$.*

Suppose M is an L^2-martingale. Then the convergence in (2.11) for $Y = M$ holds also in L^1, namely for any $t \in \mathbf{R}_+$,

$$\lim_{n\to\infty} E\left|\sum_i (M_{t^n_{i+1}} - M_{t^n_i})^2 - [M]_t\right| = 0 \tag{3.14}$$

for any sequence of partitions $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{m(n)}^n = t\}$ *of* $[0, t]$ *with* $\operatorname{mesh}(\pi^n) = \max_i(t_{i+1}^n - t_i^n) \to 0$. *Furthermore,*

$$E([M]_t) = E(M_t^2 - M_0^2). \tag{3.15}$$

If M *is continuous, then so is* $[M]$.

We check that the quadratic variation of a stopped submartingale agrees with the stopped quadratic variation. Our proof works for local L^2 martingales which is sufficient for our needs. The result is true for all local martingales. The statement for all local martingales can be derived from the estimates in the proof of Theorem 3.27 in [6], specifically from limit (3.26) on page 70 in [6].

Lemma 3.28. *Let* M *be a right-continuous* L^2*-martingale or local* L^2*-martingale. Let* τ *be a stopping time. Then* $[M^\tau] = [M]^\tau$*, in the sense that these processes are indistinguishable.*

Proof. The processes $[M^\tau]$ and $[M]^\tau$ are right-continuous. Hence indistinguishability follows from proving almost sure equality at all fixed times.

Step 1. We start with a discrete stopping time τ whose values form an unbounded, increasing sequence $u_1 < u_2 < \cdots < u_j \nearrow \infty$. Fix t and consider a sequence of partitions $\pi^n = \{0 = t_0^n < t_1^n < \cdots < t_{m(n)}^n = t\}$ of $[0, t]$ with $\operatorname{mesh}(\pi^n) \to 0$ as $n \to \infty$. For any u_j,

$$\sum_i (M_{u_j \wedge t_{i+1}^n} - M_{u_j \wedge t_i^n})^2 \to [M]_{u_j \wedge t} \quad \text{in probability, as } n \to \infty.$$

We can replace the original sequence π^n with a subsequence along which this convergence is almost sure for all j. We denote this new sequence again by π^n. (The random variables above are the same for all j large enough so that $u_j > t$, so there are really only finitely many distinct j for which the limit needs to happen.)

Fix an ω at which the convergence happens. Let $u_j = \tau(\omega)$. Then the above limit gives

$$\begin{aligned}
[M^\tau]_t(\omega) &= \lim_{n\to\infty} \sum_i \big(M^\tau_{t_{i+1}^n}(\omega) - M^\tau_{t_i^n}(\omega)\big)^2 \\
&= \lim_{n\to\infty} \sum_i \big(M_{\tau \wedge t_{i+1}^n}(\omega) - M_{\tau \wedge t_i^n}(\omega)\big)^2 \\
&= \lim_{n\to\infty} \sum_i \big(M_{u_j \wedge t_{i+1}^n}(\omega) - M_{u_j \wedge t_i^n}(\omega)\big)^2 \\
&= [M]_{u_j \wedge t}(\omega) = [M]_{\tau \wedge t}(\omega) = [M]^\tau_t(\omega).
\end{aligned}$$

The meaning of the first equality sign above is that we know $[M^\tau]_t$ is given by this limit, since according to the existence theorem, $[M^\tau]_t$ is given as a limit in probability along any sequence of partitions with vanishing mesh.

We have shown that $[M^\tau]_t = [M]^\tau_t$ almost surely for a discrete stopping time τ.

Step 2. Let τ be an arbitrary stopping time, but assume M is an L^2-martingale. Let $\tau_n = 2^{-n}(\lfloor 2^n\tau\rfloor + 1)$ be the discrete approximation that converges to τ from the right. Apply (2.20) to $X = M^{\tau_n}$ and $Y = M^\tau$, take expectations, use (3.15), and apply Schwarz inequality to get

$$\begin{aligned}
&E\bigl\{\,|\,[M^{\tau_n}]_t - [M^\tau]_t\,|\,\bigr\}\\
&\le E\bigl\{[M^{\tau_n} - M^\tau]_t\bigr\} + 2E\bigl\{[M^{\tau_n} - M^\tau]_t^{1/2}[M^\tau]_t^{1/2}\bigr\}\\
&\le E\bigl\{(M_t^{\tau_n} - M_t^\tau)^2\bigr\} + 2E\bigl\{[M^{\tau_n} - M^\tau]_t\bigr\}^{1/2}E\bigl\{[M^\tau]_t\bigr\}^{1/2}\\
&= E\bigl\{(M_{\tau_n\wedge t} - M_{\tau\wedge t})^2\bigr\} + 2E\bigl\{(M_{\tau_n\wedge t} - M_{\tau\wedge t})^2\bigr\}^{1/2}E\bigl\{M^2_{\tau\wedge t}\bigr\}^{1/2}\\
&\le \bigl(E\{M^2_{\tau_n\wedge t}\} - E\{M^2_{\tau\wedge t}\}\bigr) + 2\bigl(E\{M^2_{\tau_n\wedge t}\} - E\{M^2_{\tau\wedge t}\}\bigr)^{1/2}E\bigl\{M_t^2\bigr\}^{1/2}.
\end{aligned}$$

In the last step we used (3.3) in two ways: For a martingale it gives equality, and so

$$\begin{aligned}
E\bigl\{(M_{\tau_n\wedge t} - M_{\tau\wedge t})^2\bigr\} &= E\{M^2_{\tau_n\wedge t}\} - 2E\{\,E(M_{\tau_n\wedge t}|\mathcal{F}_{\tau\wedge t})M_{\tau\wedge t}\} + E\{M^2_{\tau\wedge t}\}\\
&= E\{M^2_{\tau_n\wedge t}\} - E\{M^2_{\tau\wedge t}\}.
\end{aligned}$$

Second, we applied (3.3) to the submartingale M^2 to get

$$E\bigl\{M^2_{\tau\wedge t}\bigr\}^{1/2} \le E\bigl\{M_t^2\bigr\}^{1/2}.$$

The string of inequalities allows us to conclude that $[M^{\tau_n}]_t$ converges to $[M^\tau]_t$ in L^1 as $n \to \infty$, if we can show that

$$E\{M^2_{\tau_n\wedge t}\} \to E\{M^2_{\tau\wedge t}\}. \tag{3.16}$$

To argue this last limit, first note that by right-continuity, $M^2_{\tau_n\wedge t} \to M^2_{\tau\wedge t}$ almost surely. By optional stopping (3.6),

$$0 \le M^2_{\tau_n\wedge t} \le E(M_t^2|\mathcal{F}_{\tau_n\wedge t}).$$

This inequality and Lemma B.16 from the Appendix imply that the sequence $\{M^2_{\tau_n\wedge t} : n \in \mathbf{N}\}$ is uniformly integrable. Under uniform integrability, the almost sure convergence implies convergence of the expectations (3.16).

To summarize, we have shown that $[M^{\tau_n}]_t \to [M^\tau]_t$ in L^1 as $n \to \infty$. By Step 1, $[M^{\tau_n}]_t = [M]_{\tau_n\wedge t}$ which converges to $[M]_{\tau\wedge t}$ by right-continuity of the process $[M]$. Putting these together, we get the almost sure equality $[M^\tau]_t = [M]_{\tau\wedge t}$ for L^2-martingales.

Step 3. Lastly a localization step. Let $\{\sigma_k\}$ be stopping times such that $\sigma_k \nearrow \infty$ and M^{σ_k} is an L^2-martingale for each k. By Step 2

$$[M^{\sigma_k \wedge \tau}]_t = [M^{\sigma_k}]_{\tau \wedge t}.$$

On the event $\{\sigma_k > t\}$, throughout the time interval $[0, t]$, $M^{\sigma_k \wedge \tau}$ agrees with M^τ and M^{σ_k} agrees with M. Hence the corresponding sums of squared increments agree also. In the limits of vanishing mesh we have almost surely $[M^{\sigma_k \wedge \tau}]_t = [M^\tau]_t$, and also $[M^{\sigma_k}]_s = [M]_s$ for all $s \in [0, t]$ by right-continuity. We can take $s = \tau \wedge t$, and this way we get the required equality $[M^\tau]_t = [M]_{\tau \wedge t}$. □

Theorem 3.29. (a) *If M is a right-continuous L^2-martingale, then $M_t^2 - [M]_t$ is a martingale.*

(b) *If M is a right-continuous local L^2-martingale, then $M_t^2 - [M]_t$ is a local martingale.*

Proof. Part (a). Let $s < t$ and $A \in \mathcal{F}_s$. Let $0 = t_0 < \cdots < t_m = t$ be a partition of $[0, t]$, and assume that s is a partition point, say $s = t_\ell$.

$$E\big[\mathbf{1}_A\big(M_t^2 - M_s^2 - [M]_t + [M]_s\big)\big]$$

$$= E\bigg[\mathbf{1}_A\bigg(\sum_{i=\ell}^{m-1}(M_{t_{i+1}}^2 - M_{t_i}^2) - [M]_t + [M]_s\bigg)\bigg] \tag{3.17a}$$

$$= E\bigg[\mathbf{1}_A\bigg(\sum_{i=\ell}^{m-1}(M_{t_{i+1}} - M_{t_i})^2 - [M]_t + [M]_s\bigg)\bigg]$$

$$= E\bigg[\mathbf{1}_A\bigg(\sum_{i=0}^{m-1}(M_{t_{i+1}} - M_{t_i})^2 - [M]_t\bigg)\bigg] \tag{3.17b}$$

$$+ E\bigg[\mathbf{1}_A\bigg([M]_s - \sum_{i=0}^{\ell-1}(M_{t_{i+1}} - M_{t_i})^2\bigg)\bigg]. \tag{3.17c}$$

The second equality above follows from

$$E\big[M_{t_{i+1}}^2 - M_{t_i}^2 \;\big|\; \mathcal{F}_{t_i}\big] = E\big[(M_{t_{i+1}} - M_{t_i})^2 \;\big|\; \mathcal{F}_{t_i}\big].$$

To apply this, the expectation on line (3.17a) has to be taken apart, the conditioning applied to individual terms, and then the expectation put back together. Letting the mesh of the partition tend to zero makes the expectations on lines (3.17b)–(3.17c) vanish by the L^1 convergence in (2.11) for L^2-martingales.

In the limit we have

$$E\big[\mathbf{1}_A\big(M_t^2 - [M]_t\big)\big] = E\big[\mathbf{1}_A\big(M_s^2 - [M]_s\big)\big]$$

for an arbitrary $A \in \mathcal{F}_s$, which implies the martingale property.

(b) Let $X = M^2 - [M]$ for the local L^2-martingale M. Let $\{\tau_k\}$ be a localizing sequence for M. By part (a), $(M^{\tau_k})_t^2 - [M^{\tau_k}]_t$ is a martingale. Since $[M^{\tau_k}]_t = [M]_{t\wedge\tau_k}$ by Lemma 3.28, this martingale is the same as $M^2_{t\wedge\tau_k} - [M]_{t\wedge\tau_k} = X_t^{\tau_k}$. Thus $\{\tau_k\}$ is a localizing sequence for X. ☐

From Theorem 3.27 it follows also that the covariation $[M, N]$ of two right-continuous local martingales M and N exists. As a difference of increasing processes, $[M, N]$ is a finite variation process.

Lemma 3.30. *Let M and N be cadlag L^2-martingales or local L^2-martingales. Let τ be a stopping time. Then $[M^\tau, N] = [M^\tau, N^\tau] = [M, N]^\tau$.*

Proof. $[M^\tau, N^\tau] = [M, N]^\tau$ follows from the definition (2.12) and Lemma 3.28. For the first equality claimed, consider a partition of $[0, t]$. If $0 < \tau \le t$, let ℓ be the index such that $t_\ell < \tau \le t_{\ell+1}$. Then

$$\sum_i (M^\tau_{t_{i+1}} - M^\tau_{t_i})(N_{t_{i+1}} - N_{t_i}) = (M_\tau - M_{t_\ell})(N_{t_{\ell+1}} - N_\tau)\mathbf{1}_{\{0<\tau\le t\}}$$
$$+ \sum_i (M^\tau_{t_{i+1}} - M^\tau_{t_i})(N^\tau_{t_{i+1}} - N^\tau_{t_i}).$$

(If $\tau = 0$ the equality above is still true, for both sides vanish.) Let the mesh of the partition tend to zero. With cadlag paths, the term after the equality sign converges almost surely to $(M_\tau - M_{\tau-})(N_\tau - N_\tau)\mathbf{1}_{\{0<\tau\le t\}} = 0$. The convergence of the sums gives $[M^\tau, N] = [M^\tau, N^\tau]$. ☐

Theorem 3.31. (a) *If M and N are right-continuous L^2-martingales, then $MN - [M, N]$ is a martingale.*

(b) *If M and N are right-continuous local L^2-martingales, then $MN - [M, N]$ is a local martingale.*

Proof. Apply (2.14) to write

$$MN - [M, N] = \tfrac{1}{2}\{(M + N)^2 - [M + N]\} - \tfrac{1}{2}\{M^2 - [M]\} - \tfrac{1}{2}\{N^2 - [N]\}.$$

Both (a) and (b) now follow from Theorem 3.29. ☐

As the last issue we extend the existence results to semimartingales.

Corollary 3.32. *Let M be a cadlag local martingale, V a cadlag FV process, $M_0 = V_0 = 0$, and $Y = Y_0 + M + V$ the cadlag semimartingale. Then the cadlag quadratic variation process $[Y]$ exists and satisfies*

$$\begin{aligned}[Y]_t &= [M]_t + 2[M, V]_t + [V]_t \\ &= [M]_t + 2\sum_{s\in(0,t]} \Delta M_s \Delta V_s + \sum_{s\in(0,t]} (\Delta V_s)^2. \end{aligned} \tag{3.18}$$

Furthermore, $[Y^\tau] = [Y]^\tau$ *for any stopping time* τ *and the covariation* $[X, Y]$ *exists for any pair of cadlag semimartigales.*

Proof. We already know the existence and properties of $[M]$. According to Lemma A.10 the two sums on line (3.18) converge absolutely. Thus the process given by line (3.18) is a cadlag process. (Recall Example 1.13.) Theorem 3.27 and Lemma A.10 combined imply that line (3.18) is the limit (in probability) of sums $\sum_i (Y_{t_i} - Y_{t_{i-1}})^2$ as $\text{mesh}(\pi) \to 0$. Denote the process on line (3.18) temporarily by U_t. It follows from the limits that $U_s \le U_t$ a.s. for each pair of times $s < t$, and hence simultaneously for all rational $s < t$. By taking limits, the cadlag property of paths extends monotonicity from rational times to all times. Thus U is an increasing process and gives a version of $[Y]$ with nondecreasing paths. This proves the existence of $[Y]$.

$[Y^\tau] = [Y]^\tau$ follows by looking at line (3.18) term by term and by using Lemma 3.28. Since quadratic variation exists for semimartingales, so does $[X, Y] = [(X + Y)/2] - [(X - Y)/2]$. □

3.5. Doob-Meyer decomposition

In addition to the quadratic variation $[M]$ there is another increasing process with similar notation, the so-called *predictable quadratic variation* $\langle M \rangle$, associated to a square-integrable martingale M. We will not use $\langle M \rangle$ much in this text, except in Chapter 9 on stochastic partial differential equations. For the sake of completeness we address the relationship between $[M]$ and $\langle M \rangle$. It turns out that for continuous square-integrable martingales $[M]$ and $\langle M \rangle$ coincide, so in that context one can use them interchangeably. In particular, books that restrict their treatment of stochastic integration to continuous integrators need only discuss $\langle M \rangle$.

Throughout this section we work with a fixed probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$ assumed to satisfy the usual conditions (complete, right-continuous).

In order to state a precise definition we need to introduce the *predictable* σ*-algebra* $\mathcal{P}$ on the space $\mathbf{R}_+ \times \Omega$. $\mathcal{P}$ is the sub-σ-algebra of $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$ generated by *left*-continuous adapted processes. More precisely, $\mathcal{P}$ is generated by events of the form $\{(t, \omega) : X_t(\omega) \in B\}$ where X is an adapted, left-continuous process and $B \in \mathcal{B}_{\mathbf{R}}$. Such a process is measurable, even progressively measurable (Lemma 2.4), so these events lie in $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$. There are other ways of generating $\mathcal{P}$. For example, continuous processes would also do. Left-continuity has the virtue of focusing on the "predictability": if we know X_s for all $s < t$ then we can "predict" the value X_t. A thorough discussion of $\mathcal{P}$ has to wait till Section 5.1 where stochastic integration with respect to cadlag martingales is introduced. Any $\mathcal{P}$-measurable function $X : \mathbf{R}_+ \times \Omega \to \mathbf{R}$ is called a *predictable process.*

Here is the existence and uniqueness statement. It is a special case of the Doob-Meyer decomposition.

Theorem 3.33. *Assume the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions. Let M be a right-continuous square-integrable martingale. Then there is a unique predictable process $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a martingale.*

If M is a right-continuous local L^2-martingale, then there is a unique predictable process $\langle M \rangle$ such that $M^2 - \langle M \rangle$ is a local martingale.

Uniqueness above means uniqueness up to indistinguishability. $\langle M \rangle$ is called the *predictable quadratic variation.* For two such processes we can define the *predictable covariation* by $\langle M, N \rangle = \frac{1}{4}\langle M + N \rangle - \frac{1}{4}\langle M - N \rangle$. From Theorem 3.29 and the uniqueness of $\langle M \rangle$ it follows that if $[M]$ is predictable then $[M] = \langle M \rangle$.

Proposition 3.34. *Assume the filtration satisfies the usual conditions.*

(a) *Suppose M is a continuous square-integrable martingale. Then $\langle M \rangle = [M]$.*

(b) *Suppose M is a right-continuous square-integrable martingale with stationary independent increments: for all $s, t \geq 0$, $M_{s+t} - M_s$ is independent of $\mathcal{F}_s$ and has the same distribution as $M_t - M_0$. Then $\langle M \rangle_t = t \cdot E[M_1^2 - M_0^2]$.*

Proof. Part (a). By Theorems 3.27 and 3.29, $[M]$ is a continuous, increasing process such that $M^2 - [M]$ is a martingale. Continuity implies that $[M]$ is predictable. Uniqueness of $\langle M \rangle$ implies $\langle M \rangle = [M]$.

Part (b). The deterministic, continuous function $t \mapsto tE[M_1^2 - M_0^2]$ is predictable. For any $t > 0$ and integer k

$$E[M_{kt}^2 - M_0^2] = \sum_{j=0}^{k-1} E[M_{(j+1)t}^2 - M_{jt}^2] = \sum_{j=0}^{k-1} E[(M_{(j+1)t} - M_{jt})^2]$$
$$= kE[(M_t - M_0)^2] = kE[M_t^2 - M_0^2].$$

Using this twice, for any rational k/n,

$$E[M_{k/n}^2 - M_0^2] = kE[M_{1/n}^2 - M_0^2] = (k/n)E[M_1^2 - M_0^2].$$

Given an irrational $t > 0$, pick rationals $q_m \searrow t$. Fix $T \geq q_m$. By right-continuity of paths, $M_{q_m} \to M_t$ almost surely. Uniform integrability of $\{M_{q_m}^2\}$ follows by the submartingale property

$$0 \leq M_{q_m}^2 \leq E[M_T^2 | \mathcal{F}_{q_m}]$$

and Lemma B.16. Uniform integrability gives convergence of expectations $E[M_{q_m}^2] \to E[M_t^2]$. Applying this above gives

$$E[M_t^2 - M_0^2] = tE[M_1^2 - M_0^2].$$

Now we can check the martingale property.

$$\begin{aligned} E[M_t^2|\mathcal{F}_s] &= M_s^2 + E[M_t^2 - M_s^2|\mathcal{F}_s] = M_s^2 + E[(M_t - M_s)^2|\mathcal{F}_s] \\ &= M_s^2 + E[(M_{t-s} - M_0)^2] = M_s^2 + E[M_{t-s}^2 - M_0^2] \\ &= M_s^2 + (t-s)E[M_1^2 - M_0^2]. \end{aligned}$$ □

This proposition was tailored to handle our two main examples.

Example 3.35. For a standard Brownian motion $\langle B\rangle_t = [B]_t = t$. For a compensated Poisson process $M_t = N_t - \alpha t$,

$$\langle M\rangle_t = tE[M_1^2] = tE[(N_1 - \alpha)^2] = \alpha t.$$

We continue this discussion for a while, although what comes next makes no appearance later on. It is possible to introduce $\langle M\rangle$ without reference to $\mathcal{P}$. We do so next, and then state the Doob-Meyer decomposition. Recall Definition 2.17 of an increasing process

Definition 3.36. An increasing process A is *natural* if for every bounded cadlag martingale $M = \{M(t) : 0 \le t < \infty\}$,

$$E\int_{(0,t]} M(s)dA(s) = E\int_{(0,t]} M(s-)dA(s) \quad \text{for } 0 < t < \infty. \tag{3.19}$$

The expectation–integral on the left in condition (3.36) is interpreted in the following way. First for a fixed ω, the function $s \mapsto M(s,\omega)$ is integrated against the (positive) Lebesgue-Stieltjes measure of the function $s \mapsto A(s,\omega)$. The resulting quantity is a measurable function of ω (Exercise 3.13). Then this function is averaged over the probability space. Similar interpretation on the right in (3.36), of course. The expectations in (3.36) can be infinite. For a fixed ω,

$$\left|\int_{(0,t]} M(s)dA(s)\right| \le \sup_s |M(s)|A(t) < \infty$$

so the random integral is finite.

Lemma 3.37. *Let A be an increasing process and M a bounded cadlag martingale. If A is continuous then (3.19) holds.*

Proof. A cadlag path $s \mapsto M(s,\omega)$ has at most countably many discontinuities. If A is continuous, the Lebesgue-Stieltjes measure Λ_A gives no mass to singletons: $\Lambda_A\{s\} = A(s) - A(s-) = 0$, and hence no mass to a countable set either. Consequently

$$\int_{(0,t]} \big(M(s) - M(s-)\big)\, dA(s) = 0. \qquad \square$$

Much more is true. In fact, an increasing process is natural if and only if it is predictable. A proof can be found in Chapter 25 of [10]. Consequently the characterizing property of $\langle M\rangle$ can be taken to be naturalness rather than predictability.

Definition 3.38. For $0 < u < \infty$, let $\mathcal{T}_u$ be the collection of stopping times τ that satisfy $\tau \le u$. A process X is of *class DL* if the random variables $\{X_\tau : \tau \in \mathcal{T}_u\}$ are uniformly integrable for each $0 < u < \infty$.

The main example is the following.

Lemma 3.39. *A right-continuous nonnegative submartingale is of class DL.*

Proof. Let X be a right-continuous nonnegative submartingale, and $0 < u < \infty$. By (3.3)

$$0 \le X_\tau \le E[X_u|\mathcal{F}_\tau].$$

By Lemma B.16 the collection of all conditional expectations on the right is uniformly integrable. Consequently these inequalities imply the uniform integrability of the collection $\{X_\tau : \tau \in \mathcal{T}_u\}$. $\square$

Here is the main theorem. For a proof see Theorem 1.4.10 in [11].

Theorem 3.40. (Doob-Meyer Decomposition) *Assume the underlying filtration is complete and right-continuous. Let X be a right-continuous submartingale of class DL. Then there is an increasing natural process A, unique up to indistinguishability, such that $X - A$ is a martingale.*

Let us return to a right-continuous square-integrable martingale M. We can now equivalently define $\langle M\rangle$ as the unique increasing, natural process such that $M_t^2 - \langle M\rangle_t$ is a martingale, given by the Doob-Meyer decomposition.

3.6. Spaces of martingales

Stochastic integrals will be constructed as limits. To get the desirable path properties for the integrals, it is convenient to take these limits in a space of stochastic processes rather than simply in terms of individual random variables. This is the purpose of introducing two spaces of martingales.

Given a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$, let $\mathcal{M}_2$ denote the space of square-integrable cadlag martingales on this space with respect to $\{\mathcal{F}_t\}$. The subspace of $\mathcal{M}_2$ of martingales with continuous paths is $\mathcal{M}_2^c$. $\mathcal{M}_2$ and $\mathcal{M}_2^c$ are both linear spaces.

We measure the size of a martingale $M \in \mathcal{M}_2$ with the quantity

$$\|M\|_{\mathcal{M}_2} = \sum_{k=1}^{\infty} 2^{-k}\big(1 \wedge \|M_k\|_{L^2(P)}\big). \tag{3.20}$$

$\|M_k\|_{L^2(P)} = E[\,|M_k|^2]^{1/2}$ is the L^2 norm of M_k. $\|\cdot\|_{\mathcal{M}_2}$ is not a norm because $\|\alpha M\|_{\mathcal{M}_2}$ is not necessarily equal to $|\alpha| \cdot \|M\|_{\mathcal{M}_2}$ for a real number α. But the triangle inequality

$$\|M + N\|_{\mathcal{M}_2} \le \|M\|_{\mathcal{M}_2} + \|N\|_{\mathcal{M}_2}$$

is valid, and follows from the triangle inequality of the L^2 norm and because

$$1 \wedge (a + b) \le 1 \wedge a \;+\; 1 \wedge b \qquad \text{for } a, b \ge 0.$$

Hence we can define a metric, or distance function, between two martingales $M, N \in \mathcal{M}_2$ by

$$d_{\mathcal{M}_2}(M, N) = \|M - N\|_{\mathcal{M}_2}. \tag{3.21}$$

A technical issue arises here. A basic property of a metric is that the distance between two elements is zero iff these two elements coincide. But with the above definition we have $d_{\mathcal{M}_2}(M, N) = 0$ if M and N are indistinguishable, even if they are not exactly equal as functions. So if we were to precisely follow the axiomatics of metric spaces, indistinguishable martingales should actually be regarded as equal. The mathematically sophisticated way of doing this is to regard $\mathcal{M}_2$ not as a space of processes but as a space of equivalence classes

$$\{M\} = \{N : N \text{ is a square-integrable cadlag martingale on } (\Omega, \mathcal{F}, P), \\ \text{and } M \text{ and } N \text{ are indistinguishable}\}$$

Fortunately this technical point does not cause any difficulties. We shall continue to regard the elements of $\mathcal{M}_2$ as processes in our discussion, and remember that two indistinguishable processes are really two "representatives" of the same underlying process.

Theorem 3.41. *Assume the underlying probability space $(\Omega, \mathcal{F}, P)$ and the filtration $\{\mathcal{F}_t\}$ complete. Let indistinguishable processes be interpreted as equal. Then $\mathcal{M}_2$ is a complete metric space under the metric $d_{\mathcal{M}_2}$. The subspace $\mathcal{M}_2^c$ is closed, and hence a complete metric space in its own right.*

Proof. Suppose $M \in \mathcal{M}_2$ and $\|M\|_{\mathcal{M}_2} = 0$. Then $E[M_k^2] = 0$ for each $k \in \mathbf{N}$. Since M_t^2 is a submartingale, $E[M_t^2] \le E[M_k^2]$ for $t \le k$, and consequently $E[M_t^2] = 0$ for all $t \geq 0$. In particular, for each fixed t, $P\{M_t = 0\} = 1$. A countable union of null sets is a null set, and so there exists an event $\Omega_0 \subseteq \Omega$ such that $P(\Omega_0) = 1$ and $M_q(\omega) = 0$ for all $\omega \in \Omega_0$ and $q \in \mathbf{Q}_+$. By right-continuity, then $M_t(\omega) = 0$ for all $\omega \in \Omega_0$ and $t \in \mathbf{R}_+$. This shows that M is indistinguishable from the identically zero process.

The above paragraph shows that $d_{\mathcal{M}_2}(M, N) = 0$ implies $M = N$, in the sense that M and N are indistinguishable. We already observed above that $d_{\mathcal{M}_2}$ satisfies the triangle inequality. The remaining property of a metric is the symmetry $d_{\mathcal{M}_2}(M, N) = d_{\mathcal{M}_2}(N, M)$ which is true by the definition.

To check completeness, let $\{M^{(n)} : n \in \mathbf{N}\}$ be a Cauchy sequence in the metric $d_{\mathcal{M}_2}$ in the space $\mathcal{M}_2$. We need to show that there exists $M \in \mathcal{M}_2$ such that $\|M^{(n)} - M\|_{\mathcal{M}_2} \to 0$ as $n \to \infty$.

For any $t \leq k \in \mathbf{N}$, first because $(M_t^{(m)} - M_t^{(n)})^2$ is a submartingale, and then by the definition (3.20),

$$\begin{aligned} 1 \wedge E\big[(M_t^{(m)} - M_t^{(n)})^2\big]^{1/2} &\leq 1 \wedge E\big[(M_k^{(m)} - M_k^{(n)})^2\big]^{1/2} \\ &\leq 2^k \|M^{(m)} - M^{(n)}\|_{\mathcal{M}_2}. \end{aligned}$$

It follows that for each fixed t, $\{M_t^{(n)}\}$ is a Cauchy sequence in $L^2(P)$. By the completeness of the space $L^2(P)$, for each $t \geq 0$ there exists a random variable $Y_t \in L^2(P)$ defined by the mean-square limit

$$\lim_{n\to\infty} E\big[(M_t^{(n)} - Y_t)^2\big] = 0. \tag{3.22}$$

Take $s < t$ and $A \in \mathcal{F}_s$. Let $n \to \infty$ in the equality $E[\mathbf{1}_A M_t^{(n)}] = E[\mathbf{1}_A M_s^{(n)}]$. Mean-square convergence guarantees the convergence of the expectations, and in the limit

$$E[\mathbf{1}_A Y_t] = E[\mathbf{1}_A Y_s]. \tag{3.23}$$

We could already conclude here that $\{Y_t\}$ is a martingale, but $\{Y_t\}$ is not our ultimate limit because we need the cadlag path property.

To get a cadlag limit we use a Borel-Cantelli argument. By inequality (3.8),

$$P\Big\{\sup_{0\leq t\leq k} |M_t^{(m)} - M_t^{(n)}| \geq \varepsilon\Big\} \leq \varepsilon^{-2} E\big[(M_k^{(m)} - M_k^{(n)})^2\big]. \tag{3.24}$$

This enables us to choose a subsequence $\{n_k\}$ such that

$$P\Big\{\sup_{0\leq t\leq k} |M_t^{(n_{k+1})} - M_t^{(n_k)}| \geq 2^{-k}\Big\} \leq 2^{-k}. \tag{3.25}$$

To achieve this, start with $n_0 = 1$, and assuming n_{k-1} has been chosen, pick $n_k > n_{k-1}$ so that

$$\|M^{(m)} - M^{(n)}\|_{\mathcal{M}_2} \leq 2^{-3k}$$

for $m, n \geq n_k$. Then for $m \geq n_k$,

$$1 \wedge E\big[(M_k^{(m)} - M_k^{(n_k)})^2\big]^{1/2} \leq 2^k \|M^{(m)} - M^{(n_k)}\|_{\mathcal{M}_2} \leq 2^{-2k},$$

and the minimum with 1 is superfluous since $2^{-2k} < 1$. Substituting this back into (3.24) with $\varepsilon = 2^{-k}$ gives (3.25) with 2^{-2k} on the right-hand side.

By the Borel-Cantelli lemma, there exists an event Ω_1 with $P(\Omega_1) = 1$ such that for $\omega \in \Omega_1$,

$$\sup_{0 \leq t \leq k} |M_t^{(n_{k+1})}(\omega) - M_t^{(n_k)}(\omega)| < 2^{-k}$$

for all but finitely many k's. It follows that the sequence of cadlag functions $t \mapsto M_t^{(n_k)}(\omega)$ are Cauchy in the uniform metric over any bounded time interval $[0, T]$. By Lemma A.5 in the Appendix, for each $T < \infty$ there exists a cadlag process $\{N_t^{(T)}(\omega) : 0 \leq t \leq T\}$ such that $M_t^{(n_k)}(\omega)$ converges to $N_t^{(T)}(\omega)$ uniformly on the time interval $[0, T]$, as $k \to \infty$, for any $\omega \in \Omega_1$. $N_t^{(S)}(\omega)$ and $N_t^{(T)}(\omega)$ must agree for $t \in [0, S \wedge T]$, since both are limits of the same sequence. Thus we can define one cadlag function $t \mapsto M_t(\omega)$ on $\mathbf{R}_+$ for $\omega \in \Omega_1$, such that $M_t^{(n_k)}(\omega)$ converges to $M_t(\omega)$ uniformly on each bounded time interval $[0, T]$. To have M defined on all of Ω, set $M_t(\omega) = 0$ for $\omega \notin \Omega_1$.

The event Ω_1 lies in $\mathcal{F}_t$ by the assumption of completeness of the filtration. Since $M_t^{(n_k)} \to M_t$ on Ω_1 while $M_t = 0$ on Ω_1^c, it follows that M_t is $\mathcal{F}_t$-measurable. The almost sure limit M_t and the L^2 limit Y_t of the sequence $\{M_t^{(n_k)}\}$ must coincide almost surely. Consequently (3.23) becomes

$$E[\mathbf{1}_A M_t] = E[\mathbf{1}_A M_s] \tag{3.26}$$

for all $A \in \mathcal{F}_s$ and gives the martingale property for M.

To summarize, M is now a square-integrable cadlag martingale, in other words an element of $\mathcal{M}_2$. The final piece, namely $\|M^{(n)} - M\|_{\mathcal{M}_2} \to 0$, follows because we can replace Y_t by M_t in (3.22) due to the almost sure equality $M_t = Y_t$.

If all $M^{(n)}$ are continuous martingales, the uniform convergence above produces a continuous limit M. This shows that $\mathcal{M}_2^c$ is a closed subspace of $\mathcal{M}_2$ under the metric $d_{\mathcal{M}_2}$. ☐

By adapting the argument above from equation (3.24) onwards, we get this useful consequence of convergence in $\mathcal{M}_2$.

Lemma 3.42. *Suppose $\|M^{(n)} - M\|_{\mathcal{M}_2} \to 0$ as $n \to \infty$. Then for each $T < \infty$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\Big\{ \sup_{0 \leq t \leq T} |M_t^{(n)} - M_t| \geq \varepsilon \Big\} = 0. \tag{3.27}$$

Furthermore, there exists a subsequence $\{M^{(n_k)}\}$ *and an event* Ω_0 *such that* $P(\Omega_0) = 1$ *and for each* $\omega \in \Omega_0$ *and* $T < \infty$,

$$\lim_{n\to\infty} \sup_{0\le t\le T} \bigl|M_t^{(n_k)}(\omega) - M_t(\omega)\bigr| = 0. \tag{3.28}$$

When (3.27) holds for all $T < \infty$ and $\varepsilon > 0$, it is called *uniform convergence in probability on compact intervals.*

We shall write $\mathcal{M}_{2,\mathrm{loc}}$ for the space of cadlag local L^2-martingales with respect to a given filtration $\{\mathcal{F}_t\}$ on a given probability space $(\Omega, \mathcal{F}, P)$. We do not introduce a distance function on this space.

Exercises

Exercise 3.1. Create a simple example that shows that $E(M_\tau|\mathcal{F}_\sigma) = M_\sigma$ cannot hold for general random times $\sigma \le \tau$ that are not stopping times.

Exercise 3.2. Let B be a standard Brownian motion and $\tau = \inf\{t \ge 0 : B_t = 1\}$. Show that $P(\tau < \infty) = 1$, $E(B_{\tau\wedge n}) = E(B_0)$ for all $n \in \mathbf{N}$ but $E(B_\tau) = E(B_0)$ fails. In other words, optional stopping does not work for all stopping times.

Exercise 3.3. Let τ be a stopping time and M a right-continuous martingale. Corollaries 3.8 and 3.9 imply that M^τ is a martingale for both filtrations $\{\mathcal{F}_t\}$ and $\{\mathcal{F}_{t\wedge\tau}\}$. This exercise shows that any martingale with respect to $\{\mathcal{F}_{t\wedge\tau}\}$ is also a martingale with respect to $\{\mathcal{F}_t\}$.

So suppose $\{X_t\}$ is a martingale with respect to $\{\mathcal{F}_{t\wedge\tau}\}$. That is, X_t is $\mathcal{F}_{t\wedge\tau}$-measurable and integrable, and $E(X_t\,|\,\mathcal{F}_{s\wedge\tau}) = X_s$ for $s < t$.

(a) Show that for $s < t$, $\mathbf{1}\{\tau \le s\}X_t = \mathbf{1}\{\tau \le s\}X_s$. *Hint.* $\mathbf{1}\{\tau \le s\}X_t$ is $\mathcal{F}_{s\wedge\tau}$-measurable. Multiply the martingale property by $\mathbf{1}\{\tau \le s\}$.

(b) Show that $\{X_t\}$ is a martingale with respect to $\{\mathcal{F}_t\}$. *Hint.* With $s < t$ and $A \in \mathcal{F}_s$, start with $E(\mathbf{1}_A X_t) = E(\mathbf{1}_A\mathbf{1}\{\tau \le s\}X_t) + E(\mathbf{1}_A\mathbf{1}\{\tau > s\}X_t)$.

Exercise 3.4. (Brownian motion with a random speed.) Let B_t be a standard Brownian motion with filtration $\mathcal{F}_t^B = \sigma\{B_s : s \in [0,t]\}$. Let U be a nonnegative finite random variable, independent of $\mathcal{F}_\infty^B$, and such that $E(U^{1/2}) = \infty$. Define the filtration $\mathcal{G}_t = \sigma\{\mathcal{F}_t^B, \sigma(U)\}$. Show that for each $s \ge 0$, sU is a stopping time under filtration $\{\mathcal{G}_t\}$. Let $X_t = B_{tU}$, a process adapted to $\mathcal{F}_t = \mathcal{G}_{tU}$. ($\mathcal{G}_{tU}$ is defined as in (2.4).) Show that X_t is not a martingale but it is a local martingale. *Hints.* You need Exercise (2.22). Compute $E|X_t|$. There are localizing stopping times that take only values 0 and ∞, or you can look at Exercise 3.6 below.

Exercise 3.5. (a) Suppose X is an adapted process and $\tau_1, \dots, \tau_k$ stopping times such that $X^{\tau_1}, \dots, X^{\tau_k}$ are all martingales. Show that then $X^{\tau_1 \vee \cdots \vee \tau_k}$ is a martingale. *Hint.* In the case $k = 2$ write $X^{\tau_1 \vee \tau_2}$ in terms of X^{τ_1}, X^{τ_2} and $X^{\tau_1 \wedge \tau_2}$.

(b) Let M be an adapted process, and suppose there exists a sequence of stopping times $\rho_n \nearrow \infty$ (a.s.) such that M^{ρ_n} is a local martingale for each n. Show that then M is a local martingale.

Exercise 3.6. Let X be a continuous local martingale. Show that X can be localized by the stopping times $\nu_n = \inf\{t \ge 0 : |X_t| \ge n\}$.

Exercise 3.7. Let X be a nonnegative local martingale such that $EX_0 < \infty$. Show that then X is a supermartingale, and X is a martingale iff $EX_t = EX_0$ for all $t \ge 0$. *Hint.* Consider processes of the type $Y_t = X_{\tau_n \wedge t} \wedge K$.

Exercise 3.8. Let M be a right-continuous local martingale such that $M_t^* \in L^1(P)$ for all $t \in \mathbf{R}_+$. Show that then M is a martingale. *Hint.* Let $n \to \infty$ in the equality $E[\mathbf{1}_A M_{t\wedge\tau_n}] = E[\mathbf{1}_A M_{s\wedge\tau_n}]$ for $s < t$ and $A \in \mathcal{F}_s$.

Exercise 3.9. Let M be a local martingale with localizing sequence $\{\tau_k\}$. Suppose that for each $t \in \mathbf{R}_+$, the sequence $\{M_{t\wedge\tau_n}\}_{n\in\mathbf{N}}$ is uniformly integrable. Show that then M is a martingale. Same hint as above.

Exercise 3.10. Let M be a local martingale with localizing sequence $\{\tau_k\}$. Show that $M^{n\wedge\tau_n}$ is uniformly integrable (in other words, that the family of random variables $\{M_t^{n\wedge\tau_n} : t \in \mathbf{R}_+\}$ is uniformly integrable). *Hint.* Use Lemma B.16.

Exercise 3.11. Let M be some process and $X_t = M_t - M_0$. Show that if M is a local martingale, then so is X, but the converse is not true. For the counterexample, consider simply $M_t = \xi$ for a fixed random variable ξ.

Exercise 3.12. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $\mathcal{N} = \{N \in \mathcal{F} : P(N) = 0\}$ the class of null sets, and take a random variable $X \in L^1(P)$ but not in $L^2(P)$. For $t \in \mathbf{R}_+$ define

$$\mathcal{F}_t = \begin{cases} \sigma(\mathcal{N}), & 0 \le t < 1 \\ \mathcal{F}, & t \ge 1 \end{cases} \quad \text{and} \quad M_t = \begin{cases} EX, & 0 \le t < 1 \\ X, & t \ge 1. \end{cases}$$

Then $\{\mathcal{F}_t\}$ satisfies the usual conditions and M is a martingale but not a local L^2 martingale. *Hint.* Show that $\{\tau < 1\} \in \sigma(\mathcal{N})$ for any stopping time τ.

Exercise 3.13. Let A be an increasing process, and $\phi : \mathbf{R}_+ \times \Omega \to \mathbf{R}$ a bounded $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable function. Let $T < \infty$. Show that

$$g_\phi(\omega) = \int_{(0,T]} \phi(t, \omega)\, dA_t(\omega)$$

is an $\mathcal{F}$-measurable function. Show also that, for any $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable nonnegative function $\phi : \mathbf{R}_+ \times \Omega \to \mathbf{R}_+$,

$$g_\phi(\omega) = \int_{(0,\infty)} \phi(t,\omega)\, dA_t(\omega)$$

is an $\mathcal{F}$-measurable function. The integrals are Lebesgue-Stieltjes integrals, evaluated separately for each ω. The only point in separating the two cases is that if ϕ takes both positive and negative values, the integral over the entire interval $[0,\infty)$ might not be defined.

Hint. One can start with $\phi(t,\omega) = \mathbf{1}_{(a,b]\times\Gamma}(t,\omega)$ for $0 \le a < b < \infty$ and $\Gamma \in \mathcal{F}$. Then apply Theorem B.4 from the Appendix.

Exercise 3.14. Let $N = \{N(t) : 0 \le t < \infty\}$ be a homogeneous rate α Poisson process with respect to $\{\mathcal{F}_t\}$ and $M_t = N_t - \alpha t$ the compensated Poisson process. We have seen that the quadratic variation is $[M]_t = N_t$ while $\langle M \rangle_t = \alpha t$. It follows that N cannot be a natural increasing process. In this exercise you show that the naturalness condition fails for N.

(a) Let $\lambda > 0$. Show that

$$X(t) = \exp\{-\lambda N(t) + \alpha t(1 - e^{-\lambda})\}$$

is a martingale.

(b) Show that N is not a natural increasing process, by showing that for X defined above, the condition

$$E\int_{(0,t]} X(s)\, dN(s) = E\int_{(0,t]} X(s-)\, dN(s)$$

fails. (In case you protest that X is not a bounded martingale, fix $T > t$ and consider $X(s \wedge T)$.)

Chapter 4

Stochastic Integral with respect to Brownian Motion

As an introduction to stochastic integration we develop the stochastic integral with respect to Brownian motion. This can be done with fewer technicalities than are needed for integrals with respect to general cadlag martingales, so the basic ideas of stochastic integration are in clearer view. The same steps will be repeated in the next chapter in the development of the more general integral. For this reason we leave the routine verifications in this chapter as exercises. We develop only enough properties of the integral to enable us to get to the point where the integral of local integrands is defined. This chapter can be skipped without loss. Only Lemma 4.2 is referred to later in Section 5.5. This lemma is of technical nature and can be read independently of the rest of the chapter.

Throughout this chapter, $(\Omega, \mathcal{F}, P)$ is a fixed probability space with a filtration $\{\mathcal{F}_t\}$, and $B = \{B_t\}$ is a standard one-dimensional Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$ (Definition 2.26). We assume that $\mathcal{F}$ and each $\mathcal{F}_t$ contains all subsets of events of probability zero, an assumption that entails no loss of generality as explained in Section 2.1.

To begin, let us imagine that we are trying to define the integral $\int_0^t B_s\,dB_s$ through an approximation by Riemann sums. The next calculation reveals that, contrary to the familiar Riemann and Stieltjes integrals with reasonably regular functions, the choice of point of evaluation in a partition interval matters.

Lemma 4.1. *Fix a number $u \in [0,1]$. Given a partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{m(\pi)} = t\}$, let $s_i = (1-u)t_i + ut_{i+1}$, and define*

$$S(\pi) = \sum_{i=0}^{m(\pi)-1} B_{s_i}(B_{t_{i+1}} - B_{t_i}).$$

Then

$$\lim_{\text{mesh}(\pi)\to 0} S(\pi) = \tfrac{1}{2}B_t^2 - \tfrac{1}{2}t + ut \quad \text{in } L^2(P).$$

Proof. First check the algebra identity

$$b(a-c) = \frac{a^2}{2} - \frac{c^2}{2} - \frac{(a-c)^2}{2} + (b-c)^2 + (a-b)(b-c).$$

Applying this,

$$\begin{aligned} S(\pi) &= \tfrac{1}{2}B_t^2 - \tfrac{1}{2}\sum_i (B_{t_{i+1}} - B_{t_i})^2 + \sum_i (B_{s_i} - B_{t_i})^2 \\ &\qquad + \sum_i (B_{t_{i+1}} - B_{s_i})(B_{s_i} - B_{t_i}) \\ &\equiv \tfrac{1}{2}B_t^2 - S_1(\pi) + S_2(\pi) + S_3(\pi) \end{aligned}$$

where the last equality defines the sums $S_1(\pi)$, $S_2(\pi)$, and $S_3(\pi)$. By Proposition 2.42,

$$\lim_{\text{mesh}(\pi)\to 0} S_1(\pi) = \tfrac{1}{2}t \quad \text{in } L^2(P).$$

For the second sum,

$$E[S_2(\pi)] = \sum_i (s_i - t_i) = u\sum_i (t_{i+1} - t_i) = ut,$$

and

$$\begin{aligned} \text{Var}[S_2(\pi)] &= \sum_i \text{Var}[(B_{s_i} - B_{t_i})^2] = 2\sum_i (s_i - t_i)^2 \\ &\le 2\sum_i (t_{i+1} - t_i)^2 \le 2t\,\text{mesh}(\pi) \end{aligned}$$

which vanishes as $\text{mesh}(\pi) \to 0$. The factor 2 above comes from Gaussian properties: if X is a mean zero normal with variance σ^2, then

$$\text{Var}[X^2] = E[X^4] - \big(E[X^2]\big)^2 = 3\sigma^4 - \sigma^4 = 2\sigma^2.$$

The vanishing of the variance of $S_2(\pi)$ as $\text{mesh}(\pi) \to 0$ is equivalent to

$$\lim_{\text{mesh}(\pi)\to 0} S_2(\pi) = ut \quad \text{in } L^2(P).$$

Lastly, we show that $S_3(\pi)$ vanishes in $L^2(P)$ as $\text{mesh}(\pi) \to 0$.

$$\begin{aligned} E[S_3(\pi)^2] &= E\Big[\Big(\sum_i (B_{t_{i+1}} - B_{s_i})(B_{s_i} - B_{t_i})\Big)^2\Big] \\ &= \sum_i E\big[(B_{t_{i+1}} - B_{s_i})^2(B_{s_i} - B_{t_i})^2\big] \\ &\quad + \sum_{i\neq j} E\big[(B_{t_{i+1}} - B_{s_i})(B_{s_i} - B_{t_i})(B_{t_{j+1}} - B_{s_j})(B_{s_j} - B_{t_j})\big] \\ &= \sum_i (t_{i+1} - s_i)(s_i - t_i) \le \sum_i (t_{i+1} - t_i)^2 \le t\,\text{mesh}(\pi) \end{aligned}$$

which again vanishes as $\text{mesh}(\pi) \to 0$. □

According to Proposition 2.28, there is a unique choice that makes the limit of $S(\pi)$ into a martingale, namely $u = 0$, in other words taking $s_i = t_i$, the initial point of the partition interval. This is the choice for the *Itô integral*. After developing some background we revisit this calculation in Example 4.9 and establish the Itô integral

$$\int_0^t B_s\,dB_s = \tfrac{1}{2}B_t^2 - \tfrac{1}{2}t.$$

The choice $u = \frac{1}{2}$ leads to the *Stratonovich integral* given by

$$\int_0^t B_s \circ dB_s = \tfrac{1}{2}B_t^2.$$

A virtue of the Stratonovich integral is that the rules of ordinary calculus apply, as in the example above. But for developing the theory the Itô integral reigns supreme. We shall revisit the Stratonovich integral in some exercises later on.

We turn to develop the Itô stochastic integral with respect to Brownian motion. The first issue is to describe the spaces of stochastic processes X that serve as integrands in the integral $\int_0^t X\,dB$.

For a measurable process X, the L^2-norm over the set $[0,T] \times \Omega$ is

$$\|X\|_{L^2([0,T]\times\Omega)} = \Big(E\int_{[0,T]} |X(t,\omega)|^2 dt\Big)^{1/2}. \tag{4.1}$$

Let $\mathcal{L}_2(B)$ denote the collection of all measurable, adapted processes X such that

$$\|X\|_{L^2([0,T]\times\Omega)} < \infty$$

for all $T < \infty$. A metric on $\mathcal{L}_2(B)$ is defined by $d_{\mathcal{L}_2}(X,Y) = \|X - Y\|_{\mathcal{L}_2(B)}$ where

$$\|X\|_{\mathcal{L}_2(B)} = \sum_{k=1}^{\infty} 2^{-k}\bigl(1 \wedge \|X\|_{L^2([0,k]\times\Omega)}\bigr). \tag{4.2}$$

As for $\|\cdot\|_{\mathcal{M}_2}$ in Section 3.6 we use the norm notation even though $\|\cdot\|_{\mathcal{L}_2(B)}$ is not a genuine norm. The triangle inequality

$$\|X + Y\|_{\mathcal{L}_2(B)} \le \|X\|_{\mathcal{L}_2(B)} + \|Y\|_{\mathcal{L}_2(B)}$$

is valid, and this gives the triangle inequality

$$d_{\mathcal{L}_2}(X,Y) \le d_{\mathcal{L}_2}(X,Z) + d_{\mathcal{L}_2}(Z,Y)$$

required for $d_{\mathcal{L}_2}(X,Y)$ to be a genuine metric.

To have a metric, one also needs the property $d_{\mathcal{L}_2}(X,Y) = 0$ iff $X = Y$. We have to adopt the point of view that two processes X and Y are considered "equal" if the set of points (t,ω) where $X(t,\omega) \ne Y(t,\omega)$ has $m \otimes P$-measure zero. Equivalently,

$$\int_0^\infty P\{X(t) \ne Y(t)\}\,dt = 0. \tag{4.3}$$

In particular processes that are indistinguishable, or modifications of each other have to be considered equal under this interpretation.

The symmetry $d_{\mathcal{L}_2}(X,Y) = d_{\mathcal{L}_2}(Y,X)$ is immediate from the definition. So with the appropriate meaning assigned to equality, $\mathcal{L}_2(B)$ is a metric space. Convergence $X_n \to X$ in $\mathcal{L}_2(B)$ is equivalent to $X_n \to X$ in $L^2([0,T]\times\Omega)$ for each $T < \infty$.

The symbol B reminds us that $\mathcal{L}_2(B)$ is a space of integrands for stochastic integration with respect to Brownian motion.

The finite mean square requirement for membership in $\mathcal{L}_2(B)$ is restrictive. For example, it excludes some processes of the form $f(B_t)$ where f is a smooth but rapidly growing function. Consequently from $\mathcal{L}_2(B)$ we move to a wider class of processes denoted by $\mathcal{L}(B)$, where the mean square requirement is imposed only locally and only on integrals over the time variable. Precisely, $\mathcal{L}(B)$ is the class of measurable, adapted processes X such that

$$P\Bigl\{\omega : \int_0^T X(t,\omega)^2\,dt < \infty \quad \text{for all } T < \infty\Bigr\} = 1. \tag{4.4}$$

This will be the class of processes X for which the stochastic integral process with respect to Brownian motion, denoted by

$$(X \cdot B)_t = \int_0^t X_s\,dB_s$$

will ultimately be defined.

The development of the integral starts from a class of processes for which the integral can be written down directly. There are several possible starting places. Here is our choice.

A *simple predictable process* is a process of the form

$$X(t,\omega) = \xi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{n-1} \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t) \tag{4.5}$$

where n is finite integer, $0 = t_0 = t_1 < t_2 < \cdots < t_n$ are time points, and for $0 \le i \le n-1$, ξ_i is a bounded $\mathcal{F}_{t_i}$-measurable random variable on $(\Omega, \mathcal{F}, P)$. Predictability refers to the fact that the value X_t can be "predicted" from $\{X_s : s < t\}$. Here this point is rather simple because X is left-continuous so $X_t = \lim X_s$ as $s \nearrow t$. In the next chapter we need to deal seriously with the notion of predictability but in this chapter it is not really needed. We use the term only to be consistent with what comes later. The value ξ_0 at $t = 0$ is irrelevant both for the stochastic integral of X and for approximating general processes. We include it so that the value $X(0,\omega)$ is not artificially restricted.

A key point is that processes in $\mathcal{L}_2(B)$ can be approximated by simple predictable processes in the $\mathcal{L}_2(B)$-distance. We split this approximation into two steps.

Lemma 4.2. *Suppose X is a bounded, measurable, adapted process. Then there exists a sequence $\{X_n\}$ of simple predictable processes such that, for any $0 < T < \infty$,*

$$\lim_{n\to\infty} E\int_0^T \big|X_n(t) - X(t)\big|^2\,dt = 0.$$

Proof. We begin by showing that, given $T < \infty$, we can find simple predictable processes $Y_k^{(T)}$ that vanish outside $[0,T]$ and satisfy

$$\lim_{k\to\infty} E\int_0^T \big|Y_k^{(T)}(t) - X(t)\big|^2\,dt = 0. \tag{4.6}$$

Extend X to $\mathbf{R}\times\Omega$ by defining $X(t,\omega) = 0$ for $t < 0$. For each $n \in \mathbf{N}$ and $s \in [0,1]$, define

$$Z^{n,s}(t,\omega) = \sum_{j\in\mathbf{Z}} X(s + 2^{-n}j, \omega)\mathbf{1}_{(s+2^{-n}j,\, s+2^{-n}(j+1)]}(t)\cdot \mathbf{1}_{[0,T]}(t).$$

$Z^{n,s}$ is a simple predictable process. It is jointly measurable as a function of the triple (s,t,ω), so it can be integrated over all three variables. Fubini's theorem allows us to perform these integrations in any order we please.

We claim that

$$\lim_{n\to\infty} E\int_0^T dt\int_0^1 ds\,\bigl|Z^{n,s}(t)-X(t)\bigr|^2=0. \tag{4.7}$$

This limit relies on the property called L^p-continuity (Proposition A.18).

To prove (4.7), start by considering a fixed ω and rewrite the double integral as follows:

$$\begin{aligned}
&\int_0^T dt\int_0^1 ds\,\bigl|Z^{n,s}(t,\omega)-X(t,\omega)\bigr|^2\\
&=\int_0^T dt\sum_{j\in\mathbf{Z}}\int_0^1 ds\,\bigl|X(s+2^{-n}j,\omega)-X(t,\omega)\bigr|^2\mathbf{1}_{(s+2^{-n}j,s+2^{-n}(j+1)]}(t)\\
&=\int_0^T dt\sum_{j\in\mathbf{Z}}\int_0^1 ds\,\bigl|X(s+2^{-n}j,\omega)-X(t,\omega)\bigr|^2\mathbf{1}_{[t-2^{-n}(j+1),t-2^{-n}j)}(s).
\end{aligned}$$

For a fixed t, the s-integral vanishes unless

$$0<t-2^{-n}j \text{ and } t-2^{-n}(j+1)<1,$$

which is equivalent to $2^n(t-1)-1<j<2^nt$. For each fixed t and j, change variables in the s-integral: let $h=t-s-2^{-n}j$. Then $s\in[t-2^{-n}(j+1),t-2^{-n}j)$ iff $h\in(0,2^{-n}]$. These steps turn the integral into

$$\begin{aligned}
&\int_0^T dt\sum_{j\in\mathbf{Z}}\mathbf{1}_{\{2^n(t-1)-1<j<2^nt\}}\int_0^{2^{-n}} dh\,\bigl|X(t-h,\omega)-X(t,\omega)\bigr|^2\\
&\qquad\le(2^n+1)\int_0^{2^{-n}} dh\int_0^T dt\,\bigl|X(t-h,\omega)-X(t,\omega)\bigr|^2.
\end{aligned}$$

The last upper bound follows because there are at most 2^n+1 j-values that satisfy the restriction $2^n(t-1)-1<j<2^nt$. Now take expectations through the inequalities. We get

$$\begin{aligned}
&E\int_0^T dt\int_0^1 ds\,\bigl|Z^{n,s}(t)-X(t)\bigr|^2\,dt\\
&\qquad\le(2^n+1)\int_0^{2^{-n}} dh\left\{E\int_0^T dt\,\bigl|X(t-h,\omega)-X(t,\omega)\bigr|^2\right\}.
\end{aligned}$$

The last line vanishes as $n\to\infty$ for these reasons: First,

$$\lim_{h\to 0}\int_0^T dt\,\bigl|X(t-h,\omega)-X(t,\omega)\bigr|^2=0$$

for each fixed ω by virtue of L^p-continuity (Proposition A.18). Since X is bounded, the expectations converge by dominated convergence:

$$\lim_{h\to 0}E\int_0^T dt\,\bigl|X(t-h,\omega)-X(t,\omega)\bigr|^2=0.$$

Last,

$$\lim_{n\to\infty}(2^n+1)\int_0^{2^{-n}} dh\left\{E\int_0^T dt\,\big|X(t-h,\omega)-X(t,\omega)\big|^2\right\}=0$$

follows from this general fact, which we leave as an exercise: if $f(x)\to 0$ as $x\to 0$, then

$$\frac{1}{\varepsilon}\int_0^\varepsilon f(x)\,dx\to 0 \quad \text{as } \varepsilon\to 0.$$

We have justified (4.7).

We can restate (4.7) by saying that the function

$$\phi_n(s)=E\int_0^T dt\,\big|Z^{n,s}(t)-X(t)\big|^2\,dt$$

satisfies $\phi_n\to 0$ in $L^1[0,1]$. Consequently a subsequence $\phi_{n_k}(s)\to 0$ for Lebesgue-almost every $s\in[0,1]$. Fix any such s. Define $Y_k^{(T)}=Z^{n_k,s}$, and we have (4.6).

To complete the proof, create the simple predictable processes $\{Y_k^{(m)}\}$ for all $T=m\in\mathbf{N}$. For each m, pick k_m such that

$$E\int_0^m dt\,\big|Y_{k_m}^{(m)}(t)-X(t)\big|^2\,dt<\frac{1}{m}.$$

Then $X_m=Y_{k_m}^{(m)}$ satisfies the requirement of the lemma. □

Proposition 4.3. *Suppose $X\in\mathcal{L}_2(B)$. Then there exists a sequence of simple predictable processes $\{X_n\}$ such that $\|X-X_n\|_{\mathcal{L}_2(B)}\to 0$.*

Proof. Let $X^{(k)}=(X\wedge k)\vee(-k)$. Since $|X^{(k)}-X|\le|X|$ and $|X^{(k)}-X|\to 0$ pointwise on $\mathbf{R}_+\times\Omega$,

$$\lim_{k\to\infty}E\int_0^m\big|X^{(k)}(t)-X(t)\big|^2\,dt=0$$

for each $m\in\mathbf{N}$. This is equivalent to $\|X-X^{(k)}\|_{\mathcal{L}_2(B)}\to 0$. Given $\varepsilon>0$, pick k such that $\|X-X^{(k)}\|_{\mathcal{L}_2(B)}\le\varepsilon/2$. Since $X^{(k)}$ is a bounded process, the previous lemma gives a simple predictable process Y such that $\|X^{(k)}-Y\|_{\mathcal{L}_2(B)}\le\varepsilon/2$. By the triangle inequality $\|X-Y\|_{\mathcal{L}_2(B)}\le\varepsilon$. Repeat this argument for each $\varepsilon=1/n$, and let X_n be the Y thus selected. This gives the approximating sequence $\{X_n\}$. □

We are ready to proceed to the construction of the stochastic integral. There are three main steps.

(i) First an explicit formula is given for the integral $X\cdot B$ of a simple predictable process X. This integral will be a continuous L^2-martingale.

(ii) A general process X in $\mathcal{L}_2(B)$ is approximated by simple processes X_n. One shows that the integrals $X_n \cdot B$ of the approximating simple processes converge to a uniquely defined continuous L^2-martingale which is then defined to be the stochastic integral $X \cdot B$.

(iii) A localization step is used to get from integrands in $\mathcal{L}_2(B)$ to integrands in $\mathcal{L}(B)$. The integral $X \cdot B$ is a continuous local L^2-martingale

The lemmas needed along the way for this development make valuable exercises. So we give only hints for the proofs, and urge the first-time reader to give them a try. These same properties are proved again with full detail in the next chapter when we develop the more general integral. The proofs for the Brownian case are very similar to those for the general case.

We begin with the integral of simple processes. For a simple predictable process of the type (4.5), the *stochastic integral* is the process $X \cdot B$ defined by

$$(X \cdot B)_t(\omega) = \sum_{i=1}^{n-1} \xi_i(\omega)\big(B_{t_{i+1}\wedge t}(\omega) - B_{t_i \wedge t}(\omega)\big). \tag{4.8}$$

Note that our convention is such that the value of X at $t = 0$ does not influence the integral. We also write $I(X) = X \cdot B$ when we need a symbol for the mapping $I : X \mapsto X \cdot B$.

Let $\mathcal{S}_2$ denote the space of simple predictable processes. It is a subspace of $\mathcal{L}_2(B)$. An element X of $\mathcal{S}_2$ can be represented in the form (4.5) in many different ways. We need to check that the integral $X \cdot B$ depends only on the process X and not on the particular representation. Also, we need to know that $\mathcal{S}_2$ is a linear space, and that the integral $I(X)$ is a linear map on $\mathcal{S}_2$.

Lemma 4.4. (a) *Suppose the process X in (4.5) also satisfies*

$$X_t(\omega) = \eta_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{j=1}^{m-1} \eta_j(\omega)\mathbf{1}_{(s_i, s_{i+1}]}(t)$$

for all (t, ω), where $0 = s_0 = s_1 < s_2 < \cdots < s_m < \infty$ and η_j is $\mathcal{F}_{s_j}$-measurable for $0 \le j \le m-1$. Then for each (t, ω),

$$\sum_{i=1}^{n-1} \xi_i(\omega)\big(B_{t_{i+1}\wedge t}(\omega) - B_{t_i \wedge t}(\omega)\big) = \sum_{j=1}^{m-1} \eta_i(\omega)\big(B_{s_{i+1}\wedge t}(\omega) - B_{s_i \wedge t}(\omega)\big).$$

In other words, $X \cdot B$ is independent of the representation.

(b) $\mathcal{S}_2$ *is a linear space, in other words for* $X, Y \in \mathcal{S}_2$ *and reals* α *and* β, $\alpha X + \beta Y \in \mathcal{S}_2$. *The integral satisfies*

$$(\alpha X + \beta Y) \cdot B = \alpha(X \cdot B) + \beta(Y \cdot B).$$

Hints for proof. Let $\{u_k\} = \{s_j\} \cup \{t_i\}$ be the common refinement of the partitions $\{s_j\}$ and $\{t_i\}$. Rewrite both representations of X in terms of $\{u_k\}$. The same idea can be used for part (b) to write two arbitrary simple processes in terms of a common partition, which makes adding them easy. □

Next we need some continuity properties for the integral. Recall the distance measure $\|\cdot\|_{\mathcal{M}_2}$ defined for continuous L^2-martingales by (3.20).

Lemma 4.5. *Let* $X \in \mathcal{S}_2$. *Then* $X \cdot B$ *is a continuous square-integrable martingale with respect to the original filtration* $\{\mathcal{F}_t\}$. *We have these isometries:*

$$E\big[(X \cdot B)_t^2\big] = E\int_0^t X_s^2\, ds \quad \text{for all } t \geq 0, \tag{4.9}$$

and

$$\|X \cdot B\|_{\mathcal{M}_2} = \|X\|_{\mathcal{L}_2(B)}. \tag{4.10}$$

Hints for proof. To show that $X \cdot B$ is a martingale, start by proving this statement: if $u < v$ and ξ is a bounded $\mathcal{F}_u$-measurable random variable, then $Z_t = \xi(B_{t\wedge v} - B_{t\wedge u})$ is a martingale.

To prove (4.9), first square:

$$\begin{aligned}(X \cdot B)_t^2 = \sum_{i=1}^{n-1} \xi_i^2 \big(B_{t\wedge t_{i+1}} - B_{t\wedge t_i}\big)^2 \\ + 2\sum_{i<j} \xi_i \xi_j (B_{t\wedge t_{i+1}} - B_{t\wedge t_i})(B_{t\wedge t_{j+1}} - B_{t\wedge t_j}).\end{aligned}$$

Then compute the expectations of all terms. □

From the isometry property we can deduce that simple process approximation gives approximation of stochastic integrals.

Lemma 4.6. *Let* $X \in \mathcal{L}_2(B)$. *Then there is a unique continuous* L^2-*martingale* Y *such that, for any sequence of simple predictable processes* $\{X_n\}$ *such that*

$$\|X - X_n\|_{\mathcal{L}_2(B)} \to 0,$$

we have

$$\|Y - X_n \cdot B\|_{\mathcal{M}_2} \to 0.$$

Hints for proof. It all follows from these facts: an approximating sequence of simple predictable processes exists for each process in $\mathcal{L}_2(B)$, a convergent sequence in a metric space is a Cauchy sequence, a Cauchy sequence in a complete metric space converges, the space $\mathcal{M}_2^c$ of continuous L^2-martingales is complete, the isometry (4.10), and the triangle inequality. □

Note that uniqueness of the process Y defined in the lemma means uniqueness up to indistinguishability: any process $\widetilde{Y}$ indistinguishable from Y also satisfies $\|\widetilde{Y} - X_n \cdot B\|_{\mathcal{M}_2} \to 0$.

Now we can state the definition of the integral of $\mathcal{L}_2(B)$-integrands with respect to Brownian motion.

Definition 4.7. Let B be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ with respect to a filtration $\{\mathcal{F}_t\}$. For any measurable adapted process $X \in \mathcal{L}_2(B)$, the *stochastic integral* $I(X) = X \cdot B$ is the square-integrable continuous martingale that satisfies

$$\lim_{n\to\infty} \|X \cdot B - X_n \cdot B\|_{\mathcal{M}_2} = 0$$

for any sequence $X_n \in \mathcal{S}_2$ of simple predictable processes such that

$$\|X - X_n\|_{\mathcal{L}_2(B)} \to 0.$$

The process $I(X)$ is unique up to indistinguishability. Alternative notation for the stochastic integral is the familiar

$$\int_0^t X_s \, dB_s = (X \cdot B)_t.$$

The reader familiar with more abstract principles of analysis should note that the extension of the stochastic integral $X \cdot B$ from $X \in \mathcal{S}_2$ to $X \in \mathcal{L}_2(B)$ is an instance of a general, classic argument. A uniformly continuous map from a metric space into a complete metric space can always be extended to the closure of its domain (Lemma A.3). If the spaces are linear, the linear operations are continuous, and the map is linear, then the extension is a linear map too (Lemma A.4). In this case the map is $X \mapsto X \cdot B$, first defined for $X \in \mathcal{S}_2$. Uniform continuity follows from linearity and (4.10). Proposition 4.3 implies that the closure of $\mathcal{S}_2$ in $\mathcal{L}_2(B)$ is all of $\mathcal{L}_2(B)$.

Some books first define the integral $(X \cdot B)_t$ at a fixed time t as a map from $L^2([0,t] \times \Omega, m \otimes P)$ into $L^2(P)$, utilizing the completeness of L^2-spaces. Then one needs a separate argument to show that the integrals defined for different times t can be combined into a continuous martingale $t \mapsto (X \cdot B)_t$. We defined the integral directly as a map into the space of martingales $\mathcal{M}_2^c$ to avoid the extra argument. Of course, we did not really save work. We just did part of the work earlier when we proved that $\mathcal{M}_2^c$ is a complete space (Theorem 3.41).

Example 4.8. In the definition (4.5) of the simple predictable process we required the ξ_i bounded because this will be convenient later. For this section it would have been more convenient to allow square-integrable ξ_i. So let us derive the integral for that case. Let

$$X(t) = \sum_{i=1}^{m-1} \eta_i \mathbf{1}_{(s_i, s_{i+1}]}(t)$$

where $0 \le s_1 < \cdots < s_m$ and each $\eta_i \in L^2(P)$ is $\mathcal{F}_{s_i}$-measurable. Check that a sequence of approximating simple processes is given by

$$X_k(t) = \sum_{i=1}^{m-1} \eta_i^{(k)} \mathbf{1}_{(s_i, s_{i+1}]}(t)$$

with truncated variables $\eta_i^{(k)} = (\eta_i \wedge k) \vee (-k)$. And then that

$$\int_0^t X(s)\, dB_s = \sum_{i=1}^{m-1} \eta_i (B_{t\wedge s_{i+1}} - B_{t\wedge s_i}).$$

There is something to check here because it is not immediately obvious that the terms on the right above are square-integrable. See Exercise 4.4.

Example 4.9. One can check that Brownian motion itself is an element of $\mathcal{L}_2(B)$. Let $t_i^n = i2^{-n}$ and

$$X_n(t) = \sum_{i=0}^{2^n n-1} B_{t_i^n} \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t).$$

$X_n \notin \mathcal{S}_2$ but it can be used to approximate B. By Example 4.8

$$\int_0^t X_n(s)\, dB_s = \sum_{i=1}^{2^n n-1} B_{t_i^n} (B_{t\wedge t_{i+1}^n} - B_{t\wedge t_i^n}).$$

For any $T < n$,

$$\begin{aligned} E \int_0^T |X_n(s) - B_s|^2\, dt &\le \sum_{i=0}^{2^n n-1} \int_{t_i^n}^{t_{i+1}^n} E\big[(B_{t_i^n} - B_s)^2\big]\, ds \\ &= \sum_{i=0}^{2^n n-1} \tfrac{1}{2}(t_{i+1}^n - t_i^n)^2 = \tfrac{1}{2} n 2^{-n}. \end{aligned}$$

Thus X_n converges to B in $\mathcal{L}_2(B)$ as $n \to \infty$. By the isometry (4.12) in the next Proposition, this integral converges to $\int_0^t B_s\,dB_s$ in L^2 as $n \to \infty$, so by Lemma 4.1,

$$\int_0^t B_s\,dB_s = \tfrac{1}{2}B_t^2 - \tfrac{1}{2}t.$$

Before developing the integral further, we record some properties.

Proposition 4.10. *Let $X, Y \in \mathcal{L}_2(B)$.*

(a) *Linearity carries over:*

$$(\alpha X + \beta Y) \cdot B = \alpha(X \cdot B) + \beta(Y \cdot B).$$

(b) *We have again the isometries*

$$E\big[(X \cdot B)_t^2\big] = E\int_0^t X_s^2\,ds \quad \text{for all } t \geq 0, \tag{4.11}$$

and

$$\|X \cdot B\|_{\mathcal{M}_2} = \|X\|_{\mathcal{L}_2(B)}. \tag{4.12}$$

In particular, if $X, Y \in \mathcal{L}_2(B)$ are $m \otimes P$-equivalent in the sense (4.3), then $X \cdot B$ and $Y \cdot B$ are indistinguishable.

(c) *Suppose τ is a stopping time such that $X(t,\omega) = Y(t,\omega)$ for $t \leq \tau(\omega)$. Then for almost every ω, $(X \cdot B)_t(\omega) = (Y \cdot B)_t(\omega)$ for $t \leq \tau(\omega)$.*

Hints for proof. Parts (a)–(b): These properties are inherited from the integrals of the approximating simple processes X_n. One needs to justify taking limits in Lemma 4.4(b) and Lemma 4.5.

The proof of part (c) is different from the one that is used in the next chapter. So we give here more details than in previous proofs.

By considering $Z = X - Y$, it suffices to prove that if $Z \in \mathcal{L}_2(B)$ satisfies $Z(t,\omega) = 0$ for $t \leq \tau(\omega)$, then $(Z \cdot B)_t(\omega) = 0$ for $t \leq \tau(\omega)$.

Assume first that Z is bounded, so $|Z(t,\omega)| \leq C$. Pick a sequence $\{Z_n\}$ of simple predictable processes that converge to Z in $\mathcal{L}_2(B)$. Let Z_n be of the generic type (recall (4.5))

$$Z_n(t,\omega) = \sum_{i=1}^{m(n)-1} \xi_i^n(\omega)\mathbf{1}_{(t_i^n, t_{i+1}^n]}(t).$$

(To approximate a process in $\mathcal{L}_2(B)$ the $t = 0$ term in (4.5) is not needed because values at $t = 0$ do not affect dt-integrals.) We may assume $|\xi_i^n| \leq C$ always, for if this is not the case, replacing ξ_i^n by $(\xi_i^n \wedge C) \vee (-C)$ will only improve the approximation. Define another sequence of simple predictable processes by

$$\widetilde{Z}_n(t) = \sum_{i=1}^{m(n)-1} \xi_i^n \mathbf{1}\{\tau \le t_i^n\} \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t).$$

We claim that

$$\widetilde{Z}_n \to Z \text{ in } \mathcal{L}_2(B). \tag{4.13}$$

To prove (4.13), note first that $Z_n \mathbf{1}\{\tau < t\} \to Z\mathbf{1}\{\tau < t\} = Z$ in $\mathcal{L}_2(B)$. So it suffices to show that

$$Z_n \mathbf{1}\{\tau < t\} - \widetilde{Z}_n \to 0 \text{ in } \mathcal{L}_2(B). \tag{4.14}$$

Estimate

$$\begin{aligned} \big|Z_n(t)\mathbf{1}\{\tau < t\} - \widetilde{Z}_n(t)\big| &\le C \sum_i \big|\mathbf{1}\{\tau < t\} - \mathbf{1}\{\tau \le t_i^n\}\big| \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t) \\ &\le C \sum_i \mathbf{1}\{t_i^n < \tau < t_{i+1}^n\} \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t). \end{aligned}$$

Integrate over $[0, T] \times \Omega$, to get

$$\begin{aligned} E \int_0^T &\big|Z_n(t)\mathbf{1}\{\tau < t\} - \widetilde{Z}_n(t)\big|^2 \, dt \\ &\le C^2 \sum_i P\{t_i^n < \tau < t_{i+1}^n\} \int_0^T \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t) \, dt \\ &\le C^2 \max\{T \wedge t_{i+1}^n - T \wedge t_i^n : 1 \le i \le m(n) - 1\}. \end{aligned}$$

We can artificially add partition points t_i^n to each Z_n so that this last quantity converges to 0 as $n \to \infty$, for each fixed T. This verifies (4.14), and thereby (4.13).

The integral of $\widetilde{Z}_n$ is given explicitly by

$$(\widetilde{Z}_n \cdot B)_t = \sum_{i=1}^{m(n)-1} \xi_i^n \mathbf{1}\{\tau \le t_i^n\}(B_{t \wedge t_{i+1}^n} - B_{t \wedge t_i^n}).$$

By inspecting each term, we see that $(\widetilde{Z}_n \cdot B)_t = 0$ if $t \le \tau$. By the definition of the integral and (4.13), $\widetilde{Z}_n \cdot B \to Z \cdot B$ in $\mathcal{M}_2^c$. Then by Lemma 3.42 there exists a subsequence $\widetilde{Z}_{n_k} \cdot B$ and an event Ω_0 of full probability such that, for each $\omega \in \Omega_0$ and $T < \infty$,

$$(\widetilde{Z}_{n_k} \cdot B)_t(\omega) \to (Z \cdot B)_t(\omega) \quad \text{uniformly for } 0 \le t \le T.$$

For any $\omega \in \Omega_0$, in the limit $(Z \cdot B)_t(\omega) = 0$ for $t \le \tau(\omega)$. Part (c) has been proved for a bounded process.

To complete the proof, given $Z \in \mathcal{L}_2(B)$, let $Z^{(k)}(t, \omega) = (Z(t, \omega) \wedge k) \vee (-k)$, a bounded process in $\mathcal{L}_2(B)$ with the same property $Z^{(k)}(t, \omega) = 0$ if

$t \le \tau(\omega)$. Apply the previous step to $Z^{(k)}$ and justify what happens in the limit. □

Next we extend the integral to integrands in $\mathcal{L}(B)$. Given a process $X \in \mathcal{L}(B)$, define the stopping times

$$\tau_n(\omega) = \inf\Big\{ t \ge 0 : \int_0^t X(s,\omega)^2\,ds \ge n \Big\}. \tag{4.15}$$

These are stopping times by Corollary 2.10 because the function

$$t \mapsto \int_0^t X(s,\omega)^2\,ds$$

is continuous for each ω in the event in the definition (4.4). By this same continuity, if $\tau_n(\omega) < \infty$,

$$\int_0^\infty X(s,\omega)^2 \mathbf{1}\{s \le \tau_n(\omega)\}\,ds = \int_0^{\tau_n(\omega)} X(s,\omega)^2\,ds = n.$$

Let

$$X_n(t,\omega) = X(t,\omega)\mathbf{1}\{t \le \tau_n(\omega)\}.$$

Adaptedness of X_n follows from $\{t \le \tau_n\} = \{\tau_n < t\}^c \in \mathcal{F}_t$. The function $(t,\omega) \mapsto \mathbf{1}\{t \le \tau_n(\omega)\}$ is $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable by Exercise 4.3, hence X_n is a measurable process. Together these properties say that $X_n \in \mathcal{L}_2(B)$, and the stochastic integrals $X_n \cdot B$ are well-defined.

The goal is to show that there is a uniquely defined limit of the processes $X_n \cdot B$ as $n \to \infty$, and this will then serve as the definition of $X \cdot B$.

Lemma 4.11. *For almost every ω, $(X_m \cdot B)_t(\omega) = (X_n \cdot B)_t(\omega)$ for all $t \le \tau_m(\omega) \wedge \tau_n(\omega)$.*

Proof. Immediate from Proposition 4.10(c). □

The lemma says that, for a given (t,ω), once n is large enough so that $\tau_n(\omega) \ge t$, the value $(X_n \cdot B)_t(\omega)$ does not change with n. The definition (4.4) guarantees that $\tau_n(\omega) \nearrow \infty$ for almost every ω. These ingredients almost justify the next extension of the stochastic integral to $\mathcal{L}(B)$.

Definition 4.12. Let B be a Brownian motion on a probability space $(\Omega, \mathcal{F}, P)$ with respect to a filtration $\{\mathcal{F}_t\}$, and $X \in \mathcal{L}(B)$. Let Ω_0 be the event of full probability on which $\tau_n \nearrow \infty$ and the conclusion of Lemma 4.11 holds for all pairs m, n. The *stochastic integral* $X \cdot B$ is defined for $\omega \in \Omega_0$ by

$$(X \cdot B)_t(\omega) = (X_n \cdot B)_t(\omega) \quad \text{for any } n \text{ such that } \tau_n(\omega) \ge t. \tag{4.16}$$

For $\omega \notin \Omega_0$ define $(X \cdot B)_t(\omega) \equiv 0$. The process $X \cdot B$ is a continuous local L^2-martingale.

To justify the claim that $X \cdot B$ is a local L^2-martingale, just note that $\{\tau_n\}$ serves as a localizing sequence:

$$(X \cdot B)_t^{\tau_n} = (X \cdot B)_{t \wedge \tau_n} = (X_n \cdot B)_{t \wedge \tau_n} = (X_n \cdot B)_t^{\tau_n},$$

so $(X \cdot B)^{\tau_n} = (X_n \cdot B)^{\tau_n}$, which is an L^2-martingale by Corollary 3.8. The above equality also implies that $(X \cdot B)_t(\omega)$ is continuous for $t \in [0, \tau_n(\omega)]$, which contains any given interval $[0, T]$ if n is taken large enough.

It seems somewhat arbitrary to base the definition of the stochastic integral on the particular stopping times $\{\tau_n\}$. The property that enabled us to define $X \cdot B$ by (4.16) was that $X(t)\mathbf{1}\{t \le \tau_n\}$ is a process in the space $\mathcal{L}_2(B)$ for all n. Let us make this into a new definition.

Definition 4.13. Let X be an adapted, measurable process. A nondecreasing sequence of stopping times $\{\sigma_n\}$ is a *localizing sequence* for X if $X(t)\mathbf{1}\{t \le \sigma_n\}$ is in $\mathcal{L}_2(B)$ for all n, and $\sigma_n \nearrow \infty$ with probability one.

One can check that $X \in \mathcal{L}(B)$ if and only if X has a localizing sequence $\{\sigma_n\}$ (Exercise 4.6). Lemma 4.11 and Definition 4.12 work equally well with $\{\tau_n\}$ replaced by an arbitrary localizing sequence $\{\sigma_n\}$. Fix such a sequence $\{\sigma_n\}$ and define $\widetilde{X}_n(t) = \mathbf{1}\{t \le \sigma_n\}X(t)$. Let Ω_1 be the event of full probability on which $\sigma_n \nearrow \infty$ and for all pairs m, n, $(\widetilde{X}_m \cdot B)_t = (\widetilde{X}_n \cdot B)_t$ for $t \le \sigma_m \wedge \sigma_n$. (In other words, the conclusion of Lemma 4.11 holds for $\{\sigma_n\}$.) Let Y be the process defined by

$$Y_t(\omega) = (\widetilde{X}_n \cdot B)_t(\omega) \quad \text{for any } n \text{ such that } \sigma_n(\omega) \ge t, \tag{4.17}$$

for $\omega \in \Omega_1$, and identically zero outside Ω_1.

Lemma 4.14. *$Y = X \cdot B$ in the sense of indistinguishability.*

Proof. Let Ω_2 be the intersection of the full-probability events Ω_0 and Ω_1 defined previously above (4.16) and (4.17). By applying Proposition 4.10(c) to the stopping time $\sigma_n \wedge \tau_n$ and the processes X_n and $\widetilde{X}_n$, we conclude that for almost every $\omega \in \Omega_2$, if $t \le \sigma_n(\omega) \wedge \tau_n(\omega)$,

$$Y_t(\omega) = (\widetilde{X}_n \cdot B)_t(\omega) = (X_n \cdot B)_t(\omega) = (X \cdot B)_t(\omega).$$

Since $\sigma_n(\omega) \wedge \tau_n(\omega) \nearrow \infty$, the above equality holds almost surely for all $0 \le t < \infty$. □

This lemma tells us that for $X \in \mathcal{L}(B)$ the stochastic integral $X \cdot B$ can be defined in terms of any localizing sequence of stopping times.

Exercises

Exercise 4.1. Show by example that it is possible to have $E\int_0^1 |X_t|\,dt < \infty$ but still $E|X_t| = \infty$ for some fixed time $t \in [0,1]$.

Exercise 4.2. Let X be a continuous, adapted process such that

$$E\Big(\sup_{t\in[0,T]} |X_t|^2\Big) < \infty \qquad \text{for all } T < \infty.$$

Let $\pi^n = \{0 = s_0^n < s_1^n < s_2^n < \cdots\}$ be any sequence of partitions of $\mathbf{R}_+$ such that, for each fixed n, $s_i^n \nearrow \infty$ as $i \nearrow \infty$ while mesh $\pi^n = \sup_i (s_{i+1}^n - s_i^n) \to 0$ as $n \to \infty$. Let g_n be the truncation $g_n(x) = (x \wedge n) \vee (-n)$. Show that the processes

$$X_n(t) = \sum_{i\geq 0} g_n(X(s_i^n))\mathbf{1}_{(s_i^n,\, s_{i+1}^n]}(t)$$

are simple predictable processes that satisfy $\|X_n - X\|_{\mathcal{L}_2(B)} \to 0$.

Exercise 4.3. Show that for any $[0,\infty]$-valued measurable function Y on $(\Omega, \mathcal{F})$, the set $\{(s,\omega) \in \mathbf{R}_+ \times \Omega : Y(\omega) > s\}$ is $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable.

Hint. Start with a simple Y. Show that if $Y_n \nearrow Y$ pointwise, then $\{(s,\omega) : Y(\omega) > s\} = \bigcup_n \{(s,\omega) : Y_n(\omega) > s\}$.

Exercise 4.4. Suppose $\eta \in L^2(P)$ is $\mathcal{F}_s$ measurable and $t > s$. Show that

$$E\big[\eta^2 (B_t - B_s)^2\big] = E\big[\eta^2\big] \cdot E\big[(B_t - B_s)^2\big]$$

by truncating and using monotone convergence. In particular, this implies that $\eta(B_t - B_s) \in L^2(P)$.

Complete the details in Example 4.8. You need to show first that $X_k \to X$ in $\mathcal{L}_2(B)$, and then that

$$\int_0^t X_k(s)\,dB_s \to \sum_{i=1}^{m-1} \eta_i (B_{t\wedge s_{i+1}} - B_{t\wedge s_i}) \quad \text{in } \mathcal{M}_2^c.$$

Exercise 4.5. Show that B_t^2 is a process in $\mathcal{L}_2(B)$ and evaluate

$$\int_0^t B_s^2\,dB_s.$$

Hint. Follow the example of $\int_0^t B_s\,dB_s$. Answer: $\frac{1}{3}B_t^3 - \int_0^t B_s\,ds$.

Exercise 4.6. Let X be an adapted, measurable process. Show that $X \in \mathcal{L}(B)$ if and only if X has a localizing sequence $\{\sigma_n\}$ in the sense of Definition 4.13.

Exercise 4.7. (Integral of a step function in $\mathcal{L}(B)$.) Fix $0 = t_0 < t_1 < \cdots < t_M < \infty$, and random variables $\eta_0, \dots, \eta_{M-1}$. Assume that η_i is almost surely finite and $\mathcal{F}_{t_i}$-measurable, but make no integrability assumption. Define

$$g(s,\omega) = \sum_{i=0}^{M-1} \eta_i(\omega) \mathbf{1}_{(t_i, t_{i+1}]}(s).$$

The task is to show that $g \in \mathcal{L}(B)$ (virtually immediate) and that

$$\int_0^t g(s)\, dB_s = \sum_{i=0}^{M-1} \eta_i (B_{t_{i+1} \wedge t} - B_{t_i \wedge t})$$

as one would expect.

Hints. Show that $\sigma_n(\omega) = \inf\{t : |g(t,\omega)| \geq n\}$ defines a localizing sequence of stopping times. (Recall the convention $\inf \emptyset = \infty$.) Show that $g_n(t,\omega) = g(t,\omega)\mathbf{1}\{t \leq \sigma_n(\omega)\}$ is also a simple predictable process with the same partition. Then we know what the approximating integrals

$$Y_n(t,\omega) = \int_0^t g_n(s,\omega)\, dB_s(\omega)$$

look like.

Exercise 4.8. Let f be a (nonrandom) Borel function on $[0,T]$ such that $\int_0^T |f|^2\, dt < \infty$. Find the distribution of the random variable $\int_0^T f(t)\, dB_t(\omega)$.

Chapter 5

Stochastic Integration of Predictable Processes

The main goal of this chapter is the definition of the stochastic integral $\int_{(0,t]} X(s)\,dY(s)$ where the integrator Y is a cadlag semimartingale and X is a locally bounded predictable process. The most important special case is the one where the integrand is of the form $X(t-)$ for some cadlag process X. In this case the stochastic integral $\int_{(0,t]} X(s-)\,dY(s)$ can be realized as the limit of Riemann sums

$$S(t) = \sum_{i=0}^{\infty} X(s_i)\big(Y(s_{i+1} \wedge t) - Y(s_i \wedge t)\big)$$

when the mesh of the partition $\{s_i\}$ tends to zero. The convergence is then uniform on compact time intervals, and happens in probability. Random partitions of stopping times can also be used.

These results will be reached in Section 5.3. Before the semimartingale integral we explain predictable processes and construct the integral with respect to L^2-martingales and local L^2-martingales. Right-continuity of the filtration $\{\mathcal{F}_t\}$ is not needed until we define the integral with respect to a semimartingale. And even there it is needed only for guaranteeing that the semimartingale has a decomposition whose local martingale part is a local L^2-martingale. Right-continuity of $\{\mathcal{F}_t\}$ is not needed for the arguments that establish the integral.

5.1. Square-integrable martingale integrator

Throughout this section, we consider a fixed probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$. M is a square-integrable cadlag martingale relative to the filtration $\{\mathcal{F}_t\}$. We assume that the probability space and the filtration are complete. In other words, if $N \in \mathcal{F}$ has $P(N) = 0$, then every subset $F \subseteq N$ is a member of $\mathcal{F}$ and each $\mathcal{F}_t$. Right-continuity of the filtration $\{\mathcal{F}_t\}$ is not assumed unless specifically stated.

5.1.1. Predictable processes. *Predictable rectangles* are subsets of $\mathbf{R}_+ \times \Omega$ of the type $(s,t] \times F$ where $0 \le s < t < \infty$ and $F \in \mathcal{F}_s$, or of the type $\{0\} \times F_0$ where $F_0 \in \mathcal{F}_0$. $\mathcal{R}$ stands for the collection of all predictable rectangles. We regard the empty set also as a predictable rectangle, since it can be represented as $(s,t] \times \emptyset$. The σ-field generated by $\mathcal{R}$ in the space $\mathbf{R}_+ \times \Omega$ is denoted by $\mathcal{P}$ and called the *predictable σ-field*. $\mathcal{P}$ is a sub-σ-field of $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$ because $\mathcal{R} \subseteq \mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$. Any $\mathcal{P}$-measurable function X from $\mathbf{R}_+ \times \Omega$ into $\mathbf{R}$ is called a *predictable process*.

A predictable process is not only adapted to the original filtration $\{\mathcal{F}_t\}$ but also to the potentially smaller filtration $\{\mathcal{F}_{t-}\}$ defined in (2.6) [Exercise 5.1]. This gives some mathematical sense to the term "predictable", because it means that X_t is knowable from the information "immediately prior to t" represented by $\mathcal{F}_{t-}$.

Predictable processes will be the integrands for the stochastic integral. Before proceeding, let us develop additional characterizations of the σ-field $\mathcal{P}$.

Lemma 5.1. *The following σ-fields on $\mathbf{R}_+ \times \Omega$ are all equal to $\mathcal{P}$.*

(a) *The σ-field generated by all continuous adapted processes.*

(b) *The σ-field generated by all left-continuous adapted processes.*

(c) *The σ-field generated by all adapted caglad processes (that is, left-continuous processes with right limits).*

Proof. Continuous processes and caglad processes are left-continuous. Thus to show that σ-fields (a)–(c) are contained in $\mathcal{P}$, it suffices to show that all left-continuous processes are $\mathcal{P}$-measurable.

Let X be a left-continuous, adapted process. Let

$$X_n(t,\omega) = X_0(\omega)\mathbf{1}_{\{0\}}(0) + \sum_{i=0}^{\infty} X_{i2^{-n}}(\omega)\mathbf{1}_{(i2^{-n},(i+1)2^{-n}]}(t).$$

Then for $B \in \mathcal{B}_{\mathbf{R}}$,

$$\begin{aligned}\{(t,\omega) : X_n(t,\omega) \in B\} &= \{0\} \times \{\omega : X_0(\omega) \in B\} \\ &\cup \bigcup_{i=0}^{\infty} \Big\{ (i2^{-n}, (i+1)2^{-n}] \times \{\omega : X_{i2^{-n}}(\omega) \in B\} \Big\}\end{aligned}$$

which is an event in $\mathcal{P}$, being a countable union of predictable rectangles. Thus X_n is $\mathcal{P}$-measurable. By left-continuity of X, $X_n(t,\omega) \to X(t,\omega)$ as $n \to \infty$ for each fixed (t,ω). Since pointwise limits preserve measurability, X is also $\mathcal{P}$-measurable.

We have shown that $\mathcal{P}$ contains σ-fields (a)–(c).

The indicator of a predictable rectangle is itself an adapted caglad process, and by definition this subclass of caglad processes generates $\mathcal{P}$. Thus σ-field (c) contains $\mathcal{P}$. By the same reasoning, also σ-field (b) contains $\mathcal{P}$.

It remains to show that σ-field (a) contains $\mathcal{P}$. We show that all predictable rectangles lie in σ-field (a) by showing that their indicator functions are pointwise limits of continuous adapted processes.

If $X = \mathbf{1}_{\{0\}\times F_0}$ for $F_0 \in \mathcal{F}_0$, let

$$g_n(t) = \begin{cases} 1 - nt, & 0 \le t < 1/n \\ 0, & t \ge 1/n, \end{cases}$$

and then define $X_n(t,\omega) = \mathbf{1}_{F_0}(\omega) g_n(t)$. X_n is clearly continuous. For a fixed t, writing

$$X_n(t) = \begin{cases} g_n(t)\mathbf{1}_{F_0}, & 0 \le t < 1/n \\ 0, & t \ge 1/n, \end{cases}$$

and noting that $F_0 \in \mathcal{F}_t$ for all $t \ge 0$, shows that X_n is adapted. Since $X_n(t,\omega) \to X(t,\omega)$ as $n \to \infty$ for each fixed (t,ω), $\{0\} \times F_0$ lies in σ-field (a).

If $X = \mathbf{1}_{(u,v]\times F}$ for $F \in \mathcal{F}_u$, let

$$h_n(t) = \begin{cases} n(t-u), & u \le t < u + 1/n \\ 1, & u + 1/n \le t < v \\ 1 - n(t-v), & v \le t \le v + 1/n \\ 0, & t < u \text{ or } t > v + 1/n. \end{cases}$$

Consider only n large enough so that $1/n < v - u$. Define $X_n(t,\omega) = \mathbf{1}_F(\omega) h_n(t)$, and adapt the previous argument. We leave the missing details as Exercise 5.3. □

The previous lemma tells us that all continuous adapted processes, all left-continuous adapted processes, and any process that is a pointwise limit [at each (t, ω)] of a sequence of such processes, is predictable. It is important to note that left and right continuity are not treated equally in this theory. The difference arises from the adaptedness requirement. Not all right continuous processes are predictable. However, an arbitrary *deterministic* process, one that does not depend on ω, is predictable. (See Exercises 5.2 and 5.4).

Given a square-integrable cadlag martingale M, we define its *Doléans measure* μ_M on the predictable σ-field $\mathcal{P}$ by

$$\mu_M(A) = E \int_{[0,\infty)} \mathbf{1}_A(t,\omega)\, d[M]_t(\omega), \quad A \in \mathcal{P}. \tag{5.1}$$

The meaning of formula (5.1) is that first, for each fixed ω, the function $t \mapsto \mathbf{1}_A(t, \omega)$ is integrated by the Lebesgue-Stieltjes measure $\Lambda_{[M](\omega)}$ of the nondecreasing right-continuous function $t \mapsto [M]_t(\omega)$. The resulting integral is a measurable function of ω, which is then averaged over the probability space $(\Omega, \mathcal{F}, P)$ (Exercise 3.13). Recall that our convention for the measure $\Lambda_{[M](\omega)}\{0\}$ of the origin is

$$\Lambda_{[M](\omega)}\{0\} = [M]_0(\omega) - [M]_{0-}(\omega) = 0 - 0 = 0.$$

Consequently integrals over $(0, \infty)$ and $[0, \infty)$ coincide in (5.1).

Formula (5.1) would make sense for any $A \in \mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$. But we shall see that when we want to extend μ_M beyond $\mathcal{P}$ in a useful manner, formula (5.1) does not always provide the right extension. Since

$$\mu_M([0,T] \times \Omega) = E([M]_T) = E(M_T^2 - M_0^2) < \infty \tag{5.2}$$

for all $T < \infty$, the measure μ_M is σ-finite.

Example 5.2 (Brownian motion)**.** If $M = B$, standard Brownian motion, we saw in Proposition 2.42 that $[B]_t = t$. Then

$$\mu_B(A) = E \int_{[0,\infty)} \mathbf{1}_A(t,\omega)\, dt = m \otimes P(A)$$

where m denotes Lebesgue measure on $\mathbf{R}_+$. So the Doléans measure of Brownian motion is $m \otimes P$, the product of Lebesgue measure on $\mathbf{R}_+$ and the probability measure P on Ω.

Example 5.3 (Compensated Poisson process)**.** Let N be a homogeneous rate α Poisson process on $\mathbf{R}_+$ with respect to the filtration $\{\mathcal{F}_t\}$. Let $M_t = N_t - \alpha t$. We claim that the Doléans measure of μ_M is $\alpha m \otimes P$, where as above m is Lebesgue measure on $\mathbf{R}_+$. We have seen that $[M] = N$ (Example

3.26). For a predictable rectangle $A = (s,t] \times F$ with $F \in \mathcal{F}_s$,

$$\begin{aligned}\mu_M(A) &= E\int_{[0,\infty)} \mathbf{1}_A(u,\omega)d[M]_u(\omega) = E\int_{[0,\infty)} \mathbf{1}_F(\omega)\mathbf{1}_{(s,t]}(u)dN_u(\omega)\\ &= E\big[\mathbf{1}_F\cdot(N_t-N_s)\big] = E\big[\mathbf{1}_F\big]E\big[(N_t-N_s)\big]\\ &= P(F)\alpha(t-s) = \alpha m\otimes P(A).\end{aligned}$$

A crucial step above used the independence of $N_t - N_s$ and $\mathcal{F}_s$ which is part of the definition of a Poisson process. Both measures μ_M and $\alpha m \otimes P$ give zero measure to sets of the type $\{0\}\times F_0$. We have shown that μ_M and $\alpha m\otimes P$ agree on the class $\mathcal{R}$ of predictable rectangles. By Lemma B.5 they then agree on $\mathcal{P}$. For the application of Lemma B.5, note that the space $\mathbf{R}_+\times\Omega$ can be written as a countable disjoint union of predictable rectangles: $\mathbf{R}_+\times\Omega = \big(\{0\}\times\Omega\big)\cup\bigcup_{n\ge 0}(n,n+1]\times\Omega$.

For predictable processes X, we define the L^2 norm over the set $[0,T]\times\Omega$ under the measure μ_M by

$$\begin{aligned}\|X\|_{\mu_M,T} &= \bigg(\int_{[0,T]\times\Omega}|X|^2\,d\mu_M\bigg)^{1/2}\\ &= \bigg(E\int_{[0,T]}|X(t,\omega)|^2 d[M]_t(\omega)\bigg)^{1/2}.\end{aligned} \tag{5.3}$$

Let $\mathcal{L}_2 = \mathcal{L}_2(M,\mathcal{P})$ denote the collection of all predictable processes X such that $\|X\|_{\mu_M,T} < \infty$ for all $T<\infty$. A metric on $\mathcal{L}_2$ is defined by $d_{\mathcal{L}_2}(X,Y) = \|X-Y\|_{\mathcal{L}_2}$ where

$$\|X\|_{\mathcal{L}_2} = \sum_{k=1}^{\infty} 2^{-k}\big(1\wedge\|X\|_{\mu_M,k}\big). \tag{5.4}$$

$\mathcal{L}_2$ is not an L^2 space, but instead a local L^2 space of sorts. The discussion following definition (3.20) of the metric on martingales can be repeated here with obvious changes. In particular, to satisfy the requirement that $d_{\mathcal{L}_2}(X,Y)=0$ iff $X=Y$, we have to regard two processes X and Y in $\mathcal{L}_2$ as equal if

$$\mu_M\{(t,\omega) : X(t,\omega)\neq Y(t,\omega)\} = 0. \tag{5.5}$$

Let us say processes X and Y are *μ_M-equivalent* if (5.5) holds.

Example 5.4. For both Brownian motion and the compensated Poisson process, the form of μ_M tells us that a predictable process X lies in $\mathcal{L}_2$ if and only if

$$E\int_0^T X(s,\omega)^2\,ds < \infty \qquad \text{for all } T<\infty.$$

For Brownian motion this is the same integrability requirement as imposed for the space $\mathcal{L}_2(B)$ in Chapter 4, except that now we are restricted to predictable integrands while in Chapter 4 we were able to integrate more general measurable, adapted processes. This shortcoming will be fixed in Section 5.5.

Example 5.5. Suppose X is predictable and bounded on bounded time intervals, in other words there exist constants $C_T < \infty$ such that, for almost every ω and all $T < \infty$, $|X_t(\omega)| \le C_T$ for $0 \le t \le T$. Then $X \in \mathcal{L}_2(M, \mathcal{P})$ because

$$E \int_{[0,T]} X(s)^2 \, d[M]_s \le C_T^2 E\{[M]_T\} = C_T^2 E\{M_T^2 - M_0^2\} < \infty.$$

5.1.2. Construction of the stochastic integral. In this section we define the stochastic integral process $(X \cdot M)_t = \int_{(0,t]} X \, dM$ for integrands $X \in \mathcal{L}_2$. There are two steps: first an explicit definition of integrals for a class of processes with a particularly simple structure, and then an approximation step that defines the integral for a general $X \in \mathcal{L}_2$.

A *simple predictable process* is a process of the form

$$X_t(\omega) = \xi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{n-1} \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t) \tag{5.6}$$

where n is a finite integer, $0 = t_0 = t_1 < t_2 < \cdots < t_n$ are time points, ξ_i is a bounded $\mathcal{F}_{t_i}$-measurable random variable on $(\Omega, \mathcal{F}, P)$ for $0 \le i \le n-1$. We set $t_1 = t_0 = 0$ for convenience, so the formula for X covers the interval $[0, t_n]$ without leaving a gap at the origin.

Lemma 5.6. *A process of type (5.6) is predictable.*

Proof. Immediate from Lemma 5.1 and the left-continuity of X.

Alternatively, here is an elementary argument that shows that X is $\mathcal{P}$-measurable. For each ξ_i we can find $\mathcal{F}_{t_i}$-measurable simple functions

$$\eta_i^N = \sum_{j=1}^{m(i,N)} \beta_j^{i,N} \mathbf{1}_{F_j^{i,N}}$$

such that $\eta_i^N(\omega) \to \xi_i(\omega)$ as $N \to \infty$. Here $\beta_j^{i,N}$ are constants and $F_j^{i,N} \in \mathcal{F}_{t_i}$. Adding these up, we have that

$$\begin{aligned} X_t(\omega) &= \lim_{N\to\infty} \left\{ \eta_0^N(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{n-1} \eta_i^N(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t) \right\} \\ &= \lim_{N\to\infty} \left\{ \sum_{j=1}^{m(0,N)} \beta_j^{0,N} \mathbf{1}_{\{0\}\times F_j^{0,N}}(t,\omega) + \sum_{i=1}^{n-1} \sum_{j=1}^{m(i,N)} \beta_j^{i,N} \mathbf{1}_{(t_i,t_{i+1}]\times F_j^{i,N}}(t,\omega) \right\}. \end{aligned}$$

The last function is clearly $\mathcal{P}$-measurable, being a linear combination of indicator functions of predictable rectangles. Consequently X is $\mathcal{P}$-measurable as a pointwise limit of $\mathcal{P}$-measurable functions. ☐

Definition 5.7. For a simple predictable process of the type (5.6), the *stochastic integral* is the process $X \cdot M$ defined by

$$(X \cdot M)_t(\omega) = \sum_{i=1}^{n-1} \xi_i(\omega)\big(M_{t_{i+1}\wedge t}(\omega) - M_{t_i\wedge t}(\omega)\big). \tag{5.7}$$

Note that our convention is such that the value of X at $t = 0$ does not influence the integral. We also write $I(X) = X \cdot M$ when we need a symbol for the mapping $I : X \mapsto X \cdot M$.

Let $\mathcal{S}_2$ denote the subspace of $\mathcal{L}_2$ consisting of simple predictable processes. Any particular element X of $\mathcal{S}_2$ can be represented in the form (5.6) with many different choices of random variables and time intervals. The first thing to check is that the integral $X \cdot M$ depends only on the process X and not on the particular representation (5.6) used. Also, let us check that the space $\mathcal{S}_2$ is a linear space and the integral behaves linearly, since these properties are not immediately clear from the definitions.

Lemma 5.8. (a) *Suppose the process X in (5.6) also satisfies*

$$X_t(\omega) = \eta_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{j=1}^{m-1} \eta_j(\omega)\mathbf{1}_{(s_i,s_{i+1}]}(t)$$

for all (t,ω), where $0 = s_0 = s_1 < s_2 < \cdots < s_m < \infty$ and η_j is $\mathcal{F}_{s_j}$-measurable for $0 \le j \le m-1$. Then for each (t,ω),

$$\sum_{i=1}^{n-1} \xi_i(\omega)\big(M_{t_{i+1}\wedge t}(\omega) - M_{t_i\wedge t}(\omega)\big) = \sum_{j=1}^{m-1} \eta_i(\omega)\big(M_{s_{i+1}\wedge t}(\omega) - M_{s_i\wedge t}(\omega)\big).$$

In other words, $X \cdot M$ is independent of the representation.

(b) *$\mathcal{S}_2$ is a linear space, in other words for $X, Y \in \mathcal{S}_2$ and reals α and β, $\alpha X + \beta Y \in \mathcal{S}_2$. The integral satisfies*

$$(\alpha X + \beta Y) \cdot M = \alpha(X \cdot M) + \beta(Y \cdot M).$$

Proof. Part (a). We may assume $s_m = t_n$. For if say $t_n < s_m$, replace n by $n+1$, define $t_{n+1} = s_m$ and $\xi_n(\omega) = 0$, and add the term $\xi_n \mathbf{1}_{(t_n,t_{n+1}]}$ to the $\{\xi_i, t_i\}$-representation (5.6) of X. X did not change because the new term is zero. The stochastic integral $X \cdot M$ then acquires the term $\xi_n(M_{t_{n+1}\wedge t} - M_{t_n\wedge t})$ which is identically zero. Thus the new term in the representation does not change the value of either $X_t(\omega)$ or $(X \cdot M)_t(\omega)$.

Let $T = s_m = t_n$, and let $0 = u_1 < u_2 < \cdots < u_p = T$ be an ordered relabeling of the union $\{s_j : 1 \le j \le m\} \cup \{t_i : 1 \le i \le n\}$. Then for each $1 \le k \le p-1$ there are unique indices i and j such that

$$(u_k, u_{k+1}] = (t_i, t_{i+1}] \cap (s_j, s_{j+1}].$$

For $t \in (u_k, u_{k+1}]$, $X_t(\omega) = \xi_i(\omega)$ and $X_t(\omega) = \eta_j(\omega)$. So for these particular i and j, $\xi_i = \eta_j$.

The proof now follows from a reordering of the sums for the stochastic integrals.

$$\begin{aligned}
&\sum_{i=1}^{n-1} \xi_i (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) \\
&= \sum_{i=1}^{n-1} \xi_i \sum_{k=1}^{p-1} (M_{u_{k+1}\wedge t} - M_{u_k\wedge t}) \mathbf{1}\{(u_k, u_{k+1}] \subseteq (t_i, t_{i+1}]\} \\
&= \sum_{k=1}^{p-1} (M_{u_{k+1}\wedge t} - M_{u_k\wedge t}) \sum_{i=1}^{n-1} \xi_i \mathbf{1}\{(u_k, u_{k+1}] \subseteq (t_i, t_{i+1}]\} \\
&= \sum_{k=1}^{p-1} (M_{u_{k+1}\wedge t} - M_{u_k\wedge t}) \sum_{j=1}^{m-1} \eta_j \mathbf{1}\{(u_k, u_{k+1}] \subseteq (s_j, s_{j+1}]\} \\
&= \sum_{j=1}^{m-1} \eta_j \sum_{k=1}^{p-1} (M_{u_{k+1}\wedge t} - M_{u_k\wedge t}) \mathbf{1}\{(u_k, u_{k+1}] \subseteq (s_j, s_{j+1}]\} \\
&= \sum_{j=1}^{m-1} \eta_j (M_{s_{j+1}\wedge t} - M_{s_j\wedge t}).
\end{aligned}$$

Part (b). Suppose we are given two simple predictable processes

$$X_t = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{n-1} \xi_i \mathbf{1}_{(t_i, t_{i+1}]}(t)$$

and

$$Y_t = \eta_0 \mathbf{1}_{\{0\}}(t) + \sum_{j=1}^{m-1} \eta_j \mathbf{1}_{(s_i, s_{i+1}]}(t).$$

As above, we can assume that $T = t_n = s_m$ by adding a zero term to one of these processes. As above, let $\{u_k : 1 \le k \le p\}$ be the common refinement of $\{s_j : 1 \le j \le m\}$ and $\{t_i : 1 \le i \le n\}$, as partitions of $[0, T]$. Given $1 \le k \le p-1$, let $i = i(k)$ and $j = j(k)$ be the indices defined by

$$(u_k, u_{k+1}] = (t_i, t_{i+1}] \cap (s_j, s_{j+1}].$$

Define $\rho_k(\omega) = \xi_{i(k)}(\omega)$ and $\zeta_k(\omega) = \eta_{j(k)}(\omega)$. Then

$$X_t = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{p-1} \rho_k \mathbf{1}_{(u_k, u_{k+1}]}(t)$$

and

$$Y_t = \eta_0 \mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{p-1} \zeta_k \mathbf{1}_{(u_k, u_{k+1}]}(t).$$

The representation

$$\alpha X_t + \beta Y_t = (\alpha\xi_0 + \beta\eta_0)\mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{p-1}(\alpha\rho_k + \beta\zeta_k)\mathbf{1}_{(u_k, u_{k+1}]}(t)$$

shows that $\alpha X + \beta Y$ is a member of $\mathcal{S}_2$. According to part (a) proved above, we can write the stochastic integrals based on these representations, and then linearity of the integral is clear. □

To build a more general integral on definition (5.7), we need some continuity properties.

Lemma 5.9. *Let $X \in \mathcal{S}_2$. Then $X \cdot M$ is a square-integrable cadlag martingale. If M is continuous, then so is $X \cdot M$. These isometries hold: for all $t > 0$*

$$E\big[(X \cdot M)_t^2\big] = \int_{[0,t]\times\Omega} X^2 \, d\mu_M \tag{5.8}$$

and

$$\|X \cdot M\|_{\mathcal{M}_2} = \|X\|_{\mathcal{L}_2}. \tag{5.9}$$

Proof. The cadlag property for each fixed ω is clear from the definition (5.7) of $X \cdot M$, as is the continuity if M is continuous to begin with.

Linear combinations of martingales are martingales. So to prove that $X \cdot M$ is a martingale it suffices to check this statement: if M is a martingale, $u < v$ and ξ is a bounded $\mathcal{F}_u$-measurable random variable, then $Z_t = \xi(M_{t\wedge v} - M_{t\wedge u})$ is a martingale. The boundedness of ξ and integrability of M guarantee integrability of Z_t. Take $s < t$.

First, if $s < u$, then

$$\begin{aligned} E[Z_t|\mathcal{F}_s] &= E\big[\xi(M_{t\wedge v} - M_{t\wedge u}) \,\big|\, \mathcal{F}_s\big] \\ &= E\big[\xi E\{M_{t\wedge v} - M_{t\wedge u}|\mathcal{F}_u\} \,\big|\, \mathcal{F}_s\big] \\ &= 0 = Z_s \end{aligned}$$

because $M_{t\wedge v} - M_{t\wedge u} = 0$ for $t \le u$, and for $t > u$ the martingale property of M gives

$$E\{M_{t\wedge v} - M_{t\wedge u}|\mathcal{F}_u\} = E\{M_{t\wedge v}|\mathcal{F}_u\} - M_u = 0.$$

Second, if $s \ge u$, then also $t > s \ge u$, and it follows that ξ is $\mathcal{F}_s$-measurable and $M_{t\wedge u} = M_u = M_{s\wedge u}$ is also $\mathcal{F}_s$-measurable. Then

$$\begin{aligned} E[Z_t|\mathcal{F}_s] &= E\big[\xi(M_{t\wedge v} - M_{t\wedge u}) \bigm| \mathcal{F}_s\big] \\ &= \xi E\big[M_{t\wedge v} - M_{s\wedge u} \bigm| \mathcal{F}_s\big] \\ &= \xi\big(E[M_{t\wedge v}|\mathcal{F}_s] - M_{s\wedge u}\big) \\ &= \xi(M_{s\wedge v} - M_{s\wedge u}) = Z_s. \end{aligned}$$

In the last equality above, either $s \ge v$ in which case $M_{t\wedge v} = M_v = M_{s\wedge v}$ is $\mathcal{F}_s$-measurable, or $s < v$ in which case we use the martingale property of M.

We have proved that $X \cdot M$ is a martingale.

Next we prove (5.8). After squaring,

$$\begin{aligned} (X\cdot M)_t^2 = &\sum_{i=1}^{n-1} \xi_i^2 \big(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}\big)^2 \\ &+ 2\sum_{i<j} \xi_i\xi_j(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})(M_{t\wedge t_{j+1}} - M_{t\wedge t_j}). \end{aligned}$$

We claim that each term of the last sum has zero expectation. Since $i < j$, $t_{i+1} \le t_j$ and both ξ_i and ξ_j are $\mathcal{F}_{t_j}$-measurable.

$$\begin{aligned} &E\Big[\xi_i\xi_j(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})(M_{t\wedge t_{j+1}} - M_{t\wedge t_j})\Big] \\ &= E\big[\xi_i\xi_j(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})E\{M_{t\wedge t_{j+1}} - M_{t\wedge t_j}|\mathcal{F}_{t_j}\}\big] = 0 \end{aligned}$$

because the conditional expectation vanishes, either trivially if $t \le t_j$, or by the martingale property of M if $t > t_j$.

Now we can compute the mean of the square. Let $t > 0$. The key point of the next calculation is the fact that $M^2 - [M]$ is a martingale.

$$\begin{aligned} E\big[(X\cdot M)_t^2\big] &= \sum_{i=1}^{n-1} E\big[\xi_i^2\big(M_{t\wedge t_{i+1}} - M_{t\wedge t_i}\big)^2\big] \\ &= \sum_{i=1}^{n-1} E\big[\xi_i^2 E\big\{(M_{t\wedge t_{i+1}} - M_{t\wedge t_i})^2 \bigm| \mathcal{F}_{t_i}\big\}\big] \\ &= \sum_{i=1}^{n-1} E\big[\xi_i^2 E\big\{M_{t\wedge t_{i+1}}^2 - M_{t\wedge t_i}^2 \bigm| \mathcal{F}_{t_i}\big\}\big] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n-1} E\big[\xi_i^2 E\{[M]_{t\wedge t_{i+1}} - [M]_{t\wedge t_i} \,\big|\, \mathcal{F}_{t_i}\}\big] \\
&= \sum_{i=1}^{n-1} E\big[\xi_i^2 \big([M]_{t\wedge t_{i+1}} - [M]_{t\wedge t_i}\big)\big] \\
&= \sum_{i=1}^{n-1} E\Big[\xi_i^2 \int_{[0,t]} \mathbf{1}_{(t_i,t_{i+1}]}(s)\, d[M]_s\Big] \\
&= E\Big[\int_{[0,t]} \Big(\xi_0^2 \mathbf{1}_{\{0\}}(s) + \sum_{i=1}^{n-1} \xi_i^2 \mathbf{1}_{(t_i,t_{i+1}]}(s)\Big)\, d[M]_s\Big] \\
&= E\Big[\int_{[0,t]} \Big(\xi_0 \mathbf{1}_{\{0\}}(s) + \sum_{i=1}^{n-1} \xi_i \mathbf{1}_{(t_i,t_{i+1}]}(s)\Big)^2\, d[M]_s\Big] \\
&= \int_{[0,t]\times\Omega} X^2 d\mu_M.
\end{aligned}
$$

In the third last equality we added the term $\xi_0^2 \mathbf{1}_{\{0\}}(s)$ inside the $d[M]_s$-integral because this term integrates to zero (recall that $\Lambda_{[M]}\{0\} = 0$). In the second last equality we used the equality

$$
\xi_0^2 \mathbf{1}_{\{0\}}(s) + \sum_{i=1}^{n-1} \xi_i^2 \mathbf{1}_{(t_i,t_{i+1}]}(s) = \Big(\xi_0 \mathbf{1}_{\{0\}}(s) + \sum_{i=1}^{n-1} \xi_i \mathbf{1}_{(t_i,t_{i+1}]}(s)\Big)^2
$$

which is true due to the pairwise disjointness of the time intervals.

The above calculation checks that

$$
\|(X\cdot M)_t\|_{L^2(P)} = \|X\|_{\mu_M,t}
$$

for any $t > 0$. Comparison of formulas (3.20) and (5.4) then proves (5.9). $\square$

Let us summarize the message of Lemmas 5.8 and 5.9 in words. The stochastic integral $I : X \mapsto X \cdot M$ is a linear map from the space $\mathcal{S}_2$ of predictable simple processes into $\mathcal{M}_2$. Equality (5.9) says that this map is a linear isometry that maps from the subspace $(\mathcal{S}_2, d_{\mathcal{L}_2})$ of the metric space $(\mathcal{L}_2, d_{\mathcal{L}_2})$, and into the metric space $(\mathcal{M}_2, d_{\mathcal{M}_2})$. In case M is continuous, the map goes into the space $(\mathcal{M}_2^c, d_{\mathcal{M}_2})$.

A consequence of (5.9) is that if X and Y satisfy (5.5) then $X\cdot M$ and $Y\cdot M$ are indistinguishable. For example, we may have $Y_t = X_t + \zeta \mathbf{1}\{t=0\}$ for a bounded $\mathcal{F}_0$-measurable random variable ζ. Then the integrals $X\cdot M$ and $Y\cdot M$ are indistinguishable, in other words the same process. This is no different from the analytic fact that changing the value of a function f on $[a,b]$ at a single point (or even at countably many points) does not affect the value of the integral $\int_a^b f(x)\,dx$.

We come to the approximation step.

Lemma 5.10. *For any $X \in \mathcal{L}_2$ there exists a sequence $X_n \in \mathcal{S}_2$ such that $\|X - X_n\|_{\mathcal{L}_2} \to 0$.*

Proof. Let $\widetilde{\mathcal{L}}_2$ denote the class of $X \in \mathcal{L}_2$ for which this approximation is possible. Of course $\mathcal{S}_2$ itself is a subset of $\widetilde{\mathcal{L}}_2$.

Indicator functions of time-bounded predictable rectangles are of the form

$$\mathbf{1}_{\{0\}\times F_0}(t,\omega) = \mathbf{1}_{F_0}(\omega)\mathbf{1}_{\{0\}}(t),$$

or

$$\mathbf{1}_{(u,v]\times F}(t,\omega) = \mathbf{1}_{F}(\omega)\mathbf{1}_{(u,v]}(t),$$

for $F_0 \in \mathcal{F}_0$, $0 \le u < v < \infty$, and $F \in \mathcal{F}_u$. They are elements of $\mathcal{S}_2$ due to (5.2). Furthermore, since $\mathcal{S}_2$ is a linear space, it contains all simple functions of the form

$$X(t,\omega) = \sum_{i=0}^{n} c_i \mathbf{1}_{R_i}(t,\omega) \tag{5.10}$$

where $\{c_i\}$ are finite constants and $\{R_i\}$ are time-bounded predictable rectangles.

The approximation of predictable processes proceeds from constant multiples of indicator functions of predictable sets through $\mathcal{P}$-measurable simple functions to the general case.

Step 1. Let $G \in \mathcal{P}$ be an arbitrary set and $c \in \mathbf{R}$. We shall show that $X = c\mathbf{1}_G$ lies in $\widetilde{\mathcal{L}}_2$. We can assume $c \neq 0$, otherwise $X = 0 \in \mathcal{S}_2$. Again by (5.2) $c\mathbf{1}_G \in \mathcal{L}_2$ because

$$\|c\mathbf{1}_G\|_{\mu_M,T} = |c| \cdot \mu\big(G \cap ([0,T]\times\Omega)\big)^{1/2} < \infty$$

for all $T < \infty$.

Given $\varepsilon > 0$ fix n large enough so that $2^{-n} < \varepsilon/2$. Let $G_n = G \cap ([0,n]\times\Omega)$. Consider the restricted σ-algebra

$$\mathcal{P}_n = \{A \in \mathcal{P} : A \subseteq [0,n]\times\Omega\} = \{B \cap ([0,n]\times\Omega) : B \in \mathcal{P}\}.$$

$\mathcal{P}_n$ is generated by the collection $\mathcal{R}_n$ of predictable rectangles that lie in $[0,n]\times\Omega$ (Exercise 1.8 part (d)). $\mathcal{R}_n$ is a semialgebra in the space $[0,n]\times\Omega$. (For this reason it is convenient to regard $\emptyset$ as a member of $\mathcal{R}_n$.) The algebra $\mathcal{A}_n$ generated by $\mathcal{R}_n$ is the collection of all finite disjoint unions of members of $\mathcal{R}_n$ (Lemma B.1). Restricted to $[0,n]\times\Omega$, μ_M is a finite measure. Thus by Lemma B.2 there exists $R \in \mathcal{A}_n$ such that $\mu_M(G_n \triangle R) < |c|^{-2}\varepsilon^2/4$. We can write $R = R_1 \cup \cdots \cup R_p$ as a finite disjoint union of time-bounded predictable rectangles.

Let $Z = c\mathbf{1}_R$. By the disjointness,

$$Z = c\mathbf{1}_R = c\mathbf{1}\Big\{\bigcup_{i=1}^p R_i\Big\} = \sum_{i=1}^p c\mathbf{1}_{R_i}$$

so in fact Z is of type (5.10) and a member of $\mathcal{S}_2$. The $\mathcal{L}_2$-distance between $Z = c\mathbf{1}_R$ and $X = c\mathbf{1}_G$ is now estimated as follows.

$$\begin{aligned}
\|Z - X\|_{\mathcal{L}_2} &\le \sum_{k=1}^n 2^{-k}\|c\mathbf{1}_R - c\mathbf{1}_G\|_{\mu_M,k} + 2^{-n} \\
&\le \sum_{k=1}^n 2^{-k}|c|\Big(\int_{[0,k]\times\Omega} |\mathbf{1}_R - \mathbf{1}_G|^2\,d\mu_M\Big)^{1/2} + \varepsilon/2 \\
&\le |c|\Big(\int_{[0,n]\times\Omega} |\mathbf{1}_R - \mathbf{1}_{G_n}|^2\,d\mu_M\Big)^{1/2} + \varepsilon/2 \\
&= |c|\mu_M(G_n \triangle R)^{1/2} + \varepsilon/2 \le \varepsilon.
\end{aligned}$$

Above we bounded integrals over $[0,k]\times\Omega$ with the integral over $[0,n]\times\Omega$ for $1 \le k \le n$, then noted that $\mathbf{1}_G(t,\omega) = \mathbf{1}_{G_n}(t,\omega)$ for $(t,\omega) \in [0,n]\times\Omega$, and finally used the general fact that

$$|\mathbf{1}_A - \mathbf{1}_B| = \mathbf{1}_{A\triangle B}$$

for any two sets A and B.

To summarize, we have shown that given $G \in \mathcal{P}$, $c \in \mathbf{R}$, and $\varepsilon > 0$, there exists a process $Z \in \mathcal{S}_2$ such that $\|Z - c\mathbf{1}_G\|_{\mathcal{L}_2} \le \varepsilon$. Consequently $c\mathbf{1}_G \in \widetilde{\mathcal{L}}_2$.

Let us observe that $\widetilde{\mathcal{L}}_2$ is closed under addition. Let $X, Y \in \widetilde{\mathcal{L}}_2$ and $X_n, Y_n \in \mathcal{S}_2$ be such that $\|X_n - X\|_{\mathcal{L}_2}$ and $\|Y_n - Y\|_{\mathcal{L}_2}$ vanish as $n \to \infty$. Then $X_n + Y_n \in \mathcal{S}_2$ and by the triangle inequality

$$\|(X+Y) - (X_n + Y_n)\|_{\mathcal{L}_2} \le \|X_n - X\|_{\mathcal{L}_2} + \|X_n - X\|_{\mathcal{L}_2} \to 0.$$

From this and the proof for $c\mathbf{1}_G$ we conclude that all simple functions of the type

$$X = \sum_{i=1}^n c_i\mathbf{1}_{G_i}, \quad \text{with } c_i \in \mathbf{R} \text{ and } G_i \in \mathcal{P}, \tag{5.11}$$

lie in $\widetilde{\mathcal{L}}_2$.

Step 2. Let X be an arbitrary process in $\mathcal{L}_2$. Given $\varepsilon > 0$, pick n so that $2^{-n} < \varepsilon/3$. Find simple functions X_m of the type (5.11) such that $|X - X_m| \le |X|$ and $X_m(t,\omega) \to X(t,\omega)$ for all (t,ω). This is just an instance of the general approximation of measurable functions with simple functions, as for example in (1.3). Since $X \in L^2([0,n]\times\Omega, \mathcal{P}_n, \mu_M)$, Lebesgue's dominated convergence theorem implies for $1 \le k \le n$ that

$$\limsup_{m\to\infty} \|X - X_m\|_{\mu_M,k} \le \lim_{m\to\infty} \biggl(\int_{[0,n]\times\Omega} |X - X_m|^2 \, d\mu_M \biggr)^{1/2} = 0.$$

Consequently

$$\limsup_{m\to\infty} \|X - X_m\|_{\mathcal{L}_2} \le \sum_{k=1}^{n} 2^{-k} \limsup_{m\to\infty} \|X - X_m\|_{\mu_M,k} + \varepsilon/3 = \varepsilon/3.$$

Fix m large enough so that $\|X - X_m\|_{\mathcal{L}_2} \le \varepsilon/2$. Using **Step 1** find a process $Z \in \mathcal{S}_2$ such that $\|X_m - Z\|_{\mathcal{L}_2} < \varepsilon/2$. Then by the triangle inequality $\|X - Z\|_{\mathcal{L}_2} \le \varepsilon$. We have shown that an arbitrary process $X \in \mathcal{L}_2$ can be approximated by simple predictable processes in the $\mathcal{L}_2$-distance. □

Now we can state formally the definition of the stochastic integral.

Definition 5.11. Let M be a square-integrable cadlag martingale on a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}$. For any predictable process $X \in \mathcal{L}_2(M, \mathcal{P})$, the *stochastic integral* $I(X) = X \cdot M$ is the square-integrable cadlag martingale that satisfies

$$\lim_{n\to\infty} \|X \cdot M - X_n \cdot M\|_{\mathcal{M}_2} = 0$$

for every sequence $X_n \in \mathcal{S}_2$ of simple predictable processes such that $\|X - X_n\|_{\mathcal{L}_2} \to 0$. The process $I(X)$ is unique up to indistinguishability. If M is continuous, then so is $X \cdot M$.

Justification of the definition. Here is the argument that justifies the claims implicit in the definition. It is really the classic argument about extending a uniformly continuous map into a complete metric space to the closure of its domain.

Existence. Let $X \in \mathcal{L}_2$. By Lemma 5.10 there exists a sequence $X_n \in \mathcal{S}_2$ such that $\|X - X_n\|_{\mathcal{L}_2} \to 0$. From the triangle inequality it then follows that $\{X_n\}$ is a Cauchy sequence in $\mathcal{L}_2$: given $\varepsilon > 0$, choose n_0 so that $\|X - X_n\|_{\mathcal{L}_2} \le \varepsilon/2$ for $n \ge n_0$. Then if $m, n \ge n_0$,

$$\|X_m - X_n\|_{\mathcal{L}_2} \le \|X_m - X\|_{\mathcal{L}_2} + \|X - X_n\|_{\mathcal{L}_2} \le \varepsilon.$$

For $X_n \in \mathcal{S}_2$ the stochastic integral $X_n \cdot M$ was defined in (5.7). By the isometry (5.9) and the additivity of the integral,

$$\|X_m \cdot M - X_n \cdot M\|_{\mathcal{M}_2} = \|(X_m - X_n) \cdot M\|_{\mathcal{M}_2} = \|X_m - X_n\|_{\mathcal{L}_2}.$$

Consequently $\{X_n \cdot M\}$ is a Cauchy sequence in the space $\mathcal{M}_2$ of martingales. If M is continuous this Cauchy sequence lies in the space $\mathcal{M}_2^c$. By Theorem

3.41 these spaces are complete metric spaces, and consequently there exists a limit process $Y = \lim_{n\to\infty} X_n \cdot M$. This process Y we call $I(X) = X \cdot M$.

Uniqueness. Let Z_n be another sequence in $\mathcal{S}_2$ that converges to X in $\mathcal{L}_2$. We need to show that $Z_n \cdot M$ converges to the same $Y = X \cdot M$ in $\mathcal{M}_2$. This follows again from the triangle inequality and the isometry:

$$\begin{aligned}
&\|Y - Z_n \cdot M\|_{\mathcal{M}_2} \le \|Y - X_n \cdot M\|_{\mathcal{M}_2} + \|X_n \cdot M - Z_n \cdot M\|_{\mathcal{M}_2} \\
&= \|Y - X_n \cdot M\|_{\mathcal{M}_2} + \|X_n - Z_n\|_{\mathcal{L}_2} \\
&\le \|Y - X_n \cdot M\|_{\mathcal{M}_2} + \|X_n - X\|_{\mathcal{L}_2} + \|X - Z_n\|_{\mathcal{L}_2}.
\end{aligned}$$

All terms on the last line vanish as $n \to \infty$. This shows that $Z_n \cdot M \to Y$, and so there is only one process $Y = X \cdot M$ that satisfies the description of the definition.

Note that the uniqueness of the stochastic integral cannot hold in a sense stronger than indistinguishability. If W is a process that is indistinguishable from $X \cdot M$, which meant that

$$P\{\omega : W_t(\omega) = (X \cdot M)_t(\omega) \quad \text{for all } t \in \mathbf{R}_+\} = 1,$$

then W also has to be regarded as the stochastic integral. This is built into the definition of $I(X)$ as the limit: if $\|X \cdot M - X_n \cdot M\|_{\mathcal{M}_2} \to 0$ and W is indistinguishable from $X \cdot M$, then also $\|W - X_n \cdot M\|_{\mathcal{M}_2} \to 0$. ☐

The definition of the stochastic integral $X \cdot M$ feels somewhat abstract because the approximation happens in a space of processes, and it may not seem obvious how to produce the approximating predictable simple processes X_n. When X is caglad, one can use Riemann sum type approximations with X-values evaluated at left endpoints of partition intervals. To get $\mathcal{L}_2$ approximation, one must truncate the process, and then let the mesh of the partition shrink fast enough and the number of terms in the simple process grow fast enough. See Proposition 5.32 and Exercise 5.13.

We took the approximation step in the space of martingales to avoid separate arguments for the path properties of the integral. The completeness of the space of cadlag martingales and the space of continuous martingales gives immediately a stochastic integral with the appropriate path regularity.

As for the style of convergence in the definition of the integral, let us recall that convergence in the spaces of processes actually reduces back to familiar mean-square convergence. $\|X_n - X\|_{\mathcal{L}_2} \to 0$ is equivalent to having

$$\int_{[0,T]\times\Omega} |X_n - X|^2 \, d\mu_M \to 0 \quad \text{for all } T < \infty.$$

Convergence in $\mathcal{M}_2$ is equivalent to $L^2(P)$ convergence at each fixed time t: for martingales $N^{(j)}$, $N \in \mathcal{M}_2$,

$$\|N^{(j)} - N\|_{\mathcal{M}_2} \to 0$$

if and only if

$$E\big[\,(N_t^{(j)} - N_t)^2\big] \to 0 \quad \text{for each } t \ge 0.$$

In particular, at each time $t \ge 0$ the integral $(X \cdot M)_t$ is the mean-square limit of the integrals $(X_n \cdot M)_t$ of approximating simple processes. These observations are used in the extension of the isometric property of the integral.

Proposition 5.12. *Let $M \in \mathcal{M}_2$ and $X \in \mathcal{L}_2(M, \mathcal{P})$. Then we have the isometries*

$$E\big[(X \cdot M)_t^2\big] = \int_{[0,t]\times\Omega} X^2 \, d\mu_M \quad \textit{for all } t \ge 0, \tag{5.12}$$

and

$$\|X \cdot M\|_{\mathcal{M}_2} = \|X\|_{\mathcal{L}_2}. \tag{5.13}$$

In particular, if $X, Y \in \mathcal{L}_2(M, \mathcal{P})$ are μ_M-equivalent in the sense (5.5), then $X \cdot M$ and $Y \cdot M$ are indistinguishable.

Proof. As already observed, the triangle inequality is valid for the distance measures $\|\cdot\|_{\mathcal{L}_2}$ and $\|\cdot\|_{\mathcal{M}_2}$. From this we get a continuity property. Let $Z, W \in \mathcal{L}_2$.

$$\begin{aligned} \|Z\|_{\mathcal{L}_2} - \|W\|_{\mathcal{L}_2} &\le \|Z - W\|_{\mathcal{L}_2} + \|W\|_{\mathcal{L}_2} - \|W\|_{\mathcal{L}_2} \\ &\le \|Z - W\|_{\mathcal{L}_2}. \end{aligned}$$

This and the same inequality with Z and W switched give

$$\big|\, \|Z\|_{\mathcal{L}_2} - \|W\|_{\mathcal{L}_2} \big| \le \|Z - W\|_{\mathcal{L}_2}. \tag{5.14}$$

This same calculation applies to $\|\cdot\|_{\mathcal{M}_2}$ also, and of course equally well to the L^2 norms on Ω and $[0, T] \times \Omega$.

Let $X_n \in \mathcal{S}_2$ be a sequence such that $\|X_n - X\|_{\mathcal{L}_2} \to 0$. As we proved in Lemma 5.9, the isometries hold for $X_n \in \mathcal{S}$. Consequently to prove the proposition we need only let $n \to \infty$ in the equalities

$$E\big[(X_n \cdot M)_t^2\big] = \int_{[0,t]\times\Omega} X_n^2 \, d\mu_M$$

and

$$\|X_n \cdot M\|_{\mathcal{M}_2} = \|X_n\|_{\mathcal{L}_2}$$

that come from Lemma 5.9. Each term converges to the corresponding term with X_n replaced by X.

The last statement of the proposition follows because $\|X - Y\|_{\mathcal{L}_2} = 0$ iff X and Y are μ_M-equivalent, and $\|X \cdot M - Y \cdot M\|_{\mathcal{M}_2} = 0$ iff $X \cdot M$ and $Y \cdot M$ are indistinguishable. □

Remark 5.13 (Enlarging the filtration)**.** Throughout we assume that M is a cadlag martingale. By Proposition 3.2, if our original filtration $\{\mathcal{F}_t\}$ is not already right-continuous, we can replace it with the larger filtration $\{\mathcal{F}_{t+}\}$. Under the filtration $\{\mathcal{F}_{t+}\}$ we have more predictable rectangles than before, and hence $\mathcal{P}_+$ (the predictable σ-field defined in terms of $\{\mathcal{F}_{t+}\}$) is potentially larger than our original predictable σ-field $\mathcal{P}$. The relevant question is whether switching to $\{\mathcal{F}_{t+}\}$ and $\mathcal{P}_+$ gives us more processes X to integrate? The answer is essentially no. Only the value at $t = 0$ of a $\mathcal{P}_+$-measurable process differentiates it from a $\mathcal{P}$-measurable process (Exercise 5.5). And as already seen, the value X_0 is irrelevant for the stochastic integral.

5.1.3. Properties of the stochastic integral. We prove here basic properties of the L^2 integral $X \cdot M$ constructed in Definition 5.11. Many of these properties really amount to saying that the notation works the way we would expect it to work. Those properties that take the form of an equality between two stochastic integrals are interpreted in the sense that the two processes are indistinguishable. Since the stochastic integrals are cadlag processes, indistinguishability follows from showing almost sure equality at all fixed times (Lemma 2.5).

The stochastic integral $X \cdot M$ was defined as a limit $X_n \cdot M \to X \cdot M$ in $\mathcal{M}_2$-space, where $X_n \cdot M$ are stochastic integrals of approximating simple predictable processes X_n. Recall that this implies that for a fixed time t, $(X \cdot M)_t$ is the $L^2(P)$-limit of the random variables $(X_n \cdot M)_t$. And furthermore, there is uniform convergence in probability on compact intervals:

$$\lim_{n\to\infty} P\Big\{ \sup_{0\le t\le T} \big|(X_n \cdot M)_t - (X \cdot M)_t\big| \ge \varepsilon \Big\} = 0 \tag{5.15}$$

for each $\varepsilon > 0$ and $T < \infty$. By the Borel-Cantelli lemma, along some subsequence $\{n_j\}$ there is almost sure convergence uniformly on compact time intervals: for P-almost every ω

$$\lim_{n\to\infty} \sup_{0\le t\le T} \big|(X_{n_j} \cdot M)_t(\omega) - (X \cdot M)_t(\omega)\big| = 0 \quad \text{for each } T < \infty.$$

These last two statements are general properties of $\mathcal{M}_2$-convergence, see Lemma 3.42.

A product of functions of t, ω, and (t, ω) is regarded as a process in the obvious sense: for example, if X is a process, Z is a random variable and f is a function on $\mathbf{R}_+$, then fZX is the process whose value at (t, ω) is $f(t)Z(\omega)X(t, \omega)$. This just amounts to taking some liberties with the notation: we do not distinguish notationally between the function $t \mapsto f(t)$ on $\mathbf{R}_+$ and the function $(t,\omega) \mapsto f(t)$ on $\mathbf{R}_+ \times \Omega$.

Throughout these proofs, when $X_n \in \mathcal{S}_2$ approximates $X \in \mathcal{L}_2$, we write X_n generically in the form

$$X_n(t,\omega) = \xi_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{k-1} \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t). \tag{5.16}$$

We introduce the familiar integral notation through the definition

$$\int_{(s,t]} X\,dM = (X\cdot M)_t - (X\cdot M)_s \quad \text{for } 0 \le s \le t. \tag{5.17}$$

To explicitly display either the time variable (the integration variable), or the sample point ω, this notation has variants such as

$$\int_{(s,t]} X\,dM = \int_{(s,t]} X_u\,dM_u = \int_{(s,t]} X_u(\omega)\,dM_u(\omega).$$

When the martingale M is continuous, we can also write

$$\int_s^t X\,dM$$

because then including or excluding endpoints of the interval make no difference (Exercise 5.6). We shall alternate freely between the different notations for the stochastic integral, using whichever seems more clear, compact or convenient.

Since $(X\cdot M)_0 = 0$ for any stochastic integral,

$$\int_{(0,t]} X\,dM = (X\cdot M)_t.$$

It is more accurate to use the interval $(0,t]$ above rather than $[0,t]$ because the integral does not take into consideration any jump of the martingale at the origin. Precisely, if ζ is an $\mathcal{F}_0$-measurable random variable and $\widetilde{M}_t = \zeta + M_t$, then $[\widetilde{M}] = [M]$, the spaces $\mathcal{L}_2(\widetilde{M},\mathcal{P})$ and $\mathcal{L}(M,\mathcal{P})$ coincide, and $X\cdot\widetilde{M} = X\cdot M$ for each admissible integrand. (It starts with definition (5.7).)

An integral of the type

$$\int_{(u,v]} G(s,\omega)\,d[M]_s(\omega)$$

is interpreted as a path-by-path Lebesgue-Stieltjes integral (in other words, evaluated as an ordinary Lebesgue-Stieltjes integral over $(u,v]$ separately for each fixed ω).

Proposition 5.14. (a) *Linearity:*

$$(\alpha X + \beta Y) \cdot M = \alpha(X \cdot M) + \beta(Y \cdot M).$$

(b) *For any* $0 \le u \le v$,

$$\int_{(0,t]} \mathbf{1}_{[0,v]} X \, dM = \int_{(0,v\wedge t]} X \, dM \tag{5.18}$$

and

$$\begin{aligned}\int_{(0,t]} \mathbf{1}_{(u,v]} X \, dM &= (X \cdot M)_{v\wedge t} - (X \cdot M)_{u\wedge t} \\ &= \int_{(u\wedge t, v\wedge t]} X \, dM.\end{aligned} \tag{5.19}$$

The inclusion or exclusion of the origin in the interval $[0, v]$ *is immaterial because a process of the type* $\mathbf{1}_{\{0\}}(t)X(t,\omega)$ *for* $X \in \mathcal{L}_2(M, \mathcal{P})$ *is* μ_M*-equivalent to the identically zero process, and hence has zero stochastic integral.*

(c) *For* $s < t$, *we have a conditional form of the isometry:*

$$E\big[\big((X \cdot M)_t - (X \cdot M)_s\big)^2 \mid \mathcal{F}_s\big] = E\bigg[\int_{(s,t]} X_u^2 \, d[M]_u \;\bigg|\; \mathcal{F}_s\bigg]. \tag{5.20}$$

This implies that

$$(X \cdot M)_t^2 - \int_{(0,t]} X_u^2 \, d[M]_u$$

is a martingale.

Proof. Part (a). Take limits in Lemma 5.8(b).

Part (b). If $X_n \in \mathcal{S}_2$ approximate X, then

$$\mathbf{1}_{[0,v]}(t)X_n(t) = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{k-1} \xi_i \mathbf{1}_{(t_i\wedge v, t_{i+1}\wedge v]}(t)$$

are simple predictable processes that approximate $\mathbf{1}_{[0,v]}X$.

$$\begin{aligned}\big((\mathbf{1}_{[0,v]}X_n) \cdot M\big)_t &= \sum_{i=1}^{k-1} \xi_i (M_{t_{i+1}\wedge v\wedge t} - M_{t_i\wedge v\wedge t}) \\ &= (X_n \cdot M)_{v\wedge t}.\end{aligned}$$

Letting $n \to \infty$ along a suitable subsequence gives in the limit the almost sure equality

$$\big((\mathbf{1}_{[0,v]}X) \cdot M\big)_t = (X \cdot M)_{v\wedge t}$$

which is (5.18). The second part (5.19) comes from $\mathbf{1}_{(u,v]}X = \mathbf{1}_{[0,v]}X - \mathbf{1}_{[0,u]}X$, the additivity of the integral, and definition (5.17).

Part (c). First we check this for the simple process X_n in (5.16). This is essentially a redoing of the calculations in the proof of Lemma 5.9. Let $s < t$. If $s \ge s_k$ then both sides of (5.20) are zero. Otherwise, fix an index $1 \le m \le k-1$ such that $t_m \le s < t_{m+1}$. Then

$$\begin{aligned}(X_n \cdot M)_t - (X_n \cdot M)_s &= \xi_m (M_{t_{m+1}\wedge t} - M_s) + \sum_{i=m+1}^{k-1} \xi_i (M_{t_{i+1}\wedge t} - M_{t_i\wedge t}) \\ &= \sum_{i=m}^{k-1} \xi_i (M_{u_{i+1}\wedge t} - M_{u_i\wedge t})\end{aligned}$$

where we have temporarily defined $u_m = s$ and $u_i = t_i$ for $i > m$. After squaring,

$$\begin{aligned}\big((X_n \cdot M)_t - (X_n \cdot M)_s\big)^2 &= \sum_{i=m}^{k-1} \xi_i^2 \big(M_{u_{i+1}\wedge t} - M_{u_i\wedge t}\big)^2 \\ &+ 2 \sum_{m\le i<j<k} \xi_i \xi_j (M_{u_{i+1}\wedge t} - M_{u_i\wedge t})(M_{u_{j+1}\wedge t} - M_{u_j\wedge t}).\end{aligned}$$

We claim that the cross terms vanish under the conditional expectation. Since $i < j$, $u_{i+1} \le u_j$ and both ξ_i and ξ_j are $\mathcal{F}_{u_j}$-measurable.

$$\begin{aligned}&E\big[\xi_i \xi_j (M_{u_{i+1}\wedge t} - M_{u_i\wedge t})(M_{u_{j+1}\wedge t} - M_{u_j\wedge t}) \;\big|\; \mathcal{F}_s\big] \\ &= E\big[\xi_i \xi_j (M_{u_{i+1}\wedge t} - M_{u_i\wedge t}) E\{M_{u_{j+1}\wedge t} - M_{u_j\wedge t} | \mathcal{F}_{u_j}\} \;\big|\; \mathcal{F}_s\big] = 0\end{aligned}$$

because the inner conditional expectation vanishes by the martingale property of M.

Now we can compute the conditional expectation of the square.

$$\begin{aligned}E\Big[\big((X_n \cdot M)_t - (X_n \cdot M)_s\big)^2 \;\Big|\; \mathcal{F}_s\Big] &= \sum_{i=m}^{k-1} E\Big[\xi_i^2 \big(M_{u_{i+1}\wedge t} - M_{u_i\wedge t}\big)^2 \;\Big|\; \mathcal{F}_s\Big] \\ &= \sum_{i=m}^{k-1} E\Big[\xi_i^2 E\big\{(M_{u_{i+1}\wedge t} - M_{u_i\wedge t})^2 \;\big|\; \mathcal{F}_{u_i}\big\} \;\Big|\; \mathcal{F}_s\Big] \\ &= \sum_{i=m}^{k-1} E\Big[\xi_i^2 E\big\{M^2_{u_{i+1}\wedge t} - M^2_{u_i\wedge t} \;\big|\; \mathcal{F}_{u_i}\big\} \;\Big|\; \mathcal{F}_s\Big] \\ &= \sum_{i=m}^{k-1} E\Big[\xi_i^2 E\big\{[M]_{u_{i+1}\wedge t} - [M]_{u_i\wedge t} \;\big|\; \mathcal{F}_{u_i}\big\} \;\Big|\; \mathcal{F}_s\Big] \\ &= \sum_{i=m}^{k-1} E\Big[\xi_i^2 \big([M]_{u_{i+1}\wedge t} - [M]_{u_i\wedge t}\big) \;\Big|\; \mathcal{F}_s\Big]\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=m}^{k-1} E\Big[\xi_i^2 \int_{(s,t]} \mathbf{1}_{(u_i,u_{i+1}]}(u)\, d[M]_u \;\Big|\; \mathcal{F}_s\Big] \\
&= E\Big[\int_{(s,t]} \Big(\xi_0 \mathbf{1}_{\{0\}}(u) + \sum_{i=1}^{k-1} \xi_i^2 \mathbf{1}_{(t_i,t_{i+1}]}(u)\Big)\, d[M]_u \;\Big|\; \mathcal{F}_s\Big] \\
&= E\Big[\int_{(s,t]} \Big(\xi_0 \mathbf{1}_{\{0\}}(u) + \sum_{i=1}^{k-1} \xi_i \mathbf{1}_{(t_i,t_{i+1}]}(u)\Big)^2\, d[M]_u \;\Big|\; \mathcal{F}_s\Big] \\
&= E\Big[\int_{(s,t]} X_n(u,\omega)^2\, d[M]_u(\omega) \;\Big|\; \mathcal{F}_s\Big]
\end{aligned}
$$

Inside the $d[M]_u$ integral above we replaced the u_i's with t_i's because for $u \in (s,t]$, $\mathbf{1}_{(u_i,u_{i+1}]}(u) = \mathbf{1}_{(t_i,t_{i+1}]}(u)$. Also, we brought in the terms for $i < m$ because these do not influence the integral, as they are supported on $[0, t_m]$ which is disjoint from $(s,t]$.

Next let $X \in \mathcal{L}_2$ be general, and $X_n \to X$ in $\mathcal{L}_2$. The limit $n \to \infty$ is best taken with expectations, so we rewrite the conclusion of the previous calculation as

$$
E\big[((X_n \cdot M)_t - (X_n \cdot M)_s)^2 \mathbf{1}_A\big] = E\Big[\mathbf{1}_A \int_{(s,t]} X_n^2(u)\, d[M]_u\Big]
$$

for an arbitrary $A \in \mathcal{F}_s$. Rewrite this once again as

$$
E\big[(X_n \cdot M)_t^2\, \mathbf{1}_A\big] - E\big[(X_n \cdot M)_s^2\, \mathbf{1}_A\big] = \int_{(s,t]\times A} X_n^2\, d\mu_M.
$$

All terms in this equality converge to the corresponding integrals with X_n replaced by X, because $(X_n \cdot M)_t \to (X \cdot M)_t$ in $L^2(P)$ and $X_n \to X$ in $L^2((0,t] \times \Omega, \mu_M)$ (see Lemma A.16 in the Appendix for the general idea). As $A \in \mathcal{F}_s$ is arbitrary, (5.20) is proved. □

Given stopping times σ and τ we can define various *stochastic intervals*. These are subsets of $\mathbf{R}_+ \times \Omega$. Here are two examples which are elements of $\mathcal{P}$:

$$
[0, \tau] = \{(t,\omega) \in \mathbf{R}_+ \times \Omega : 0 \le t \le \tau(\omega)\},
$$

and

$$
(\sigma, \tau] = \{(t,\omega) \in \mathbf{R}_+ \times \Omega : \sigma(\omega) < t \le \tau(\omega)\}.
$$

If $\tau(\omega) = \infty$, the ω-section of $[0,\tau]$ is $[0,\infty)$, because (∞, ω) is not a point in the space $\mathbf{R}_+ \times \Omega$. If $\sigma(\omega) = \infty$ then the ω-section of $(\sigma,\tau]$ is empty. The path $t \mapsto \mathbf{1}_{[0,\tau]}(t,\omega) = \mathbf{1}_{[0,\tau(\omega)]}(t)$ is adapted and left-continuous with right limits. Hence by Lemma 5.1 the indicator function $\mathbf{1}_{[0,\tau]}$ is $\mathcal{P}$-measurable. The same goes for $\mathbf{1}_{(\sigma,\tau]}$. If X is a predictable process, then so is the product $\mathbf{1}_{[0,\tau]} X$.

Recall also the notion of a stopped process M^τ defined by $M_t^\tau = M_{\tau\wedge t}$. If $M \in \mathcal{M}_2$ then also $M^\tau \in \mathcal{M}_2$, because Lemma 3.5 implies

$$E[M_{\tau\wedge t}^2] \le 2E[M_t^2] + E[M_0^2].$$

We insert a lemma on the effect of stopping on the Doléans measure.

Lemma 5.15. *Let $M \in \mathcal{M}_2$ and τ a stopping time. Then for any $\mathcal{P}$-measurable nonnegative function Y,*

$$\int_{\mathbf{R}_+\times\Omega} Y\,d\mu_{M^\tau} = \int_{\mathbf{R}_+\times\Omega} \mathbf{1}_{[0,\tau]} Y\,d\mu_M. \tag{5.21}$$

Proof. Consider first a nondecreasing cadlag function G on $[0,\infty)$. For $u > 0$, define the stopped function G^u by $G^u(t) = G(u \wedge t)$. Then the Lebesgue-Stieltjes measures satisfy

$$\int_{(0,\infty)} h\,d\Lambda_{G^u} = \int_{(0,\infty)} \mathbf{1}_{(0,u]} h\,\Lambda_G$$

for every nonnegative Borel function h. This can be justified by the π-λ Theorem. For any interval $(a,b]$,

$$\Lambda_{G^u}(s,t] = G^u(t) - G^u(s) = G(u\wedge t) - G(u\wedge s) = \Lambda_G\big((s,t]\cap(0,u]\big).$$

Then by Lemma B.5 the measures Λ_{G^u} and $\Lambda_G(\,\cdot\,\cap(0,u])$ coincide on all Borel sets of $(0,\infty)$. The equality extends to $[0,\infty)$ if we set $G(0-) = G(0)$ so that the measure of $\{0\}$ is zero under both measures.

Now fix ω and apply the preceding. By Lemma 3.28, $[M^\tau] = [M]^\tau$, and so

$$\begin{aligned}\int_{[0,\infty)} Y(s,\omega)\,d[M^\tau]_s(\omega) &= \int_{[0,\infty)} Y(s,\omega)\,d[M]_s^\tau(\omega)\\ &= \int_{[0,\infty)} \mathbf{1}_{[0,\tau(\omega)]}(s) Y(s,\omega)\,d[M]_s(\omega).\end{aligned}$$

Taking expectation over this equality gives the conclusion. □

The lemma implies that the measure μ_{M^τ} is absolutely continuous with respect to μ_M, and furthermore that $\mathcal{L}_2(M,\mathcal{P}) \subseteq \mathcal{L}_2(M^\tau,\mathcal{P})$.

Proposition 5.16. *Let $M \in \mathcal{M}_2$, $X \in \mathcal{L}_2(M,\mathcal{P})$, and let τ be a stopping time.*

(a) *Let Z be a bounded $\mathcal{F}_\tau$-measurable random variable. Then $Z\mathbf{1}_{(\tau,\infty)}X$ and $\mathbf{1}_{(\tau,\infty)}X$ are both members of $\mathcal{L}_2(M,\mathcal{P})$, and*

$$\int_{(0,t]} Z\mathbf{1}_{(\tau,\infty)} X\,dM = Z\int_{(0,t]} \mathbf{1}_{(\tau,\infty)} X\,dM. \tag{5.22}$$

(b) *The integral behaves as follows under stopping:*

$$\big((\mathbf{1}_{[0,\tau]}X)\cdot M\big)_t = (X\cdot M)_{\tau\wedge t} = (X\cdot M^{\tau})_t. \tag{5.23}$$

(c) *Let also $N \in \mathcal{M}_2$ and $Y \in \mathcal{L}_2(N,\mathcal{P})$. Suppose there is a stopping time σ such that $X_t(\omega) = Y_t(\omega)$ and $M_t(\omega) = N_t(\omega)$ for $0 \le t \le \sigma(\omega)$. Then $(X\cdot M)_{\sigma\wedge t} = (Y\cdot N)_{\sigma\wedge t}$ for all $t \ge 0$.*

Remark 5.17. Equation (5.23) implies that τ can appear in any subset of the three locations. For example,

$$\begin{aligned}(X\cdot M)_{\tau\wedge t} &= (X\cdot M)_{\tau\wedge\tau\wedge t} = (X\cdot M^{\tau})_{\tau\wedge t}\\ &= (X\cdot M^{\tau})_{\tau\wedge\tau\wedge t} = \big((\mathbf{1}_{[0,\tau]}X)\cdot M^{\tau}\big)_{\tau\wedge t}.\end{aligned} \tag{5.24}$$

Proof. (a) $Z\mathbf{1}_{(\tau,\infty)}$ is $\mathcal{P}$-measurable because it is an adapted caglad process. (This process equals Z if $t > \tau$, otherwise it vanishes. If $t > \tau$ then $\mathcal{F}_\tau \subseteq \mathcal{F}_t$ which implies that Z is $\mathcal{F}_t$-measurable.) This takes care of the measurability issue. Multiplying $X \in \mathcal{L}_2(M,\mathcal{P})$ by something bounded and $\mathcal{P}$-measurable creates a process in $\mathcal{L}_2(M,\mathcal{P})$.

Assume first $\tau = u$, a deterministic time. Let X_n as in (5.16) approximate X in $\mathcal{L}_2$. Then

$$\mathbf{1}_{(u,\infty)}X_n = \sum_{i=1}^{k-1} \xi_i \mathbf{1}_{(u\vee t_i, u\vee t_{i+1}]}$$

approximates $\mathbf{1}_{(u,\infty)}X$ in $\mathcal{L}_2$. And

$$Z\mathbf{1}_{(u,\infty)}X_n = \sum_{i=1}^{k-1} Z\xi_i \mathbf{1}_{(u\vee t_i, u\vee t_{i+1}]}$$

are elements of $\mathcal{S}_2$ that approximate $Z\mathbf{1}_{(u,\infty)}X$ in $\mathcal{L}_2$. Their integrals are

$$\begin{aligned}\big((Z\mathbf{1}_{(u,\infty)}X_n)\cdot M\big)_t &= \sum_{i=1}^{k-1} Z\xi_i (M_{(u\vee t_{i+1})\wedge t} - M_{(u\vee t_i)\wedge t})\\ &= Z\big((\mathbf{1}_{(u,\infty)}X_n)\cdot M\big)_t.\end{aligned}$$

Letting $n \to \infty$ along a suitable subsequence gives almost sure convergence of both sides of this equality to the corresponding terms in (5.22) at time t, in the case $\tau = u$.

Now let τ be a general stopping time. Define τ^m by

$$\tau^m = \begin{cases} i2^{-m}, & \text{if } (i-1)2^{-m} \le \tau < i2^{-m} \text{ for some } 1 \le i \le 2^m m\\ \infty, & \text{if } \tau \ge m.\end{cases}$$

Pointwise $\tau^m \searrow \tau$ as $m \nearrow \infty$, and $\mathbf{1}_{(\tau^m,\infty)} \nearrow \mathbf{1}_{(\tau,\infty)}$. Both

$$\mathbf{1}_{\{(i-1)2^{-m}\le\tau<i2^{-m}\}} \quad \text{and} \quad \mathbf{1}_{\{(i-1)2^{-m}\le\tau<i2^{-m}\}}Z$$

are $\mathcal{F}_{i2^{-m}}$-measurable for each i. (The former by definition of a stopping time, the latter by Exercise 2.9.) The first part proved above applies to each such random variable with $u = i2^{-m}$.

$$\begin{aligned}(Z\mathbf{1}_{(\tau^m,\infty)}X)\cdot M &= \Bigl(\sum_{i=1}^{2^m m}\mathbf{1}_{\{(i-1)2^{-m}\le\tau<i2^{-m}\}}Z\mathbf{1}_{(i2^{-m},\infty)}X\Bigr)\cdot M\\ &= Z\sum_{i=1}^{2^m m}\mathbf{1}_{\{(i-1)2^{-m}\le\tau<i2^{-m}\}}\bigl(\mathbf{1}_{(i2^{-m},\infty)}X\bigr)\cdot M\\ &= Z\Bigl(\sum_{i=1}^{2^m m}\mathbf{1}_{\{(i-1)2^{-m}\le\tau<i2^{-m}\}}\mathbf{1}_{(i2^{-m},\infty)}X\Bigr)\cdot M\\ &= Z\bigl((\mathbf{1}_{(\tau^m,\infty)}X)\cdot M\bigr).\end{aligned}$$

Let $m\to\infty$. Because $Z\mathbf{1}_{(\tau^m,\infty)}X\to Z\mathbf{1}_{(\tau,\infty)}X$ and $\mathbf{1}_{(\tau^m,\infty)}X\to\mathbf{1}_{(\tau,\infty)}X$ in $\mathcal{L}_2$, both extreme members of the equalities above converge in $\mathcal{M}_2$ to the corresponding martingales with τ^m replaced by τ. This completes the proof of part (a).

Part (b). We prove the first equality in (5.23). Let $\tau_n = 2^{-n}(\lfloor 2^n\tau\rfloor+1)$ be the usual discrete approximation that converges down to τ as $n\to\infty$. Let $\ell(n) = \lfloor 2^n t\rfloor + 1$. Since $\tau\ge k2^{-n}$ iff $\tau_n\ge(k+1)2^{-n}$,

$$\begin{aligned}(X\cdot M)_{\tau_n\wedge t} &= \sum_{k=0}^{\ell(n)}\mathbf{1}\{\tau\ge k2^{-n}\}\bigl((X\cdot M)_{(k+1)2^{-n}\wedge t}-(X\cdot M)_{k2^{-n}\wedge t}\bigr)\\ &= \sum_{k=0}^{\ell(n)}\mathbf{1}\{\tau\ge k2^{-n}\}\int_{(0,t]}\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}X\,dM\\ &= \sum_{k=0}^{\ell(n)}\int_{(0,t]}\mathbf{1}\{\tau\ge k2^{-n}\}\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}X\,dM\\ &= \int_{(0,t]}\Bigl(\mathbf{1}_{\{0\}}X+\sum_{k=0}^{\ell(n)}\mathbf{1}\{\tau\ge k2^{-n}\}\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}X\Bigr)\,dM\\ &= \int_{(0,t]}\mathbf{1}_{[0,\tau_n]}X\,dM.\end{aligned}$$

In the calculation above, the second equality comes from (5.19), the third from (5.22) where Z is the $\mathcal{F}_{k2^{-n}}$-measurable $\mathbf{1}\{\tau\ge k2^{-n}\}$. The next to last equality uses additivity and adds in the term $\mathbf{1}_{\{0\}}X$ that integrates to zero. The last equality comes from the observation that for $s\in[0,t]$

$$\mathbf{1}_{[0,\tau_n]}(s,\omega) = \mathbf{1}_{\{0\}}(s)+\sum_{k=0}^{\ell(n)}\mathbf{1}_{\{\tau\ge k2^{-n}\}}(\omega)\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(s).$$

Now let $n \to \infty$. By right-continuity, $(X \cdot M)_{\tau_n \wedge t} \to (X \cdot M)_{\tau \wedge t}$. To show that the last term of the string of equalities converges to $\big((\mathbf{1}_{[0,\tau]}X)\cdot M\big)_t$, it suffices to show, by the isometry (5.12), that

$$\lim_{n\to\infty} \int_{[0,t]\times\Omega} \big|\, \mathbf{1}_{[0,\tau_n]}X - \mathbf{1}_{[0,\tau]}X \,\big|^2 \, d\mu_M = 0.$$

This follows from dominated convergence. The integrand vanishes as $n \to \infty$ because

$$\mathbf{1}_{[0,\tau_n]}(t,\omega) - \mathbf{1}_{[0,\tau]}(t,\omega) = \begin{cases} 0, & \tau(\omega) = \infty \\ \mathbf{1}\{\tau(\omega) < t \le \tau_n(\omega)\}, & \tau(\omega) < \infty \end{cases}$$

and $\tau_n(\omega) \searrow \tau(\omega)$. The integrand is bounded by $|X|^2$ for all n, and

$$\int_{[0,t]\times\Omega} |X|^2 \, d\mu_M < \infty$$

by the assumption $X \in \mathcal{L}_2$. This completes the proof of the first equality in (5.23).

We turn to proving the second equality of (5.23). Let $X_n \in \mathcal{S}_2$ as in (5.16) approximate X in $\mathcal{L}_2(M,\mathcal{P})$. By (5.21), $X \in \mathcal{L}_2(M^\tau,\mathcal{P})$ and the processes X_n approximate X also in $\mathcal{L}_2(M^\tau,\mathcal{P})$. Comparing their integrals, we get

$$\begin{aligned} (X_n \cdot M^\tau)_t &= \sum_i \xi_i (M^\tau_{t_{i+1}\wedge t} - M^\tau_{t_i \wedge t}) \\ &= \sum_i \xi_i (M_{t_{i+1}\wedge t \wedge \tau} - M_{t_i \wedge t\wedge\tau}) \\ &= (X_n \cdot M)_{t\wedge\tau} \end{aligned}$$

By the definition of the stochastic integral $X \cdot M^\tau$, the random variables $(X_n \cdot M^\tau)_t$ converge to $(X \cdot M^\tau)_t$ in L^2 as $n \to \infty$.

We cannot appeal to the definition of the integral to assert the convergence of $(X_n \cdot M)_{t\wedge\tau}$ to $(X \cdot M)_{t\wedge\tau}$ because the time point is random. However, martingales afford strong control of their paths. $Y_n(t) = (X_n \cdot M)_t - (X \cdot M)_t$ is an L^2 martingale with $Y_n(0) = 0$. Lemma 3.5 applied to the submartingale $Y_n^2(t)$ implies

$$E[Y_n^2(t\wedge\tau)] \le 2E[Y_n^2(t)] = 2E\Big[\big((X_n \cdot M)_t - (X \cdot M)_t\big)^2\Big]$$

and this last expectation vanishes as $n \to \infty$ by the definition of $X \cdot M$. Consequently

$$(X_n \cdot M)_{t\wedge\tau} \to (X \cdot M)_{t\wedge\tau} \quad \text{in } L^2.$$

This completes the proof of the second equality in (5.23).

Part (c). Since $\mathbf{1}_{[0,\sigma]}X = \mathbf{1}_{[0,\sigma]}Y$ and $M^\sigma = N^\sigma$,

$$(X \cdot M)_{t\wedge\sigma} = \big((\mathbf{1}_{[0,\sigma]}X) \cdot M^\sigma\big)_t = \big((\mathbf{1}_{[0,\sigma]}Y) \cdot N^\sigma\big)_t = (Y \cdot N)_{t\wedge\sigma}. \qquad \square$$

Example 5.18. Let us record some simple integrals as consequences of the properties.

(a) Let $\sigma \le \tau$ be two stopping times, and ξ a bounded $\mathcal{F}_\sigma$-measurable random variable. Define $X = \xi\mathbf{1}_{(\sigma,\tau]}$, or more explicitly,

$$X_t(\omega) = \xi(\omega)\mathbf{1}_{(\sigma(\omega),\tau(\omega)]}(t).$$

As an adapted caglad process, X is predictable. Let M be an L^2-martingale. Pick a constant $C \ge |\xi(\omega)|$. Then for any $T < \infty$,

$$\begin{aligned}\int_{[0,T]\times\Omega} X^2\, d\mu_M &= E\big\{\xi^2\big([M]_{\tau\wedge T} - [M]_{\sigma\wedge T}\big)\big\} \le C^2 E\big\{[M]_{\tau\wedge T}\big\}\\ &= C^2 E\{M^2_{\tau\wedge T}\} \le C^2 E\{M_T^2\} < \infty.\end{aligned}$$

Thus $X \in \mathcal{L}_2(M, \mathcal{P})$. By (5.22) and (5.23),

$$\begin{aligned}X \cdot M &= (\xi\mathbf{1}_{(\sigma,\infty)}\mathbf{1}_{[0,\tau]}) \cdot M = \xi\big((\mathbf{1}_{(\sigma,\infty)}\mathbf{1}_{[0,\tau]}) \cdot M\big) = \xi\big((\mathbf{1}_{[0,\tau]} - \mathbf{1}_{[0,\sigma]}) \cdot M\big)\\ &= \xi\big((\mathbf{1}\cdot M)^\tau - (\mathbf{1}\cdot M)^\sigma\big) = \xi(M^\tau - M^\sigma).\end{aligned}$$

Above we used $\mathbf{1}$ to denote the function or process that is identically one.

(b) Continuing the example, consider a sequence $0 \le \sigma_1 \le \sigma_2 \le \cdots \le \sigma_i \nearrow \infty$ of stopping times, and random variables $\{\eta_i : i \ge 1\}$ such that η_i is $\mathcal{F}_{\sigma_i}$-measurable and $C = \sup_{i,\omega}|\eta_i(\omega)| < \infty$. Let

$$X(t) = \sum_{i=1}^\infty \eta_i \mathbf{1}_{(\sigma_i,\sigma_{i+1}]}(t).$$

As a bounded caglad process, $X \in \mathcal{L}_2(M, \mathcal{P})$ for any L^2-martingale M. Let

$$X_n(t) = \sum_{i=1}^n \eta_i \mathbf{1}_{(\sigma_i,\sigma_{i+1}]}(t).$$

By part (a) of the example and the additivity of the integral,

$$X_n \cdot M = \sum_{i=1}^n \eta_i (M^{\sigma_{i+1}} - M^{\sigma_i}).$$

$X_n \to X$ pointwise. And since $|X - X_n| \le 2C$,

$$\int_{[0,T]\times\Omega} |X - X_n|^2\, d\mu_M \longrightarrow 0$$

for any $T < \infty$ by dominated convergence. Consequently $X_n \to X$ in $\mathcal{L}_2(M, \mathcal{P})$, and then by the isometry, $X_n \cdot M \to X \cdot M$ in $\mathcal{M}_2$. From the formula for $X_n \cdot M$ it is clear where it converges pointwise, and this limit must agree with the $\mathcal{M}_2$ limit. The conclusion is

$$X \cdot M = \sum_{i=1}^{\infty} \eta_i (M^{\sigma_{i+1}} - M^{\sigma_i}).$$

As the last issue of this section, we consider integrating a given process X with respect to more than one martingale.

Proposition 5.19. *Let $M, N \in \mathcal{M}_2$, $\alpha, \beta \in \mathbf{R}$, and $X \in \mathcal{L}_2(M, \mathcal{P}) \cap \mathcal{L}_2(N, \mathcal{P})$. Then $X \in \mathcal{L}_2(\alpha M + \beta N, \mathcal{P})$, and*

$$X \cdot (\alpha M + \beta N) = \alpha(X \cdot M) + \beta(X \cdot N). \tag{5.25}$$

Lemma 5.20. *For a predictable process Y,*

$$\begin{aligned} &\left\{ \int_{[0,T]\times\Omega} |Y|^2 \, d\mu_{\alpha M + \beta N} \right\}^{1/2} \\ &\qquad \le |\alpha| \left\{ \int_{[0,T]\times\Omega} |Y|^2 \, d\mu_M \right\}^{1/2} + |\beta| \left\{ \int_{[0,T]\times\Omega} |Y|^2 \, d\mu_N \right\}^{1/2}. \end{aligned}$$

Proof. The linearity

$$[\alpha M + \beta N] = \alpha^2 [M] + 2\alpha\beta[M, N] + \beta^2 [N]$$

is inherited by the Lebesgue-Stieltjes measures. By the Kunita-Watanabe inequality (2.22),

$$\begin{aligned} \int_{[0,T]} |Y_s|^2 \, d[\alpha M + \beta N]_s &= \alpha^2 \int_{[0,T]} |Y_s|^2 \, d[M]_s + 2\alpha\beta \int_{[0,T]} |Y_s|^2 \, d[M, N]_s \\ &\qquad + \beta^2 \int_{[0,T]} |Y_s|^2 \, d[N]_s \\ &\le \alpha^2 \int_{[0,T]} |Y_s|^2 \, d[M]_s + 2|\alpha||\beta| \left\{ \int_{[0,T]} |Y_s|^2 \, d[M]_s \right\}^{1/2} \left\{ \int_{[0,T]} |Y_s|^2 \, d[N]_s \right\}^{1/2} \\ &\qquad + \beta^2 \int_{[0,T]} |Y_s|^2 \, d[N]_s. \end{aligned}$$

The above integrals are Lebesgue-Stieltjes integrals over $[0, T]$, evaluated at a fixed ω. Take expectations and apply Schwarz inequality to the middle term. □

Proof of Proposition 5.19. The Lemma shows that $X \in \mathcal{L}_2(\alpha M + \beta N, \mathcal{P})$. Replace the measure μ_M in the proof of Lemma 5.10 with the measure

$\widetilde{\mu} = \mu_M + \mu_N$. The proof works exactly as before, and gives a sequence of simple predictable processes X_n such that

$$\int_{[0,T]\times\Omega} |X - X_n|^2 \, d(\mu_M + \mu_N) \to 0$$

for each $T < \infty$. This combined with the previous lemma says that $X_n \to X$ simultaneously in spaces $\mathcal{L}_2(M, \mathcal{P})$, $\mathcal{L}_2(N, \mathcal{P})$, and $\mathcal{L}_2(\alpha M + \beta N, \mathcal{P})$. (5.25) holds for X_n by the explicit formula for the integral of a simple predictable process, and the general conclusion follows by taking the limit. □

5.2. Local square-integrable martingale integrator

Recall that a cadlag process M is a local L^2-martingale if there exists a nondecreasing sequence of stopping times $\{\sigma_k\}$ such that $\sigma_k \nearrow \infty$ almost surely, and for each k the stopped process $M^{\sigma_k} = \{M_{\sigma_k \wedge t} : t \in \mathbf{R}_+\}$ is an L^2-martingale. The sequence $\{\sigma_k\}$ is a localizing sequence for M. $\mathcal{M}_{2,\text{loc}}$ is the space of cadlag, local L^2-martingales.

We wish to define a stochastic integral $X \cdot M$ where M can be a local L^2-martingale. The earlier approach via an L^2 isometry will not do because the whole point is to get rid of integrability assumptions. We start by defining the class of integrands. Even for an L^2-martingale this gives us integrands beyond the $\mathcal{L}_2$-space of the previous section.

Definition 5.21. Given a local square-integrable martingale M, let $\mathcal{L}(M, \mathcal{P})$ denote the class of predictable processes X which have the following property: there exists a sequence of stopping times $0 \le \tau_1 \le \tau_2 \le \tau_3 \le \cdots \le \tau_k \le \cdots$ such that

(i) $P\{\tau_k \nearrow \infty\} = 1$,

(ii) M^{τ_k} is a square-integrable martingale for each k, and

(iii) the process $\mathbf{1}_{[0,\tau_k]}X$ lies in $\mathcal{L}_2(M^{\tau_k}, \mathcal{P})$ for each k.

Let us call such a sequence of stopping times a *localizing sequence* for the pair (X, M).

By our earlier development, for each k the stochastic integral

$$Y^k = (\mathbf{1}_{[0,\tau_k]}X) \cdot M^{\tau_k}$$

exists as an element of $\mathcal{M}_2$. The idea will be now to exploit the consistency in the sequence of stochastic integrals Y^k, which enables us to define $(X \cdot M)_t(\omega)$ for a fixed (t, ω) by the recipe "take $Y^k_t(\omega)$ for a large enough k." First a lemma that justifies the approach.

Lemma 5.22. *Let $M \in \mathcal{M}_{2,\text{loc}}$ and let X be a predictable process. Suppose σ and τ are two stopping times such that M^σ and M^τ are cadlag L^2-martingales, $\mathbf{1}_{[0,\sigma]}X \in \mathcal{L}_2(M^\sigma, \mathcal{P})$ and $\mathbf{1}_{[0,\tau]}X \in \mathcal{L}_2(M^\tau, \mathcal{P})$. Let*

$$Z_t = \int_{(0,t]} \mathbf{1}_{[0,\sigma]}X \, dM^\sigma \quad \text{and} \quad W_t = \int_{(0,t]} \mathbf{1}_{[0,\tau]}X \, dM^\tau$$

denote the stochastic integrals, which are cadlag L^2-martingales. Then

$$Z_{t\wedge\sigma\wedge\tau} = W_{t\wedge\sigma\wedge\tau}$$

where we mean that the two processes are indistinguishable.

Proof. A short derivation based on (5.23)–(5.24) and some simple observations: $(M^\sigma)^\tau = (M^\tau)^\sigma = M^{\sigma\wedge\tau}$, and $\mathbf{1}_{[0,\sigma]}X$, $\mathbf{1}_{[0,\tau]}X$ both lie in $\mathcal{L}_2(M^{\sigma\wedge\tau}, \mathcal{P})$.

$$\begin{aligned} Z_{t\wedge\sigma\wedge\tau} &= \big((\mathbf{1}_{[0,\sigma]}X)\cdot M^\sigma\big)_{t\wedge\sigma\wedge\tau} = \big((\mathbf{1}_{[0,\tau]}\mathbf{1}_{[0,\sigma]}X)\cdot (M^\sigma)^\tau\big)_{t\wedge\sigma\wedge\tau} \\ &= \big((\mathbf{1}_{[0,\sigma]}\mathbf{1}_{[0,\tau]}X)\cdot (M^\tau)^\sigma\big)_{t\wedge\sigma\wedge\tau} = \big((\mathbf{1}_{[0,\tau]}X)\cdot M^\tau\big)_{t\wedge\sigma\wedge\tau} \\ &= W_{t\wedge\sigma\wedge\tau}. \end{aligned}$$

□

Let Ω_0 be the following event:

$$\begin{aligned} \Omega_0 = \{\omega : \tau_k(\omega) \nearrow \infty \text{ as } k \nearrow \infty, \text{ and for all } (k,m), \\ Y^k_{t\wedge\tau_k\wedge\tau_m}(\omega) = Y^m_{t\wedge\tau_k\wedge\tau_m}(\omega) \text{ for all } t \in \mathbf{R}_+.\} \end{aligned} \tag{5.26}$$

$P(\Omega_0) = 1$ by the assumption $P\{\tau_k \nearrow \infty\} = 1$, by the previous lemma, and because there are countably many pairs (k,m). To rephrase this, on the event Ω_0, if k and m are indices such that $t \le \tau_k \wedge \tau_m$, then $Y^k_t = Y^m_t$. This makes the definition below sensible.

Definition 5.23. Let $M \in \mathcal{M}_{2,\text{loc}}$, $X \in \mathcal{L}(M, \mathcal{P})$, and let $\{\tau_k\}$ be a localizing sequence for (X, M). Define the event Ω_0 as in the previous paragraph. The *stochastic integral* $X \cdot M$ is the cadlag local L^2-martingale defined as follows: on the event Ω_0 set

$$\begin{aligned} &(X\cdot M)_t(\omega) = \big((\mathbf{1}_{[0,\tau_k]}X)\cdot M^{\tau_k}\big)_t(\omega) \\ &\text{for any } k \text{ such that } \tau_k(\omega) \ge t. \end{aligned} \tag{5.27}$$

Outside the event Ω_0 set $(X \cdot M)_t = 0$ for all t.

This definition is independent of the localizing sequence $\{\tau_k\}$ in the sense that using any other localizing sequence of stopping times gives a process indistinguishable from $X \cdot M$ defined above.

Justification of the definition. The process $X\cdot M$ is cadlag on any bounded interval $[0,T]$ for the following reasons. If $\omega \notin \Omega_0$ the process is constant in time. If $\omega \in \Omega_0$, pick k large enough so that $\tau_k(\omega) > T$, and note that the path $t \mapsto (X \cdot M)_t(\omega)$ coincides with the cadlag path $t \mapsto Y^k_t(\omega)$ on the interval $[0,T]$. Being cadlag on all bounded intervals is the same as being cadlag on $\mathbf{R}_+$, so it follows that $X \cdot M$ is cadlag process.

The stopped process satisfies

$$(X\cdot M)^{\tau_k}_t = (X\cdot M)_{\tau_k\wedge t} = Y^k_{\tau_k\wedge t} = (Y^k)^{\tau_k}_t$$

because by definition $X \cdot M = Y^k$ (almost surely) on $[0, \tau_k]$. Y^k is an L^2-martingale, hence so is $(Y^k)^{\tau_k}$. Consequently $(X \cdot M)^{\tau_k}$ is a cadlag L^2-martingale. This shows that $X \cdot M$ is a cadlag local L^2-martingale.

To take up the last issue, let $\{\sigma_j\}$ be another localizing sequence of stopping times for (X, M). Let

$$W_t^j = \int_{(0,t]} \mathbf{1}_{[0,\sigma_j]} X \, dM^{\sigma_j}.$$

Corresponding to the event Ω_0 and definition (5.27) from above, based on $\{\sigma_j\}$ we define an event Ω_1 with $P(\Omega_1) = 1$, and on Ω_1 an "alternative" stochastic integral by

$$W_t(\omega) = W_t^j(\omega) \quad \text{for } j \text{ such that } \sigma_j(\omega) \geq t. \tag{5.28}$$

Lemma 5.22 implies that the processes $W^j_{t\wedge\sigma_j\wedge\tau_k}$ and $Y^k_{t\wedge\sigma_j\wedge\tau_k}$ are indistinguishable. Let Ω_2 be the set of $\omega \in \Omega_0 \cap \Omega_1$ for which

$$W^j_{t\wedge\sigma_j\wedge\tau_k}(\omega) = Y^k_{t\wedge\sigma_j\wedge\tau_k}(\omega) \quad \text{for all } t \in \mathbf{R}_+ \text{ and all pairs } (j,k).$$

$P(\Omega_2) = 1$ because it is an intersection of countably many events of probability one. We claim that for $\omega \in \Omega_2$, $(X \cdot M)_t(\omega) = W_t(\omega)$ for all $t \in \mathbf{R}_+$. Given t, pick j and k so that $\sigma_j(\omega) \wedge \tau_k(\omega) \geq t$. Then, using (5.27), (5.28) and $\omega \in \Omega_2$,

$$(X \cdot M)_t(\omega) = Y_t^k(\omega) = W_t^j(\omega) = W_t(\omega).$$

We have shown that $X \cdot M$ and W are indistinguishable, so the definition of $X \cdot M$ does not depend on the particular localizing sequence used.

We have justified all the claims made in the definition. □

Remark 5.24 (Irrelevance of the time origin)**.** The value X_0 does not affect anything above because $\mu_Z(\{0\} \times \Omega) = 0$ for any L^2-martingale Z. If a predictable X is given and $\widetilde{X}_t = \mathbf{1}_{(0,\infty)}(t) X_t$, then $\mu_Z\{X \neq \widetilde{X}\} = 0$. In particular, $\{\tau_k\}$ is a localizing sequence for (X, M) iff it is a localizing sequence for $(\widetilde{X}, M)$, and $X \cdot M = \widetilde{X} \cdot M$ if a localizing sequence exists. Also, in part (iii) of Definition 5.21 we can equivalently require that $\mathbf{1}_{(0,\tau_k]} X$ lies in $\mathcal{L}_2(M^{\tau_k}, \mathcal{P})$.

Remark 5.25 (Path continuity)**.** If the local L^2-martingale M has continuous paths to begin with, then so do M^{τ_k}, hence also the integrals $\mathbf{1}_{[0,\tau_k]} M^{\tau_k}$ have continuous paths, and the integral $X \cdot M$ has continuous paths.

Example 5.26. Compared with Example 5.4, with Brownian motion and the compensated Poisson process we can now integrate predictable processes X that satisfy

$$\int_0^T X(s,\omega)^2 \, ds < \infty \quad \text{for all } T < \infty, \text{ for } P\text{-almost every } \omega.$$

(Exercise 5.7 asks you to verify this.) Again, we know from Chapter 4 that predictability is not really needed for integrands when the integrator is Brownian motion.

Property (iii) of Definition 5.21 made the localization argument of the definition of $X \cdot M$ work. In important special cases property (iii) follows from this stronger property:

$$\text{there exist stopping times } \{\sigma_k\} \text{ such that } \sigma_k \nearrow \infty \text{ almost surely and } \mathbf{1}_{(0,\sigma_k]}X \text{ is a bounded process for each } k. \tag{5.29}$$

Let M be an arbitrary local L^2-martingale with localizing sequence $\{\nu_k\}$, and assume X is a predictable process that satisfies (5.29). A bounded process is in $\mathcal{L}_2(Z,\mathcal{P})$ for any L^2-martingale Z, and consequently $\mathbf{1}_{(0,\sigma_k]}X \in \mathcal{L}_2(M^{\nu_k},\mathcal{P})$. By Remark 5.24 the conclusion extends to $\mathbf{1}_{[0,\sigma_k]}X$. Thus the stopping times $\tau_k = \sigma_k \wedge \nu_k$ localize the pair (X, M), and the integral $X \cdot M$ is well-defined.

The time origin is left out of $\mathbf{1}_{(0,\sigma_k]}X$ because $\mathbf{1}_{[0,\sigma_k]}X$ cannot be bounded unless X_0 is bounded. This would be unnecessarily restrictive.

The next proposition lists the most obvious types of predictable processes that satisfy (5.29). In certain cases demonstrating the existence of the stopping times may require a right-continuous filtration. Then one replaces $\{\mathcal{F}_t\}$ with $\{\mathcal{F}_{t+}\}$. As observed in the beginning of Chapter 3, this can be done without losing any cadlag martingales (or local martingales).

Recall also the definition

$$X_T^*(\omega) = \sup_{0\le t\le T} |X_t(\omega)| \tag{5.30}$$

which is $\mathcal{F}_T$-measurable for any left- or right-continuous process X, provided we make the filtration complete. (See discussion after (3.10) in Chapter 3.)

Proposition 5.27. *The following cases are examples of processes with stopping times $\{\sigma_k\}$ that satisfy condition (5.29).*

(i) *X is predictable, and for each $T < \infty$ there exists a constant $C_T < \infty$ such that, with probability one, $X_t \le C_T$ for all $0 < t \le T$. Take $\sigma_k = k$.*

(ii) *X is adapted and has almost surely continuous paths. Take*

$$\sigma_k = \inf\{t \ge 0 : |X_t| \ge k\}.$$

(iii) *X is adapted, and there exists an adapted, cadlag process Y such that $X(t) = Y(t-)$ for $t > 0$. Take*

$$\sigma_k = \inf\{t > 0 : |Y(t)| \ge k \text{ or } |Y(t-)| \ge k\}.$$

(iv) *X is adapted, has almost surely left-continuous paths, and $X_T^* < \infty$ almost surely for each $T < \infty$. Assume the underlying filtration $\{\mathcal{F}_t\}$ right-continuous. Take*

$$\sigma_k = \inf\{t \ge 0 : |X_t| > k\}.$$

Remark 5.28. Category (ii) is a special case of (iii), and category (iii) is a special case of (iv). Category (iii) seems artificial but will be useful. Notice that *every* caglad X satisfies $X(t) = Y(t-)$ for the cadlag process Y defined by $Y(t) = X(t+)$, but this Y may fail to be adapted. Y is adapted if $\{\mathcal{F}_t\}$ is right-continuous. But then we find ourselves in Category (iv).

Let us state the most important special case of continuous processes as a corollary in its own right. It follows from Lemma 3.20 and from case (ii) above.

Corollary 5.29. *For any continuous local martingale M and continuous, adapted process X, the stochastic integral $X \cdot M$ is well-defined.*

Proof of Proposition 5.27. Case (i): nothing to prove.

Case (ii). By Lemma 2.10 this σ_k is a stopping time. A continuous path $t \mapsto X_t(\omega)$ is bounded on compact time intervals. Hence for almost every ω, $\sigma_k(\omega) \nearrow \infty$. Again by continuity, $|X_s| \le k$ for $0 < s \le \sigma_k$. Note that if $|X_0| > k$ then $\sigma_k = 0$, so we cannot claim $\mathbf{1}_{[0,\sigma_k]}|X_0| \le k$. This is why boundedness cannot be required at time zero.

Case (iii). By Lemma 2.9 this σ_k is a stopping time. A cadlag path is locally bounded just like a continuous path (Exercise A.1), and so $\sigma_k \nearrow \infty$. If $\sigma_k > 0$, then $|X(t)| < k$ for $t < \sigma_k$, and by left-continuity $|X(\sigma_k)| \le k$. Note that $|Y(\sigma_k)| \le k$ may fail so we cannot adapt this argument to Y.

Case (iv). By Lemma 2.7 σ_k is a stopping time since we assume $\{\mathcal{F}_t\}$ right-continuous. As in case (iii), by left-continuity $|X_s| \le k$ for $0 < s \le \sigma_k$. Given ω such that $X_T^*(\omega) < \infty$ for all $T < \infty$, we can choose $k_T > \sup_{0 \le t \le T}|X_t(\omega)|$ and then $\sigma_k(\omega) \ge T$ for $k \ge k_T$. Thus $\sigma_k \nearrow \infty$ almost surely. □

Example 5.30. Let us repeat Example 5.18 without boundedness assumptions. $0 \le \sigma_1 \le \sigma_2 \le \cdots \le \sigma_i \nearrow \infty$ are stopping times, η_i is a finite $\mathcal{F}_{\sigma_i}$-measurable random variable for $i \ge 1$, and

$$X_t = \sum_{i=1}^{\infty} \eta_i \mathbf{1}_{(\sigma_i, \sigma_{i+1}]}(t).$$

X is a caglad process, and satisfies the hypotheses of case (iii) of Proposition 5.27. We shall define a concrete localizing sequence. Fix $M \in \mathcal{M}_{2,\text{loc}}$ and let $\{\rho_k\}$ be a localizing sequence for M. Define

$$\zeta_k = \begin{cases} \sigma_j, & \text{if } \max_{1 \le i \le j-1}|\eta_i| \le k < |\eta_j| \text{ for some } j \\ \infty, & \text{if } |\eta_i| \le k \text{ for all } i. \end{cases}$$

That ζ_k is a stopping time follows directly from

$$\{\zeta_k \le t\} = \bigcup_{j=1}^{\infty} \Big(\big\{ \max_{1\le i\le j-1} |\eta_i| \le k < |\eta_j| \big\} \cap \big\{ \sigma_j \le t \big\} \Big).$$

Also $\zeta_k \nearrow \infty$ since $\sigma_i \nearrow \infty$. The stopping times $\tau_k = \rho_k \wedge \zeta_k$ localize the pair (X, M).

Truncate $\eta_i^{(k)} = (\eta_i \wedge k) \vee (-k)$. If $t \in (\sigma_i \wedge \tau_k, \sigma_{i+1} \wedge \tau_k]$ then necessarily $\sigma_i < \tau_k$. This implies $\zeta_k \ge \sigma_{i+1}$ which happens iff $\eta_\ell = \eta_\ell^{(k)}$ for $1 \le \ell \le i$. Hence

$$\mathbf{1}_{[0,\tau_k]}(t) X_t = \sum_{i=1}^{\infty} \eta_i^{(k)} \mathbf{1}_{(\sigma_i\wedge\tau_k, \sigma_{i+1}\wedge\tau_k]}(t).$$

This process is bounded, so by Example 5.18,

$$\big((\mathbf{1}_{[0,\tau_k]} X) \cdot M^{\tau_k} \big)_t = \sum_{i=1}^{\infty} \eta_i^{(k)} (M_{\sigma_{i+1}\wedge\tau_k\wedge t} - M_{\sigma_i\wedge\tau_k\wedge t}).$$

Taking k so that $\tau_k \ge t$, we get

$$(X \cdot M)_t = \sum_{i=1}^{\infty} \eta_i (M_{\sigma_{i+1}\wedge t} - M_{\sigma_i\wedge t}).$$

We use the integral notation

$$\int_{(s,t]} X \, dM = (X \cdot M)_t - (X \cdot M)_s$$

and other notational conventions exactly as for the L^2 integral. The stochastic integral with respect to a local martingale inherits the path properties of the L^2 integral, as we observe in the next proposition. Expectations and conditional expectations of $(X \cdot M)_t$ do not necessarily exist any more so we cannot even contemplate their properties.

Proposition 5.31. *Let $M, N \in \mathcal{M}_{2,\mathrm{loc}}$, $X \in \mathcal{L}(M, \mathcal{P})$, and let τ be a stopping time.*

(a) *Linearity continues to hold: if also $Y \in \mathcal{L}(M, \mathcal{P})$, then*

$$(\alpha X + \beta Y) \cdot M = \alpha (X \cdot M) + \beta (Y \cdot M).$$

(b) *Let Z be a bounded $\mathcal{F}_\tau$-measurable random variable. Then $Z\mathbf{1}_{(\tau,\infty)}X$ and $\mathbf{1}_{(\tau,\infty)}X$ are both members of $\mathcal{L}(M, \mathcal{P})$, and*

$$\int_{(0,t]} Z \mathbf{1}_{(\tau,\infty)} X \, dM = Z \int_{(0,t]} \mathbf{1}_{(\tau,\infty)} X \, dM. \tag{5.31}$$

Furthermore,

$$\big((\mathbf{1}_{[0,\tau]} X) \cdot M \big)_t = (X \cdot M)_{\tau\wedge t} = (X \cdot M^\tau)_t. \tag{5.32}$$

(c) *Let* $Y \in \mathcal{L}(N, \mathcal{P})$. *Suppose* $X_t(\omega) = Y_t(\omega)$ *and* $M_t(\omega) = N_t(\omega)$ *for* $0 \le t \le \tau(\omega)$. *Then*

$$(X \cdot M)_{\tau \wedge t} = (Y \cdot N)_{\tau \wedge t}.$$

(d) *Suppose* $X \in \mathcal{L}(M, \mathcal{P}) \cap \mathcal{L}(N, \mathcal{P})$. *Then for* $\alpha, \beta \in \mathbf{R}$, $X \in \mathcal{L}(\alpha M + \beta N, \mathcal{P})$ *and*

$$X \cdot (\alpha M + \beta N) = \alpha (X \cdot M) + \beta (X \cdot N).$$

Proof. The proofs are short exercises in localization. We show the way by doing (5.31) and the first equality in (5.32).

Let $\{\sigma_k\}$ be a localizing sequence for the pair (X, M). Then $\{\sigma_k\}$ is a localizing sequence also for the pairs $(\mathbf{1}_{(\sigma,\infty)} X, M)$ and $(Z\mathbf{1}_{(\sigma,\infty)} X, M)$. Given ω and t, pick k large enough so that $\sigma_k(\omega) \ge t$. Then by the definition of the stochastic integrals for localized processes,

$$Z\big((\mathbf{1}_{(\tau,\infty)} X) \cdot M)_t(\omega) = Z\big((\mathbf{1}_{[0,\sigma_k]} \mathbf{1}_{(\tau,\infty)} X) \cdot M^{\sigma_k})_t(\omega)$$

and

$$\big((Z\mathbf{1}_{(\tau,\infty)} X) \cdot M)_t(\omega) = \big((\mathbf{1}_{[0,\sigma_k]} Z\mathbf{1}_{(\tau,\infty)} X) \cdot M^{\sigma_k})_t(\omega).$$

The right-hand sides of the two equalities above coincide, by an application of (5.22) to the L^2-martingale M^{σ_k} and the process $\mathbf{1}_{[0,\sigma_k]} X$ in place of X. This verifies (5.31).

The sequence $\{\sigma_k\}$ works also for $(\mathbf{1}_{[0,\tau]} X, M)$. If $t \le \sigma_k(\omega)$, then

$$\begin{aligned}\big((\mathbf{1}_{[0,\tau]} X) \cdot M\big)_t &= \big((\mathbf{1}_{[0,\sigma_k]} \mathbf{1}_{[0,\tau]} X) \cdot M^{\sigma_k}\big)_t = \big((\mathbf{1}_{[0,\sigma_k]} X) \cdot M^{\sigma_k}\big)_{\tau \wedge t} \\ &= (X \cdot M)_{\tau \wedge t}.\end{aligned}$$

The first and the last equality are the definition of the local integral, the middle equality an application of (5.23). This checks the first equality in (5.32). □

We come to a very helpful result for later development. The most important processes are usually either caglad or cadlag. The next proposition shows that for left-continuous processes the integral can be realized as a limit of Riemann sum-type approximations. For future benefit we include random partitions in the result.

However, a cadlag process X is not necessarily predictable and therefore not an admissible integrand. The Poisson process is a perfect example, see Exercise 5.4. It is intuitively natural that the Poisson process cannot be predictable, for how can we predict when the process jumps? But it turns out that the Riemann sums still converge for a cadlag integrand. They just cannot converge to $X \cdot M$ because this integral might not exist. Instead, these sums converge to the integral $X_- \cdot M$ of the caglad process X_- defined by

$$X_-(0) = X(0) \quad \text{and} \quad X_-(t) = X(t-) \quad \text{for } t > 0.$$

We leave it to the reader to verify that X_- has caglad paths (Exercise 5.9).

The limit in the next proposition is not a mere curiosity. It will be important when we derive Itô's formula. Note the similarity with Lemma 1.12 for Lebesgue-Stieltjes integrals.

Proposition 5.32. *Let X be an adapted process and $M \in \mathcal{M}_{2,\text{loc}}$. Suppose $0 = \tau_0^n \le \tau_1^n \le \tau_2^n \le \tau_3^n \le \cdots$ are stopping times such that for each n, $\tau_i^n \to \infty$ almost surely as $i \to \infty$, and $\delta_n = \sup_i(\tau_{i+1}^n - \tau_i^n)$ tends to zero almost surely as $n \to \infty$. Define the process*

$$R_n(t) = \sum_{i=0}^{\infty} X(\tau_i^n)\big(M(\tau_{i+1}^n \wedge t) - M(\tau_i^n \wedge t)\big). \tag{5.33}$$

(a) *Assume X is left-continuous and satisfies (5.29). Then for each fixed $T < \infty$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\Big\{ \sup_{0\le t\le T} |R_n(t) - (X \cdot M)_t| \ge \varepsilon \Big\} = 0.$$

In other words, R_n converges to $X \cdot M$ in probability, uniformly on compact time intervals.

(b) *If X is a cadlag process, then R_n converges to $X_- \cdot M$ in probability, uniformly on compact time intervals.*

Proof. Since $X_- = X$ for a left-continuous process, we can prove parts (a) and (b) simultaneously.

Assume first $X_0 = 0$. This is convenient for the proof. At the end we lift this assumption. By left- or right-continuity X is progressively measurable (Lemma 2.4) and therefore $X(\tau_i^n)$ is $\mathcal{F}_{\tau_i^n}$-measurable on the event $\tau_i^n < \infty$ (Lemma 2.3). Define

$$Y_n(t) = \sum_{i=0}^{\infty} X(\tau_i^n)\mathbf{1}_{(\tau_i^n, \tau_{i+1}^n]}(t) - X_-(t).$$

By the hypotheses and by Example 5.30, Y_n is an element of $\mathcal{L}(M, \mathcal{P})$ and its integral is

$$Y_n \cdot M = R_n - X_- \cdot M.$$

Consequently we need to show that $Y_n \cdot M \to 0$ in probability, uniformly on compacts.

Let $\{\sigma_k\}$ be a localizing sequence for (X_-, M) such that $\mathbf{1}_{(0,\sigma_k]}X_-$ is bounded. In part (a) existence of $\{\sigma_k\}$ is a hypothesis. For part (b) apply part (iii) of Proposition 5.27. As explained there, it may happen that $X(\sigma_k)$ is not bounded, but $X(t-)$ will be bounded for $0 < t \le \sigma_k$.

Pick constants b_k such that $|X_-(t)| \le b_k$ for $0 \le t \le \sigma_k$ (here we rely on the assumption $X_0 = 0$). Define $X^{(k)} = (X \wedge b_k) \vee (-b_k)$ and

$$Y_n^{(k)}(t) = \sum_{i=0}^{\infty} X^{(k)}(\tau_i^n)\mathbf{1}_{(\tau_i^n,\tau_{i+1}^n]}(t) - X_-^{(k)}(t).$$

In forming $X_-^{(k)}(t)$, it is immaterial whether truncation follows the left limit or vice versa.

We have the equality

$$\mathbf{1}_{[0,\sigma_k]}Y_n(t) = \mathbf{1}_{[0,\sigma_k]}Y_n^{(k)}(t).$$

For the sum in Y_n this can be seen term by term:

$$\mathbf{1}_{[0,\sigma_k]}(t)\mathbf{1}_{(\tau_i^n,\tau_{i+1}^n]}(t)X(\tau_i^n) = \mathbf{1}_{[0,\sigma_k]}(t)\mathbf{1}_{(\tau_i^n,\tau_{i+1}^n]}(t)X^{(k)}(\tau_i^n)$$

because both sides vanish unless $\tau_i^n < t \le \sigma_k$ and $|X(s)| \le b_k$ for $0 \le s < \sigma_k$.

Thus $\{\sigma_k\}$ is a localizing sequence for (Y_n, M). On the event $\{\sigma_k > T\}$, for $0 \le t \le T$, by definition (5.27) and Proposition 5.16(b)–(c),

$$(Y_n \cdot M)_t = \big((\mathbf{1}_{[0,\sigma_k]}Y_n) \cdot M^{\sigma_k}\big)_t = (Y_n^{(k)} \cdot M^{\sigma_k})_t.$$

Fix $\varepsilon > 0$. In the next bound we apply martingale inequality (3.8) and the isometry (5.12).

$$\begin{aligned}
P\Big\{\sup_{0\le t\le T}|(Y_n \cdot M)_t| \ge \varepsilon\Big\} &\le P\{\sigma_k \le T\} \\
&\quad + P\Big\{\sup_{0\le t\le T}\big|\big(Y_n^{(k)} \cdot M^{\sigma_k}\big)_t\big| \ge \varepsilon\Big\} \\
&\le P\{\sigma_k \le T\} + \varepsilon^{-2}E\big[(Y_n^{(k)} \cdot M^{\sigma_k})_T^2\big] \\
&\le P\{\sigma_k \le T\} + \varepsilon^{-2}\int_{[0,T]\times\Omega}|Y_n^{(k)}(t,\omega)|^2\,\mu_{M^{\sigma_k}}(dt,d\omega).
\end{aligned}$$

Let $\varepsilon_1 > 0$. Fix k large enough so that $P\{\sigma_k \le T\} < \varepsilon_1$. As $n \to \infty$, $Y_n^{(k)}(t,\omega) \to 0$ for all t, if ω is such that the path $s \mapsto X_-(s,\omega)$ is left-continuous and the assumption $\delta_n(\omega) \to 0$ holds. This excludes at most a zero probability set of ω's, and so this convergence happens $\mu_{M^{\sigma_k}}$-almost everywhere. By the bound $|Y_n^{(k)}| \le 2b_k$ and dominated convergence,

$$\int_{[0,T]\times\Omega}|Y_n^{(k)}|^2\,d\mu_{M^{\sigma_k}} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Letting $n \to \infty$ in the last string of inequalities gives

$$\varlimsup_{n\to\infty} P\Big\{\sup_{0\le t\le T}|(Y_n \cdot M)_t| \ge \varepsilon\Big\} \le \varepsilon_1.$$

Since $\varepsilon_1 > 0$ can be taken arbitrarily small, the limit above must actually equal zero.

At this point we have proved

$$\lim_{n\to\infty} P\Big\{ \sup_{0\le t\le T} |R_n(t) - (X_- \cdot M)_t| \ge \varepsilon \Big\} = 0 \tag{5.34}$$

under the extra assumption $X_0 = 0$. Suppose $\widetilde{X}$ satisfies the hypotheses of the proposition, but $\widetilde{X}_0$ is not identically zero. Then (5.34) is valid for $X_t = \mathbf{1}_{(0,\infty)}(t)\widetilde{X}_t$. Changing value at $t = 0$ does not affect stochastic integration, so $\widetilde{X} \cdot M = X \cdot M$. Let

$$\widetilde{R}_n(t) = \sum_{i=0}^{\infty} \widetilde{X}(\tau_i^n)\big(M(\tau_{i+1}^n \wedge t) - M(\tau_i^n \wedge t)\big).$$

The conclusion follows for $\widetilde{X}$ if we can show that

$$\sup_{0\le t<\infty} |\widetilde{R}_n(t) - R_n(t)| \to 0 \quad \text{as } n \to \infty.$$

Since $\widetilde{R}_n(t) - R_n(t) = \widetilde{X}(0)\big(M(\tau_1^n \wedge t) - M(0)\big)$, we have the bound

$$\sup_{0\le t<\infty} |\widetilde{R}_n(t) - R_n(t)| \le |\widetilde{X}(0)| \cdot \sup_{0\le t\le \delta_n} |M(t) - M(0)|.$$

The last quantity vanishes almost surely as $n \to \infty$, by the assumption $\delta_n \to 0$ and the cadlag paths of M. In particular it converges to zero in probability.

To summarize, (5.34) now holds for all processes that satisfy the hypotheses. □

Remark 5.33 (Doléans measure)**.** We discuss here briefly the Doléans measure of a local L^2-martingale. It provides an alternative way to define the space $\mathcal{L}(M, \mathcal{P})$ of admissible integrands. The lemma below will be used to extend the stochastic integral beyond predictable integrands, but that point is not central to the main development, so the remainder of this section can be skipped.

Fix a local L^2-martingale M and stopping times $\sigma_k \nearrow \infty$ such that $M^{\sigma_k} \in \mathcal{M}_2$ for each k. By Theorem 3.27 the quadratic variation $[M]$ exists as a nondecreasing cadlag process. Consequently Lebesgue-Stieltjes integrals with respect to $[M]$ are well-defined. The Doléans measure μ_M can be defined for $A \in \mathcal{P}$ by

$$\mu_M(A) = E \int_{[0,\infty)} \mathbf{1}_A(t,\omega)\, d[M]_t(\omega), \tag{5.35}$$

exactly as for L^2-martingales earlier. The measure μ_M is σ-finite: the union of the stochastic intervals $\{[0, \sigma_k \wedge k] : k \in \mathbf{N}\}$ exhausts $\mathbf{R}_+ \times \Omega$, and

$$\begin{aligned}\mu_M([0, \sigma_k \wedge k]) &= E \int_{[0,\infty)} \mathbf{1}_{[0,\sigma_k(\omega)\wedge k]}(t)\, d[M]_t(\omega) = E\{[M]_{\sigma_k \wedge k}\} \\ &= E\{[M^{\sigma_k}]_k\} = E\{(M_k^{\sigma_k})^2 - (M_0^{\sigma_k})^2\} < \infty.\end{aligned}$$

Along the way we used Lemma 3.28 and then the square-integrability of M^{σ_k}.

The following alternative characterization of membership in $\mathcal{L}(M, \mathcal{P})$ will be useful for extending the stochastic integral to non-predictable integrands in Section 5.5.

Lemma 5.34. *Let M be a local L^2-martingale and X a predictable process. Then $X \in \mathcal{L}(M, \mathcal{P})$ iff there exist stopping times $\rho_k \nearrow \infty$ (a.s.) such that for each k,*

$$\int_{[0,T]\times\Omega} \mathbf{1}_{[0,\rho_k]} |X|^2 \, d\mu_M < \infty \quad \text{for all } T < \infty.$$

We leave the proof of this lemma as an exercise. The key point is that for both L^2-martingales and local L^2-martingales, and a stopping time τ, $\mu_{M^\tau}(A) = \mu_M(A \cap [0, \tau])$ for $A \in \mathcal{P}$. (Just check that the proof of Lemma 5.15 applies without change to local L^2-martingales.)

Furthermore, we leave as an exercise proof of the result that if $X, Y \in \mathcal{L}(M, \mathcal{P})$ are μ_M-equivalent, which means again that

$$\mu_M\{(t, \omega) : X(t, \omega) \neq Y(t, \omega)\} = 0,$$

then $X \cdot M = Y \cdot M$ in the sense of indistinguishability.

5.3. Semimartingale integrator

First a reminder of some terminology and results. A cadlag semimartingale is a process Y that can be written as $Y_t = Y_0 + M_t + V_t$ where M is a cadlag local martingale, V is a cadlag FV process, and $M_0 = V_0 = 0$. To define the stochastic integral, we need M to be a local L^2-martingale. If we assume the filtration $\{\mathcal{F}_t\}$ complete and right-continuous (the "usual conditions"), then by Corollary 3.22 we can always select the decomposition so that M is a local L^2-martingale. Thus usual conditions for $\{\mathcal{F}_t\}$ need to be assumed in this section, unless one works with a semimartingale Y for which it is known that M can be chosen a local L^2-martingale. If g is a function of bounded variation on $[0, T]$, then the Lebesgue-Stieltjes measure Λ_g of g exists as a signed Borel measure on $[0, T]$ (Section 1.1.9).

In this section the integrands will be predictable processes X that satisfy this condition:

$$\text{there exist stopping times } \{\sigma_n\} \text{ such that } \sigma_n \nearrow \infty \text{ almost surely and } \mathbf{1}_{(0,\sigma_n]}X \text{ is a bounded process for each } n. \tag{5.36}$$

In particular, the categories listed in Proposition 5.27 are covered. We deliberately ask for $\mathbf{1}_{(0,\sigma_n]}X$ to be bounded instead of $\mathbf{1}_{[0,\sigma_n]}X$ because X_0 might not be bounded.

Definition 5.35. Let Y be a cadlag semimartingale. Let X be a predictable process that satisfies (5.36). Then we define the integral of X with respect to Y as the process

$$\int_{(0,t]} X_s\,dY_s = \int_{(0,t]} X_s\,dM_s + \int_{(0,t]} X_s\,\Lambda_V(ds). \tag{5.37}$$

Here $Y = Y_0 + M + V$ is some decomposition of Y into a local L^2-martingale M and an FV process V,

$$\int_{(0,t]} X_s\,dM_s = (X\cdot M)_t$$

is the stochastic integral of Definition 5.23, and

$$\int_{(0,t]} X_s\,\Lambda_V(ds) = \int_{(0,t]} X_s\,dV_s$$

is the path-by-path Lebesgue-Stieltjes integral of X with respect to the function $s \mapsto V_s$. The process $\int X\,dY$ thus defined is unique up to indistinguishability and it is a semimartingale.

As before, we shall use the notations $X\cdot Y$ and $\int X\,dY$ interchangeably.

Justification of the definition. The first item to check is that the integral does not depend on the decomposition of Y chosen. Suppose $Y = Y_0 + \widetilde{M} + \widetilde{V}$ is another decomposition of Y into a local L^2-martingale $\widetilde{M}$ and an FV process $\widetilde{V}$. We need to show that

$$\int_{(0,t]} X_s\,dM_s + \int_{(0,t]} X_s\,\Lambda_V(ds) = \int_{(0,t]} X_s\,d\widetilde{M}_s + \int_{(0,t]} X_s\,\Lambda_{\widetilde{V}}(ds)$$

in the sense that the processes on either side of the equality sign are indistinguishable. By Proposition 5.31(d) and the additivity of Lebesgue-Stieltjes measures, this is equivalent to

$$\int_{(0,t]} X_s\,d(M-\widetilde{M})_s = \int_{(0,t]} X_s\,\Lambda_{\widetilde{V}-V}(ds).$$

From $Y = M + V = \widetilde{M} + \widetilde{V}$ we get

$$M - \widetilde{M} = \widetilde{V} - V$$

and this process is both a local L^2-martingale and an FV process. The equality we need is a consequence of the next proposition.

Proposition 5.36. *Suppose Z is a cadlag local L^2-martingale and an FV process. Let X be a predictable process that satisfies (5.36). Then for almost every ω*

$$\int_{(0,t]} X(s,\omega)dZ_s(\omega) = \int_{(0,t]} X(s,\omega)\Lambda_{Z(\omega)}(ds) \quad \textit{for all } 0 \le t < \infty. \tag{5.38}$$

On the left is the stochastic integral, on the right the Lebesgue-Stieltjes integral evaluated separately for each fixed ω.

Proof. Both sides of (5.38) are right-continuous in t, so it suffices to check that for each t they agree with probability 1.

Step 1. Start by assuming that Z is an L^2-martingale. Fix $0 < t < \infty$. Let

$$\mathcal{H} = \{X : \ X \text{ is a bounded predictable process and (5.38) holds for } t\}.$$

By the linearity of both integrals, $\mathcal{H}$ is a linear space.

Indicators of predictable rectangles lie in $\mathcal{H}$ because we know explicitly what integrals on both sides of (5.38) look like. If $X = \mathbf{1}_{(u,v]\times F}$ for $F \in \mathcal{F}_u$, then the left side of (5.38) equals $\mathbf{1}_F(Z_{v\wedge t} - Z_{u\wedge t})$ by the first definition (5.7) of the stochastic integral. The right-hand side equals the same thing by the definition of the Lebesgue-Steltjes integral. If $X = \mathbf{1}_{\{0\}\times F_0}$ for $F_0 \in \mathcal{F}_0$, both sides of (5.38) vanish, on the left by the definition (5.7) of the stochastic integral and on the right because the integral is over $(0,t]$ and hence excludes the origin.

Let X be a bounded predictable process, $X_n \in \mathcal{H}$ and $X_n \to X$ pointwise on $\mathbf{R}_+ \times \Omega$. We wish to argue that at least along some subsequence $\{n_j\}$, both sides of

$$\int_{(0,t]} X_n(s,\omega)\, dZ_s(\omega) = \int_{(0,t]} X_n(s,\omega)\, \Lambda_{Z(\omega)}(ds) \tag{5.39}$$

converge for almost every ω to the corresponding integrals with X. This would imply that $X \in \mathcal{H}$. Then we would have checked that the space $\mathcal{H}$ satisfies the hypotheses of Theorem B.4. (The π-system for the theorem is the class of predictable rectangles.)

On the right-hand side of (5.39) the desired convergence follows from dominated convergence. For a fixed ω, the BV function $s \mapsto Z_s(\omega)$ on $[0,t]$ can be expressed as the difference $Z_s(\omega) = f(s) - g(s)$ of two nondecreasing functions. Hence the signed measure $\Lambda_{Z(\omega)}$ is the difference of two finite positive measures: $\Lambda_{Z(\omega)} = \Lambda_f - \Lambda_g$. Then

$$\begin{aligned}
&\lim_{n\to\infty} \int_{(0,t]} X_n(s,\omega)\,\Lambda_{Z(\omega)}(ds) \\
&= \lim_{n\to\infty} \int_{(0,t]} X_n(s,\omega)\,\Lambda_f(ds) - \lim_{n\to\infty} \int_{(0,t]} X_n(s,\omega)\,\Lambda_g(ds) \\
&= \int_{(0,t]} X(s,\omega)\,\Lambda_f(ds) - \int_{(0,t]} X(s,\omega)\,\Lambda_g(ds) \\
&= \int_{(0,t]} X(s,\omega)\,\Lambda_{Z(\omega)}(ds).
\end{aligned}$$

The limits are applications of the usual dominated convergence theorem, because $-C \le X_1 \le X_n \le X \le C$ for some constant C.

Now the left side of (5.39). For a fixed $T < \infty$, by dominated convergence,

$$\lim_{n\to\infty} \int_{[0,T]\times\Omega} |X - X_n|^2 \, d\mu_Z = 0.$$

Hence $\|X_n - X\|_{\mathcal{L}_2(Z,\mathcal{P})} \to 0$, and by the isometry $X_n \cdot Z \to X \cdot Z$ in $\mathcal{M}_2$, as $n \to \infty$. Then for a fixed t, $(X_{n_j} \cdot Z)_t \to (X \cdot Z)_t$ almost surely along some subsequence $\{n_j\}$. Thus taking the limit along $\{n_j\}$ on both sides of (5.39) gives

$$\int_{(0,t]} X(s,\omega)\, dZ_s(\omega) = \int_{(0,t]} X(s,\omega)\,\Lambda_{Z(\omega)}(ds)$$

almost surely. By Theorem B.4, $\mathcal{H}$ contains all bounded $\mathcal{P}$-measurable processes.

This completes Step 1: (5.38) has been verified for the case where $Z \in \mathcal{M}_2$ and X is bounded.

Step 2. Now consider the case of a local L^2-martingale Z. By the assumption on X we may pick a localizing sequence $\{\tau_k\}$ such that Z^{τ_k} is an L^2-martingale and $\mathbf{1}_{(0,\tau_k]}X$ is bounded. Then by Step 1,

$$\int_{(0,t]} \mathbf{1}_{(0,\tau_k]}(s) X(s)\, dZ_s^{\tau_k} = \int_{(0,t]} \mathbf{1}_{(0,\tau_k]}(s) X(s)\,\Lambda_{Z^{\tau_k}}(ds). \tag{5.40}$$

We claim that on the event $\{\tau_k \ge t\}$ the left and right sides of (5.40) coincide almost surely with the corresponding sides of (5.38).

The left-hand side of (5.40) coincides almost surely with $\big((\mathbf{1}_{[0,\tau_k]}X)\cdot Z^{\tau_k}\big)_t$ due to the irrelevance of the time origin. By (5.27) this is the definition of $(X \cdot Z)_t$ on the event $\{\tau_k \ge t\}$.

On the right-hand side of (5.40) we only need to observe that if $\tau_k \ge t$, then on the interval $(0,t]$, $\mathbf{1}_{(0,\tau_k]}(s)X(s)$ coincides with $X(s)$ and $Z_s^{\tau_k}$

coincides with Z_s. So it is clear that the integrals on the right-hand sides of (5.40) and (5.38) coincide.

The union over k of the events $\{\tau_k \ge t\}$ equals almost surely the whole space Ω. Thus we have verified (5.38) almost surely, for this fixed t. $\square$

Returning to the justification of the definition, we now know that the process $\int X\,dY$ does not depend on the choice of the decomposition $Y = Y_0 + M + V$.

$X \cdot M$ is a local L^2-martingale, and for a fixed ω the function

$$t \mapsto \int_{(0,t]} X_s(\omega)\,\Lambda_{V(\omega)}(ds)$$

has bounded variation on every compact interval (Lemma 1.15). Thus the definition (5.37) provides the semimartingale decomposition of $\int X\,dY$. $\square$

As in the previous step of the development, we want to check the Riemann sum approximations. Recall that for a cadlag X, the caglad process X_- is defined by $X_-(0) = X(0)$ and $X_-(t) = X(t-)$ for $t > 0$. The hypotheses for the integrand in the next proposition are exactly the same as earlier in Proposition 5.32.

Parts (a) and (b) below could be subsumed under a single statement since $X = X_-$ for a left-continuous process. We prefer to keep them separate to avoid confusing the issue that for a cadlag process X the limit is not necessarily the stochastic integral of X. The integral $X \cdot Y$ may fail to exist, and even if it exists, it does not necessarily coincide with $X_- \cdot Y$. This is not a consequence of the stochastic aspect, but can happen also for Lebesgue-Stieltjes integrals. (Find examples!)

Proposition 5.37. *Let X be an adapted process and Y a cadlag semimartingale. Suppose $0 = \tau_0^n \le \tau_1^n \le \tau_2^n \le \tau_3^n \le \cdots$ are stopping times such that for each n, $\tau_i^n \to \infty$ almost surely as $i \to \infty$, and $\delta_n = \sup_{0 \le i < \infty}(\tau_{i+1}^n - \tau_i^n)$ tends to zero almost surely as $n \to \infty$. Define*

$$S_n(t) = \sum_{i=0}^{\infty} X(\tau_i^n)\bigl(Y(\tau_{i+1}^n \wedge t) - Y(\tau_i^n \wedge t)\bigr). \tag{5.41}$$

(a) *Assume X is left-continuous and satisfies (5.36). Then for each fixed $T < \infty$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\Bigl\{ \sup_{0 \le t \le T} |S_n(t) - (X \cdot Y)_t| \ge \varepsilon \Bigr\} = 0.$$

In other words, S_n converges to $X \cdot Y$ in probability, uniformly on compact time intervals.

(b) *If X is an adapted cadlag process, then S_n converges to $X_- \cdot Y$ in probability, uniformly on compacts.*

Proof. Pick a decomposition $Y = Y_0 + M + V$. We get the corresponding decomposition $S_n = R_n + U_n$ by defining

$$R_n(t) = \sum_{i=0}^{\infty} X(\tau_i^n)\big(M_{\tau_{i+1}^n \wedge t} - M_{\tau_i^n \wedge t}\big)$$

as in Proposition 5.32, and

$$U_n(t) = \sum_{i=0}^{\infty} X(\tau_i^n)\big(V_{\tau_{i+1}^n \wedge t} - V_{\tau_i^n \wedge t}\big).$$

Proposition 5.32 gives the convergence $R_n \to X_- \cdot M$. Lemma 1.12 applied to the Lebesgue-Stieltjes measure of $V_t(\omega)$ tells us that, for almost every fixed ω

$$\lim_{n\to\infty} \sup_{0\le t\le T} \left| U_n(t,\omega) - \int_{(0,t]} X(s-,\omega)\, \Lambda_{V(\omega)}(ds) \right| = 0.$$

The reservation "almost every ω" is needed in case there is an exceptional zero probability event on which X or V fails to have the good path properties. Almost sure convergence implies convergence in probability. □

Remark 5.38. Matrix-valued integrands and vector-valued integrators. In order to consider equations for vector-valued processes, we need to establish the (obvious componentwise) conventions regarding the integrals of matrix-valued processes with vector-valued integrators. Let a probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}$ be given. If $Q_{i,j}(t)$, $1 \le i \le d$ and $1 \le j \le m$, are predictable processes on this space, then we regard $Q(t) = (Q_{i,j}(t))$ as a $d \times m$-matrix valued predictable process. And if Y_1, $\dots$, Y_m are semimartingales on this space, then $Y = (Y_1, \dots, Y_m)^T$ is an $\mathbf{R}^m$-valued semimartingale. The stochastic integral $Q \cdot Y = \int Q\, dY$ is the $\mathbf{R}^d$-valued process whose ith component is

$$(Q \cdot Y)_i(t) = \sum_{j=1}^{n} \int_{(0,t]} Q_{i,j}(s)\, dY_j(s)$$

assuming of course that all the componentwise integrals are well-defined.

One note of caution: the definition of a d-dimensional standard Brownian motion $B(t) = [B_1(t), \dots, B_d(t)]^T$ includes the requirement that the coordinates be independent of each other. For a vector-valued semimartingale, it is only required that each coordinate be marginally a semimartingale.

5.4. Further properties of stochastic integrals

Now that we have constructed the stochastic integral, subsequent chapters on Itô's formula and stochastic differential equations develop techniques that can be applied to build models and study the behavior of those models. In this section we develop further properties of the integral as groundwork for later chapters. The content of Sections 5.4.2–5.4.4 is fairly technical and is used only in the proof of existence and uniqueness for a semimartingale equation in Section 7.4.

The usual conditions on $\{\mathcal{F}_t\}$ need to be assumed only insofar as this is needed for a definition of the integral with respect to a semimartingale. For the proofs of this section right-continuity of $\{\mathcal{F}_t\}$ is not needed. (Completeness we assume always.)

The properties listed in Proposition 5.31 extend readily to the semimartingale intergal. Each property is linear in the integrator, holds for the martingale part by the proposition, and can be checked path by path for the FV part. We state the important ones for further reference and leave the proofs as exercises.

Proposition 5.39. *Let Y and Z be semimartingales, G and H predictable processes that satisfy the local boundedness condition (5.36), and let τ be a stopping time.*

(a) *Let U be a bounded $\mathcal{F}_\tau$-measurable random variable. Then*

$$\int_{(0,t]} U\mathbf{1}_{(\tau,\infty)} G\,dY = U\int_{(0,t]} \mathbf{1}_{(\tau,\infty)} G\,dY. \tag{5.42}$$

Furthermore,

$$\big((\mathbf{1}_{[0,\tau]}G)\cdot Y\big)_t = (G\cdot Y)_{\tau\wedge t} = (G\cdot Y^\tau)_t. \tag{5.43}$$

(b) *Suppose $G_t(\omega) = H_t(\omega)$ and $Y_t(\omega) = Z_t(\omega)$ for $0 \le t \le \tau(\omega)$. Then*

$$(G\cdot Y)_{\sigma\wedge t} = (H\cdot Z)_{\sigma\wedge t}.$$

5.4.1. Jumps of a stochastic integral.

For any cadlag process Z, $\Delta Z(t) = Z(t) - Z(t-)$ denotes the jump at t. First a strengthening of the part of Proposition 2.16 that identifies the jumps of the quadratic variation. The strengthening is that the "almost surely" qualifier is not applied separately to each t.

Lemma 5.40. *Let Y be a semimartingale. Then the quadratic variation $[Y]$ exists. For almost every ω, $\Delta[Y]_t = (\Delta Y_t)^2$ for all $0 < t < \infty$.*

Proof. Fix $0 < T < \infty$. By Proposition 5.37, we can pick a sequence of partitions $\pi^n = \{t_i^n\}$ of $[0, T]$ such that the process

$$S_n(t) = 2\sum_i Y_{t_i^n}(Y_{t_{i+1}^n\wedge t} - Y_{t_i^n\wedge t}) = Y_t^2 - Y_0^2 - \sum_i (Y_{t_{i+1}^n\wedge t} - Y_{t_i^n\wedge t})^2$$

converges to the process

$$S(t) = 2\int_{(0,t]} Y(s-)\, dY(s),$$

uniformly for $t \in [0, T]$, for almost every ω. This implies the convergence of the sum of squares, so the quadratic variation $[Y]$ exists and satisfies

$$[Y]_t = Y_t^2 - Y_0^2 - S(t).$$

It is true in general that a uniform bound on the difference of two functions gives a bound on the difference of the jumps: if f and g are cadlag and $|f(x) - g(x)| \le \varepsilon$ for all x, then

$$|\Delta f(x) - \Delta g(x)| = \lim_{y\nearrow x,\, y<x} |f(x) - f(y) - g(x) + g(y)| \le 2\varepsilon.$$

Fix an ω for which the uniform convergence $S_n \to S$ holds. Then for each $t \in (0, T]$, the jump $\Delta S_n(t)$ converges to $\Delta S(t) = \Delta(Y^2)_t - \Delta[Y]_t$.

Directly from the definition of S_n one sees that $\Delta S_n(t) = 2Y_{t_k^n}\Delta Y_t$ for the index k such that $t \in (t_k^n, t_{k+1}^n]$. Here is the calculation in detail: if $s < t$ is close enough to t, also $s \in (t_k^n, t_{k+1}^n]$, and then

$$\begin{aligned}
\Delta S_n(t) &= \lim_{s\nearrow t,\, s<t} \bigl\{S_n(t) - S_n(s)\bigr\} \\
&= \lim_{s\nearrow t,\, s<t} \Bigl\{2\sum_i Y_{t_i^n}(Y_{t_{i+1}^n\wedge t} - Y_{t_i^n\wedge t}) \;-\; 2\sum_i Y_{t_i^n}(Y_{t_{i+1}^n\wedge s} - Y_{t_i^n\wedge s})\Bigr\} \\
&= \lim_{s\nearrow t,\, s<t} \bigl\{2Y_{t_k^n}(Y_t - Y_{t_k^n}) \;-\; 2Y_{t_k^n}(Y_s - Y_{t_k^n})\bigr\} \\
&= \lim_{s\nearrow t,\, s<t} 2Y_{t_k^n}(Y_t - Y_s) = 2Y_{t_k^n}\Delta Y_t.
\end{aligned}$$

By the cadlag path of Y, $\Delta S_n(t) \to 2Y_{t-}\Delta Y_t$. Equality of the two limits of $S_n(t)$ gives

$$2Y_{t-}\Delta Y_t = \Delta(Y^2)_t - \Delta[Y]_t$$

which rearranges to $\Delta[Y]_t = (\Delta Y_t)^2$. □

Theorem 5.41. *Let Y be a cadlag semimartingale, X a predictable process that satisfies the local boundedness condition (5.36), and $X \cdot Y$ the stochastic integral. Then for all ω in a set of probability one,*

$$\Delta(X \cdot Y)(t) = X(t)\Delta Y(t) \quad \textit{for all } 0 < t < \infty.$$

We prove this theorem in stages. The reader may be surprised by the appearance of the point value $X(t)$ in a statement about integrals. After all, one of the lessons of integration theory is that point values of functions are often irrelevant. But implicit in the proof below is the notion that a point value of the integrand X influences the integral exactly when the integrator Y has a jump.

Lemma 5.42. *Let M be a cadlag local L^2-martingale and $X \in \mathcal{L}(M, \mathcal{P})$. Then for all ω in a set of probability one, $\Delta(X \cdot M)(t) = X(t)\Delta M(t)$ for all $0 < t < \infty$.*

Proof. Suppose the conclusion holds if $M \in \mathcal{M}_2$ and $X \in \mathcal{L}_2(M, \mathcal{P})$. Pick a sequence $\{\sigma_k\}$ that localizes (X, M) and let $X_k = \mathbf{1}_{[0,\sigma_k]}X$. Fix ω such that definition (5.27) of the integral $X \cdot M$ works and the conclusion above holds for each integral $X_k \cdot M^{\sigma_k}$. Then if $\sigma_k > t$,

$$\Delta(X \cdot M)_t = \Delta(X_k \cdot M^{\tau_k})_t = X_k(t)\Delta M_t^{\tau_k} = X(t)\Delta M_t.$$

For the remainder of the proof we may assume that $M \in \mathcal{M}_2$ and $X \in \mathcal{L}_2(M, \mathcal{P})$. Pick simple predictable processes

$$X_n(t) = \sum_{i=1}^{m(n)-1} \xi_i^n \mathbf{1}_{(t_i^n, t_{i+1}^n]}(t)$$

such that $X_n \to X$ in $\mathcal{L}_2(M, \mathcal{P})$. By the definition of Lebesgue-Stieltjes integrals, and because by Lemma 5.40 the processes $\Delta[M]_t$ and $(\Delta M_t)^2$ are indistinguishable,

$$E\biggl[\sum_{s\in(0,T]} \bigl|X_n(s)\Delta M_s - X(s)\Delta M_s\bigr|^2\biggr] = E\biggl[\sum_{s\in(0,T]} \bigl|X_n(s) - X(s)\bigr|^2 \Delta[M]_s\biggr]$$
$$\le E\int_{(0,T]} \bigl|X_n(s) - X(s)\bigr|^2 \, d[M]_s$$

and by hypothesis the last expectation vanishes as $n \to \infty$. Thus there exists a subsequence n_j along which almost surely

$$\lim_{j\to\infty} \sum_{s\in(0,T]} \bigl|X_{n_j}(s)\Delta M_s - X(s)\Delta M_s\bigr|^2 = 0.$$

In particular, on this event of full probability, for any $t \in (0, T]$

$$\lim_{j\to\infty} \bigl|X_{n_j}(t)\Delta M_t - X(t)\Delta M_t\bigr|^2 = 0. \tag{5.44}$$

On the other hand, $X_{n_j} \cdot M \to X \cdot M$ in $\mathcal{M}_2$ implies that along a further subsequence (which we denote by the same n_j), almost surely,

$$\lim_{j\to\infty} \sup_{t\in[0,T]} \bigl|(X_{n_j} \cdot M)_t - (X \cdot M)_t\bigr| = 0. \tag{5.45}$$

Fix an ω on which both almost sure limits (5.44) and (5.45) hold. For any $t \in (0, T]$, the uniform convergence in (5.45) implies $\Delta(X_{n_j} \cdot M)_t \to \Delta(X \cdot M)_t$. Also, since a path of $X_{n_j} \cdot M$ is a step function, $\Delta(X_{n_j} \cdot M)_t = X_{n_j}(t)\Delta M_t$. (The last two points were justified explicitly in the proof of Lemma 5.40 above.) Combining these with the limit (5.44) shows that, for this fixed ω and all $t \in (0, T]$,

$$\Delta(X \cdot M)_t = \lim_{j\to\infty} \Delta(X_{n_j} \cdot M)_t = \lim_{j\to\infty} X_{n_j}(t)\Delta M_t = X(t)\Delta M_t.$$

In other words, the conclusion holds almost surely on $[0, T]$. To finish, consider countably many T that increase up to ∞. □

Lemma 5.43. *Let f be a bounded Borel function and U a BV function on $[0, T]$. Denote the Lebesgue-Stieltjes integral by*

$$(f \cdot U)_t = \int_{(0,t]} f(s)\, dU(s).$$

Then $\Delta(f \cdot U)_t = f(t)\Delta U(t)$ for all $0 < t \le T$.

Proof. By the rules concerning Lebesgue-Stieltjes integration,

$$\begin{aligned}(f \cdot U)_t - (f \cdot U)_s &= \int_{(s,t]} f(r)\, dU(r) \\ &= \int_{(s,t)} f(r)\, dU(r) + f(t)\Delta U(t)\end{aligned}$$

and

$$\left| \int_{(s,t)} f(s)\, dU(s) \right| \le \|f\|_\infty \Lambda_{V_U}(s, t).$$

V_U is the total variation function of U. As $s \nearrow t$, the set (s, t) decreases down to the empty set. Since Λ_{V_U} is a finite positive measure on $[0, T]$, $\Lambda_{V_U}(s, t) \searrow 0$ as $s \nearrow t$. □

Proof of Theorem 5.41 follows from combining Lemmas 5.42 and 5.43. We introduce the following notation for the left limit of a stochastic integral:

$$\int_{(0,t)} H\, dY = \lim_{s\nearrow t,\, s<t} \int_{(0,s]} H\, dY \tag{5.46}$$

and then we have this identity:

$$\int_{(0,t]} H\, dY = \int_{(0,t)} H\, dY + H(t)\Delta Y(t). \tag{5.47}$$

5.4.2. Convergence theorem for stochastic integrals. The next theorem is a dominated convergence theorem of sorts for stochastic integrals. We will use it in the existence proof for solutions of stochastic differential equations.

Theorem 5.44. *Let $\{H_n\}$ be a sequence of predictable processes and G_n a sequence of nonnegative adapted cadlag processes. Assume $|H_n(t)| \le G_n(t-)$ for all $0 < t < \infty$, and the running maximum*

$$G_n^*(T) = \sup_{0 \le t \le T} G_n(t)$$

converges to zero in probability, for each fixed $0 < T < \infty$. Then for any cadlag semimartingale Y, $H_n \cdot Y \to 0$ in probability, uniformly on compact time intervals.

Proof. Let $Y = Y_0 + M + U$ be a decomposition of Y into a local L^2-martingale M and an FV process U. We show that both terms in $H_n \cdot Y = H_n \cdot M + H_n \cdot U$ converge to zero in probability, uniformly on compact time intervals. The FV part is immediate:

$$\sup_{0 \le t \le T} \left| \int_{(0,t]} H_n(s)\, dU(s) \right| \le \sup_{0 \le t \le T} \int_{(0,t]} |H_n(s)|\, dV_U(s) \le G_n^*(T) V_U(T)$$

where we applied inequality (1.15). Since $V_U(T)$ is a finite random variable, the last bound above converges to zero in probability.

For the local martingale part $H_n \cdot M$, pick a sequence of stopping times $\{\sigma_k\}$ that localizes M. Let

$$\rho_n = \inf\{t \ge 0 : G_n(t) \ge 1\} \wedge \inf\{t > 0 : G_n(t-) \ge 1\}.$$

These are stopping times by Lemma 2.9. By left-continuity, $G_n(t-) \le 1$ for $0 < t \le \rho_n$. For any $T < \infty$,

$$P\{\rho_n \le T\} \le P\{G_n^*(T) > 1/2\} \to 0 \quad \text{as } n \to \infty.$$

Let $H_n^{(1)} = (H_n \wedge 1) \vee (-1)$ denote the bounded process obtained by truncation. If $t \le \sigma_k \wedge \rho_n$ then $H_n \cdot M = H_n^{(1)} \cdot M^{\sigma_k}$ by part (c) of Proposition 5.31. As a bounded process $H_n^{(1)} \in \mathcal{L}_2(M^{\sigma_k}, \mathcal{P})$, so by martingale inequality

(3.8) and the stochastic integral isometry (5.12),

$$\begin{aligned}
P\Big\{ \sup_{0\le t\le T} |(H_n\cdot M)_t| \ge \varepsilon \Big\} &\le P\{\sigma_k \le T\} + P\{\rho_n \le T\} \\
&\qquad + P\Big\{ \sup_{0\le t\le T} \big|\big(H_n^{(1)}\cdot M^{\sigma_k}\big)_t\big| \ge \varepsilon \Big\} \\
&\le P\{\sigma_k \le T\} + P\{\rho_n \le T\} + \varepsilon^{-2} E\big[(H_n^{(1)}\cdot M^{\sigma_k})_T^2\big] \\
&= P\{\sigma_k \le T\} + P\{\rho_n \le T\} + \varepsilon^{-2} E\int_{[0,T]} |H_n^{(1)}(t)|^2\, d[M^{\sigma_k}]_t \\
&\le P\{\sigma_k \le T\} + P\{\rho_n \le T\} + \varepsilon^{-2} E\int_{[0,T]} \big(G_n(t-)\wedge 1\big)^2\, d[M^{\sigma_k}]_t \\
&\le P\{\sigma_k \le T\} + P\{\rho_n \le T\} + \varepsilon^{-2} E\Big\{ [M^{\sigma_k}]_T \big(G_n^*(T)\wedge 1\big)^2 \Big\}.
\end{aligned}$$

As k and T stay fixed and $n\to\infty$, the last expectation above tends to zero. This follows from the dominated convergence theorem under convergence in probability (Theorem B.12). The integrand is bounded by the integrable random variable $[M^{\sigma_k}]_T$. Given $\delta > 0$, pick $K > \delta$ so that

$$P\big\{ [M^{\sigma_k}]_T \ge K \big\} < \delta/2.$$

Then

$$P\Big\{ [M^{\sigma_k}]_T \big(G_n^*(T)\wedge 1\big)^2 \ge \delta \Big\} \le \delta/2 + P\big\{ G_n^*(T) \ge \sqrt{\delta/K}\, \big\}$$

where the last probability vanishes as $n\to\infty$, by the assumption that $G_n^*(T)\to 0$ in probability.

Returning to the string of inequalities and letting $n\to\infty$ gives

$$\varlimsup_{n\to\infty} P\Big\{ \sup_{0\le t\le T} |(H_n\cdot M)_t| \ge \varepsilon \Big\} \le P\{\sigma_k \le T\}.$$

This last bound tends to zero as $k\to\infty$, and so we have proved that $H_n\cdot M\to 0$ in probability, uniformly on $[0,T]$. ☐

5.4.3. Restarting at a stopping time. Let Y be a cadlag semimartingale and G a predictable process that satisfies the local boundedness condition (5.36). Let σ be a bounded stopping time with respect to the underlying filtration $\{\mathcal{F}_t\}$. We take a bounded stopping time in order to get moment bounds from Lemma 3.5. Define a new filtration $\bar{\mathcal{F}}_t = \mathcal{F}_{\sigma+t}$. Let $\bar{\mathcal{P}}$ be the predictable σ-field under the filtration $\{\bar{\mathcal{F}}_t\}$. In other words, $\bar{\mathcal{P}}$ is the σ-field generated by sets of the type $(u,v]\times\Gamma$ for $\Gamma\in\bar{\mathcal{F}}_u$ and $\{0\}\times\Gamma_0$ for $\Gamma_0\in\bar{\mathcal{F}}_0$.

Define new processes

$$\bar{Y}(t) = Y(\sigma+t) - Y(\sigma) \quad\text{ and }\quad \bar{G}(s) = G(\sigma+s).$$

For $\bar{Y}$ we could define just as well $\bar{Y}(t) = Y(\sigma+t)$ without changing the statements below. The reason is that $\bar{Y}$ appears only as an integrator, so only its increments matter. Sometimes it might be convenient to have initial value zero, that is why we subtract $Y(\sigma)$ from $\bar{Y}(t)$.

Theorem 5.45. *Let σ be a bounded stopping time with respect to $\{\mathcal{F}_t\}$. Under the filtration $\{\bar{\mathcal{F}}_t\}$, the process $\bar{G}$ is predictable and $\bar{Y}$ is an adapted semimartingale. We have this equality of stochastic integrals:*

$$\begin{aligned}\int_{(0,t]} \bar{G}(s)\, d\bar{Y}(s) &= \int_{(\sigma,\sigma+t]} G(s)\, dY(s) \\ &= \int_{(0,\sigma+t]} G(s)\, dY(s) - \int_{(0,\sigma]} G(s)\, dY(s).\end{aligned} \tag{5.48}$$

The second equality in (5.48) is the definition of the integral over $(\sigma, \sigma + t]$. The proof of Theorem 5.45 follows after two lemmas.

Lemma 5.46. *For any $\mathcal{P}$-measurable function G, $\bar{G}(t,\omega) = G(\sigma(\omega) + t, \omega)$ is $\bar{\mathcal{P}}$-measurable.*

Proof. Let $\mathcal{U}$ be the space of $\mathcal{P}$-measurable functions G such that $\bar{G}$ is $\bar{\mathcal{P}}$-measurable. $\mathcal{U}$ is a linear space and closed under pointwise limits since these operations preserve measurability. Since any $\mathcal{P}$-measurable function is a pointwise limit of bounded $\mathcal{P}$-measurable functions, it suffices to show that $\mathcal{U}$ contains all bounded $\mathcal{P}$-measurable functions. According to Theorem B.4 we need to check that $\mathcal{U}$ contains indicator functions of predictable rectangles. We leave the case $\{0\} \times \Gamma_0$ for $\Gamma_0 \in \mathcal{F}_0$ to the reader.

Let $\Gamma \in \mathcal{F}_u$ and $G = \mathbf{1}_{(u,v]\times\Gamma}$. Then

$$\bar{G}(t) = \mathbf{1}_{(u,v]}(\sigma + t)\mathbf{1}_{\Gamma}(\omega) = \begin{cases} \mathbf{1}_{\Gamma}(\omega), & u < \sigma + t \le v \\ 0, & \text{otherwise.} \end{cases}$$

For a fixed ω, $\bar{G}(t)$ is a caglad process. By Lemma 5.1, $\bar{\mathcal{P}}$-measurability of $\bar{G}$ follows if it is adapted to $\{\bar{\mathcal{F}}_t\}$. Since $\{\bar{G}(t) = 1\} = \Gamma \cap \{u < \sigma + t \le v\}$, $\bar{G}$ is adapted to $\{\bar{\mathcal{F}}_t\}$ if $\Gamma \cap \{u < \sigma + t \le v\} \in \bar{\mathcal{F}}_t$. This is true by Lemma 2.3 because u, v and $\sigma + t$ can be regarded as stopping times and $\bar{\mathcal{F}}_t = \mathcal{F}_{\sigma+t}$. $\square$

Lemma 5.47. *Let M be a local L^2-martingale with respect to $\{\mathcal{F}_t\}$. Then $\bar{M}_t = M_{\sigma+t} - M_\sigma$ is a local L^2-martingale with respect to $\{\bar{\mathcal{F}}_t\}$. If $G \in \mathcal{L}(M, \mathcal{P})$, then $\bar{G} \in \mathcal{L}(\bar{M}, \bar{\mathcal{P}})$, and*

$$(\bar{G} \cdot \bar{M})_t = (G \cdot M)_{\sigma+t} - (G \cdot M)_\sigma.$$

Proof. Let $\{\tau_k\}$ localize (G, M). Let $\nu_k = (\tau_k - \sigma)^+$. For any $0 \le t < \infty$,

$$\{\nu_k \le t\} = \{\tau_k \le \sigma + t\} \in \mathcal{F}_{\sigma+t} = \bar{\mathcal{F}}_t \quad \text{by Lemma 2.3(ii)}$$

so ν_k is a stopping time for the filtration $\{\bar{\mathcal{F}}_t\}$. If $\sigma \le \tau_k$,

$$\bar{M}^{\nu_k}_t = \bar{M}_{\nu_k \wedge t} = M_{\sigma+(\nu_k\wedge t)} - M_\sigma \quad = M_{(\sigma+t)\wedge\tau_k} - M_{\sigma\wedge\tau_k} = M^{\tau_k}_{\sigma+t} - M^{\tau_k}_\sigma.$$

If $\sigma > \tau_k$, then $\nu_k = 0$ and $\bar{M}^{\nu_k}_t = \bar{M}_0 = M_\sigma - M_\sigma = 0$. The earlier formula also gives zero in case $\sigma > \tau_k$, so in both cases we can write

$$\bar{M}^{\nu_k}_t = M^{\tau_k}_{\sigma+t} - M^{\tau_k}_\sigma. \tag{5.49}$$

Since M^{τ_k} is an L^2-martingale, the right-hand side above is an L^2-martingale with respect to $\{\bar{\mathcal{F}}_t\}$. See Corollary 3.9 and point (ii) after the corollary. Consequently $\bar{M}^{\nu_k}$ is an L^2-martingale with respect to $\{\bar{\mathcal{F}}_t\}$, and hence $\bar{M}$ is a local L^2-martingale.

Next we show that $\{\nu_k\}$ localizes $(\bar{G}, \bar{M})$. We fix k temporarily and write $Z = M^{\tau_k}$ to lessen the notational burden. Then (5.49) shows that

$$\bar{M}^{\nu_k}_t = \bar{Z}_t \equiv Z_{\sigma+t} - Z_\sigma,$$

in other words obtained from Z by the operation of restarting at σ. We turn to look at the measure $\mu_{\bar{Z}} = \mu_{\bar{M}^{\nu_k}}$. Let $\Gamma \in \mathcal{F}_u$. Then $\mathbf{1}_{(u,v]\times\Gamma}$ is $\mathcal{P}$-measurable, while $\mathbf{1}_{(u,v]\times\Gamma}(\sigma+t,\omega)$ is $\bar{\mathcal{P}}$-measurable.

$$\begin{aligned}
\int \mathbf{1}_\Gamma(\omega)\mathbf{1}_{(u,v]}(\sigma+t)\,\mu_{\bar{Z}}(dt,d\omega) &= E\Big[\mathbf{1}_\Gamma(\omega)\big([\bar{Z}]_{(v-\sigma)^+} - [\bar{Z}]_{(u-\sigma)^+}\big)\Big] \\
&= E\Big[\mathbf{1}_\Gamma(\omega)\big(\bar{Z}_{(v-\sigma)^+} - \bar{Z}_{(u-\sigma)^+}\big)^2\Big] \\
&= E\Big[\mathbf{1}_\Gamma(\omega)\big(Z_{\sigma+(v-\sigma)^+} - Z_{\sigma+(u-\sigma)^+}\big)^2\Big] \\
&= E\Big[\mathbf{1}_\Gamma(\omega)\big([Z]_{\sigma\vee v} - [Z]_{\sigma\vee u}\big)\Big] \\
&= E\Big[\mathbf{1}_\Gamma(\omega)\int \mathbf{1}_{(\sigma(\omega),\infty)}(t)\mathbf{1}_{(u,v]}(t)\,d[Z]_t\Big] \\
&= \int \mathbf{1}_{(\sigma,\infty)}(t,\omega)\mathbf{1}_\Gamma(\omega)\mathbf{1}_{(u,v]}(t)\,d\mu_Z.
\end{aligned}$$

Several observations go into the calculation. First, $(u-\sigma)^+$ is a stopping time for $\{\bar{\mathcal{F}}_t\}$, and $\Gamma \in \bar{\mathcal{F}}_{(u-\sigma)^+}$. We leave checking these as exercises. This justifies using optional stopping for the martingale $\bar{Z}^2 - [\bar{Z}]$. Next note that $\sigma + (u-\sigma)^+ = \sigma \vee u$, and again the martingale property is justified by $\Gamma \in \mathcal{F}_{\sigma\vee u}$. Finally $(\sigma\vee u\,,\,\sigma\vee v] = (\sigma,\infty)\cap(u,v]$.

From predictable rectangles $(u,v]\times\Gamma \in \mathcal{P}$ this integral identity extends to nonnegative predictable processes X to give, with $\bar{X}(t) = X(\sigma+t)$,

$$\int \bar{X}\,d\mu_{\bar{Z}} = \int \mathbf{1}_{(\sigma,\infty)} X\,d\mu_Z. \tag{5.50}$$

Let $T < \infty$ and apply this to $X = \mathbf{1}_{(\sigma,\sigma+T]}(t)\mathbf{1}_{(0,\tau_k]}(t)|G(t)|^2$. Then

$$\bar{X}(t) = \mathbf{1}_{[0,T]}(t)\mathbf{1}_{(0,\nu_k]}(t)\bar{G}(t)$$

and

$$\int_{[0,T]\times\Omega} \mathbf{1}_{(0,\nu_k]}(t)|\bar{G}(t)|^2\, d\mu_{\bar{Z}} = \int \mathbf{1}_{(\sigma,\sigma+T]}(t)\mathbf{1}_{(0,\tau_k]}(t)|G(t)|^2\, d\mu_Z < \infty,$$

where the last inequality is a consequence of the assumptions that $\{\tau_k\}$ localizes (G, M) and σ is bounded.

To summarize thus far: we have shown that $\{\nu_k\}$ localizes $(\bar{G}, \bar{M})$. This checks $\bar{G} \in \mathcal{L}(\bar{M}, \bar{\mathcal{P}})$.

Fix again k and continue denoting the L^2-martingales by $Z = M^{\tau_k}$ and $\bar{Z} = \bar{M}^{\nu_k}$. Consider a simple $\mathcal{P}$-predictable process

$$H_n(t) = \sum_{i=0}^{m-1} \xi_i \mathbf{1}_{(u_i,u_{i+1}]}(t).$$

Let k denote the index that satisfies $u_{k+1} > \sigma \ge u_k$. (If there is no such k then $\bar{H}_n = 0$.) Then

$$\bar{H}_n(t) = \sum_{i\ge k} \xi_i \mathbf{1}_{(u_i-\sigma,u_{i+1}-\sigma]}(t).$$

The stochastic integral is

$$\begin{aligned}(\bar{H}_n \cdot \bar{Z})_t &= \sum_{i\ge k} \xi_i\bigl(\bar{Z}_{(u_{i+1}-\sigma)\wedge t} - \bar{Z}_{(u_i-\sigma)^+\wedge t}\bigr)\\ &= \sum_{i>k} \xi_i\bigl(Z_{\sigma+(u_{i+1}-\sigma)\wedge t} - Z_{\sigma+(u_i-\sigma)\wedge t}\bigr)\\ &\qquad + \xi_k\bigl(Z_{\sigma+(u_{k+1}-\sigma)\wedge t} - Z_\sigma\bigr).\end{aligned}$$

The $i = k$ term above develops differently from the others because $\bar{Z}_{(u_k-\sigma)^+\wedge t} = \bar{Z}_0 = 0$. Observe that

$$\begin{cases}\sigma + (u_i - \sigma)\wedge t = (\sigma + t)\wedge u_i & \text{for } i > k\\ (\sigma + t)\wedge u_i = \sigma \wedge u_i = u_i & \text{for } i \le k.\end{cases}$$

Now continue from above:

$$\begin{aligned}(\bar{H}_n \cdot \bar{Z})_t &= \sum_{i>k} \xi_i\bigl(Z_{(\sigma+t)\wedge u_{i+1}} - Z_{(\sigma+t)\wedge u_i}\bigr)\\ &\qquad + \xi_k\bigl(Z_{(\sigma+t)\wedge u_{k+1}} - Z_\sigma\bigr)\\ &= \sum_i \xi_i\bigl(Z_{(\sigma+t)\wedge u_{i+1}} - Z_{(\sigma+t)\wedge u_i}\bigr) - \sum_i \xi_i\bigl(Z_{\sigma\wedge u_{i+1}} - Z_{\sigma\wedge u_i}\bigr)\\ &= (H_n \cdot Z)_{\sigma+t} - (H_n \cdot Z)_\sigma.\end{aligned}$$

Next we take an arbitrary process $H \in \mathcal{L}_2(Z, \mathcal{P})$, and wish to show

$$(\bar{H} \cdot \bar{Z})_t = (H \cdot Z)_{\sigma+t} - (H \cdot Z)_\sigma. \tag{5.51}$$

Take a sequence $\{H_n\}$ of simple predictable processes such that $H_n \to H$ in $\mathcal{L}_2(Z, \mathcal{P})$. Equality (5.50), along with the boundedness of σ, then implies that $\bar{H}_n \to \bar{H}$ in $\mathcal{L}_2(\bar{Z}, \bar{\mathcal{P}})$. Consequently we get the convergence $\bar{H}_n \cdot \bar{Z} \to \bar{H} \cdot \bar{Z}$ and $H_n \cdot Z \to H \cdot Z$ in probability, uniformly on compact time intervals. Identity (5.51) was verified above for simple predictable processes, hence the convergence passes it on to general processes. Boundedness of σ is used here because then the time arguments $\sigma + t$ and σ on the right side of (5.51) stay bounded. Since $(H \cdot Z)_\sigma = \big((\mathbf{1}_{[0,\sigma]}H) \cdot Z\big)_{\sigma+t}$ we can rewrite (5.51) as

$$(\bar{H} \cdot \bar{Z})_t = \big((\mathbf{1}_{(\sigma,\infty)}H) \cdot Z\big)_{\sigma+t}. \tag{5.52}$$

At this point we have proved the lemma in the L^2 case and a localization argument remains. Given t, pick $\nu_k > t$. Then $\tau_k > \sigma + t$. Use the fact that $\{\nu_k\}$ and $\{\tau_k\}$ are localizing sequences for their respective integrals.

$$\begin{aligned}(\bar{G} \cdot \bar{M})_t &= \big((\mathbf{1}_{(0,\nu_k]}\bar{G}) \cdot \bar{M}^{\nu_k}\big)_t \\ &= \big((\mathbf{1}_{(\sigma,\infty)}\mathbf{1}_{[0,\tau_k]}G) \cdot M^{\tau_k}\big)_{\sigma+t} \\ &= \big((\mathbf{1}_{(\sigma,\infty)}G) \cdot M\big)_{\sigma+t} \\ &= (G \cdot M)_{\sigma+t} - (G \cdot M)_\sigma.\end{aligned}$$

This completes the proof of the lemma. □

Proof of Theorem 5.45. $\bar{Y}$ is a semimartingale because by Lemma 5.47 the restarting operation preserves the local L^2-martingale part, while the FV part is preserved by a direct argument. If Y were an FV process, the integral identity (5.48) would be evident as a path-by-path identity. And again Lemma 5.47 takes care of the local L^2-martingale part of the integral, so (5.48) is proved for a semimartingale Y. □

5.4.4. Stopping just before a stopping time. Let τ be a stopping time and Y a cadlag process. The process $Y^{\tau-}$ is defined by

$$Y^{\tau-}(t) = \begin{cases} Y(0), & t = 0 \text{ or } \tau = 0 \\ Y(t), & 0 < t < \tau \\ Y(\tau-), & 0 < \tau \le t. \end{cases} \tag{5.53}$$

In other words, the process Y has been stopped just prior to the stopping time. This type of stopping is useful for processes with jumps. For example, if

$$\tau = \inf\{t \ge 0 : |Y(t)| \ge r \text{ or } |Y(t-)| \ge r\}$$

then $|Y^\tau| \le r$ may fail if Y jumped exactly at time τ, but $|Y^{\tau-}| \le r$ is true.

For continuous processes $Y^{\tau-}$ and Y^τ coincide. More precisely, the relation between the two stoppings is that

$$Y^\tau(t) = Y^{\tau-}(t) + \Delta Y(\tau)\mathbf{1}\{t \ge \tau\}.$$

In other words, only a jump of Y at τ can produce a difference, and that is not felt until t reaches τ.

The next example shows that stopping just before τ can fail to preserve the martingale property. But it does preserve a semimartingale, because a single jump can be moved to the FV part, as evidenced in the proof of the lemma after the example.

Example 5.48. Let N be a rate α Poisson process and $M_t = N_t - \alpha t$. Let τ be the time of the first jump of N. Then $N^{\tau-} = 0$, from which $M_t^{\tau-} = -\alpha t \mathbf{1}_{\{\tau > t\}} - \alpha\tau\mathbf{1}_{\{\tau \le t\}}$. This process cannot be a martingale because $M_0^{\tau-} = 0$ while $M_t^{\tau-} < 0$ for all t. One can also check that the expectation of $M_t^{\tau-}$ is not constant in t, which violates martingale behavior.

Lemma 5.49. *Let Y be a semimartingale and τ a stopping time. Then $Y^{\tau-}$ is a semimartingale.*

Proof. Let $Y = Y_0 + M + U$ be a decomposition of Y into a local L^2-martingale M and an FV process U. Then

$$Y_t^{\tau-} = Y_0 + M_t^{\tau} + U_t^{\tau-} - \Delta M_\tau \mathbf{1}\{t \ge \tau\}. \tag{5.54}$$

M^τ is a local L^2-martingale, and the remaining part

$$G_t = U_t^{\tau-} - \Delta M_\tau \mathbf{1}\{t \ge \tau\}$$

is an FV process. □

Next we state some properties of integrals stopped just before τ.

Proposition 5.50. *Let Y and Z be semimartingales, G and J predictable processes locally bounded in the sense (5.36), and τ a stopping time.*

(a) $(G \cdot Y)^{\tau-} = (\mathbf{1}_{[0,\tau]}G) \cdot (Y^{\tau-}) = G \cdot (Y^{\tau-})$.

(b) *If $G = J$ on $[0, \tau]$ and $Y = Z$ on $[0, \tau)$, then $(G \cdot Y)^{\tau-} = (J \cdot Z)^{\tau-}$.*

Proof. It suffices to check (a), as (b) is an immediate consequence. Using the semimartingale decomposition (5.54) for $Y^{\tau-}$ we have

$$\begin{aligned}(\mathbf{1}_{[0,\tau]}G) \cdot Y^{\tau-} &= (\mathbf{1}_{[0,\tau]}G) \cdot M^\tau + (\mathbf{1}_{[0,\tau]}G) \cdot U^{\tau-} - G(\tau)\Delta M_\tau \mathbf{1}_{[\tau,\infty)} \\ &= (G \cdot M)^\tau + (G \cdot U)^{\tau-} - G(\tau)\Delta M_\tau \mathbf{1}_{[\tau,\infty)},\end{aligned}$$

and the same conclusion also without the factor $\mathbf{1}_{[0,\tau]}$:

$$\begin{aligned}G \cdot Y^{\tau-} &= G \cdot M^\tau + G \cdot U^{\tau-} - G(\tau)\Delta M_\tau \mathbf{1}_{[\tau,\infty)} \\ &= (G \cdot M)^\tau + (G \cdot U)^{\tau-} - G(\tau)\Delta M_\tau \mathbf{1}_{[\tau,\infty)}.\end{aligned}$$

In the steps above, the equalities

$$(\mathbf{1}_{[0,\tau]}G) \cdot M^\tau = G \cdot M^\tau = (G \cdot M)^\tau.$$

were obtained from (5.32). The equality

$$(\mathbf{1}_{[0,\tau]}G)\cdot U^{\tau-} = G\cdot U^{\tau-} = (G\cdot U)^{\tau-}$$

comes from path by path Lebesgue-Stieltjes integration: since $U^{\tau-}$ froze just prior to τ, its Lebesgue-Stieltjes measure satisfies $\Lambda_{U^{\tau-}}(A)=0$ for any Borel set $A\subseteq[\tau,\infty)$, and so for any bounded Borel function g and any $t\ge\tau$,

$$\int_{[0,t]} g\,d\Lambda_{U^{\tau-}} = \int_{[0,\tau)} g\,d\Lambda_{U^{\tau-}}.$$

Lastly, note that by Theorem 5.41

$$\begin{aligned}(G\cdot M)^{\tau}_t - G(\tau)\Delta M_\tau \mathbf{1}_{[\tau,\infty)}(t) &= (G\cdot M)^{\tau}_t - \Delta(G\cdot M)_\tau \mathbf{1}_{[\tau,\infty)}(t)\\ &= (G\cdot M)^{\tau-}_t.\end{aligned}$$

Applying this change above gives

$$(\mathbf{1}_{[0,\tau]}G)\cdot Y^{\tau-} = (G\cdot M)^{\tau-} + (G\cdot U)^{\tau-} = (G\cdot Y)^{\tau-}$$

and completes the proof. □

5.5. Integrator with absolutely continuous Doléans measure

This section stands somewhat apart from the main development, and can be skipped.

In Chapter 4 we saw that any measurable, adapted process can be integrated with respect to Brownian motion, provided the process is locally in $\mathcal{L}_2$. On the other hand, the integral of Sections 5.1 and 5.2 is restricted to predictable processes. How does the more general Brownian integral fit in the general theory? Do we need to develop separately a more general integral for other individual processes besides Brownian motion?

In this section we partially settle the issue by showing that when the Doléans measure μ_M of a local L^2-martingale M is absolutely continuous with respect to $m\otimes P$, then all measurable adapted processes can be integrated (subject to the localization requirement). In particular, this applies to Brownian motion to give all the integrals defined in Section 4, and also applies to the compensated Poisson process $M_t = N_t - \alpha t$.

The extension is somewhat illusory though. We do not obtain genuinely new integrals. Instead we show that every measurable adapted process X is equivalent, in a sense to be made precise below, to a predictable process $\bar{X}$. Then we define $X\cdot M = \bar{X}\cdot M$. This will be consistent with our earlier definitions because it was already the case that μ_M-equivalent processes have equal stochastic integrals.

For the duration of this section, fix a local L^2-martingale M with Doléans measure μ_M. The Doléans measure of a local L^2-martingale is defined exactly as for L^2-martingales, see Remark 5.33. We assume that μ_M is absolutely continuous with respect to $m \otimes P$ on the predictable σ-algebra $\mathcal{P}$. Recall that absolute continuity, abbreviated $\mu_M \ll m \otimes P$, means that if $m \otimes P(D) = 0$ for some $D \in \mathcal{P}$, then $\mu_M(D) = 0$.

Let $\mathcal{P}^*$ be the σ-algebra generated by the predictable σ-algebra $\mathcal{P}$ and all sets $D \in \mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$ with $m \otimes P(D) = 0$. Equivalently, as checked in Exercise 1.8(e),

$$\begin{aligned} \mathcal{P}^* = \{G \in \mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F} : &\text{ there exists } A \in \mathcal{P} \\ &\text{such that } m \otimes P(G \triangle A) = 0\}. \end{aligned} \tag{5.55}$$

Definition 5.51. Suppose μ_M is absolutely continuous with respect to $m \otimes P$ on the predictable σ-algebra $\mathcal{P}$. By the Radon-Nikodym Theorem, there exists a $\mathcal{P}$-measurable function $f_M \ge 0$ on $\mathbf{R}_+ \times \Omega$ such that

$$\mu_M(A) = \int_A f_M \, d(m \otimes P) \quad \text{for } A \in \mathcal{P}. \tag{5.56}$$

Define a measure μ_M^* on the σ-algebra $\mathcal{P}^*$ by

$$\mu_M^*(G) = \int_G f_M \, d(m \otimes P) \quad \text{for } G \in \mathcal{P}^*. \tag{5.57}$$

The measure μ_M^* is an extension of μ_M from $\mathcal{P}$ to the larger σ-algebra $\mathcal{P}^*$.

Firstly, definition (5.57) makes sense because f_M and G are $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$-measurable, so they can be integrated against the product measure $m \otimes P$. Second, if $G \in \mathcal{P}$, comparison of (5.56) and (5.57) shows $\mu_M^*(G) = \mu_M(G)$, so μ_M^* is an extension of μ_M.

Note also the following. Suppose G and A are in the relationship (5.55) that characterizes $\mathcal{P}^*$. Then $\mu_M^*(G) = \mu_M(A)$. To see this, write

$$\begin{aligned} \mu_M^*(G) &= \mu_M^*(A) + \mu_M^*(G \setminus A) - \mu_M^*(A \setminus G) \\ &= \mu_M(A) + \int_{G \setminus A} f_M \, d(m \otimes P) - \int_{A \setminus G} f_M \, d(m \otimes P) \\ &= \mu_M(A) \end{aligned}$$

where the last equality follows from

$$m \otimes P(G \setminus A) + m \otimes P(A \setminus G) = m \otimes P(G \triangle A) = 0.$$

The key facts that underlie the extension of the stochastic integral are assembled in the next lemma.

Lemma 5.52. *Let X be an adapted, measurable process. Then there exists a $\mathcal{P}$-measurable process $\bar{X}$ such that*

$$m \otimes P\{(t,\omega) \in \mathbf{R}_+ \times \Omega : X(t,\omega) \neq \bar{X}(t,\omega)\} = 0. \tag{5.58}$$

In particular, all measurable adapted processes are $\mathcal{P}^*$-measurable.

Under the assumption $\mu_M \ll m \otimes P$, we also have

$$\mu_M^*\{(t,\omega) \in \mathbf{R}_+ \times \Omega : X(t,\omega) \neq \bar{X}(t,\omega)\} = 0. \tag{5.59}$$

Following our earlier conventions, we say that X and $\bar{X}$ are μ_M^*-equivalent.

Proof. Let X be a bounded, adapted, measurable process. By Lemma 4.2 there exists a sequence of simple predictable processes X_n such that

$$E\int_{[0,T]\times\Omega} |X - X_n|^2 \, ds \to 0 \tag{5.60}$$

for all $T < \infty$.

We claim that there exists a subsequence $\{X_{n_k}\}$ such that $X_{n_k}(t,\omega) \to X(t,\omega)$ $m \otimes P$-almost everywhere on $\mathbf{R}_+ \times \Omega$. We perform this construction with the usual diagonal argument. In general, L^2 convergence implies almost everywhere convergence along some subsequence. Thus from (5.60) for $T = 1$ we can extract a subsequence $\{X_{n_j^1} : j \in \mathbf{N}\}$ such that $X_{n_j^1} \to X$ $m \otimes P$-almost everywhere on the set $[0,1] \times \Omega$. Inductively, suppose we have a subsequence $\{X_{n_j^\ell} : j \in \mathbf{N}\}$ such that $\lim_{j\to\infty} X_{n_j^\ell} = X$ $m \otimes P$-almost everywhere on the set $[0,\ell] \times \Omega$. Then apply (5.60) for $T = \ell + 1$ to extract a subsequence $\{n_j^{\ell+1} : j \in \mathbf{N}\}$ of $\{n_j^\ell : j \in \mathbf{N}\}$ such that $\lim_{j\to\infty} X_{n_j^{\ell+1}} = X$ $m \otimes P$-almost everywhere on the set $[0, \ell+1] \times \Omega$.

From the array $\{n_j^\ell : \ell, j \in \mathbf{N}\}$ thus constructed, take the diagonal $n_k = n_k^k$ for $k \in \mathbf{N}$. For any ℓ, $\{n_k : k \geq \ell\}$ is a subsequence of $\{n_j^\ell : j \in \mathbf{N}\}$, and consequently $X_{n_k} \to X$ $m \otimes P$-almost everywhere on the set $[0,\ell] \times \Omega$. Let

$$A = \{(t,\omega) \in \mathbf{R}_+ \times \Omega : \; X_{n_k}(t,\omega) \to X(t,\omega) \text{ as } k \to \infty\}.$$

The last observation on $\{n_k\}$ implies that

$$A^c = \bigcup_{\ell=1}^{\infty} \{(t,\omega) \in [0,\ell] \times \Omega : \; X_{n_k}(t,\omega) \text{ does not converge to } X(t,\omega)\}$$

is a countable union of sets of $m \otimes P$-measure zero. Thus $X_{n_k} \to X$ $m \otimes P$-almost everywhere on $\mathbf{R}_+ \times \Omega$.

Set

$$\bar{X}(t,\omega) = \limsup_{k\to\infty} X_{n_k}(t,\omega).$$

The processes X_{n_k} are $\mathcal{P}$-measurable by construction. Consequently $\bar{X}$ is $\mathcal{P}$-measurable. On the set A, $\bar{X} = X$. Since $m \otimes P(A^c) = 0$, (5.58) follows for this X.

For a Borel set $B \in \mathcal{B}_{\mathbf{R}}$,

$$\begin{aligned}\{(t,\omega) \in \mathbf{R}_+ \times \Omega : X(t,\omega) \in B\} \\ &= \{(t,\omega) \in A : X(t,\omega) \in B\} \cup \{(t,\omega) \in A^c : X(t,\omega) \in B\} \\ &= \{(t,\omega) \in A : \bar{X}(t,\omega) \in B\} \cup \{(t,\omega) \in A^c : X(t,\omega) \in B\}.\end{aligned}$$

This expresses $\{X \in B\}$ as a union of a set in $\mathcal{P}$ and a set in $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$ with $m \otimes P$-measure zero. So $\{X \in B\} \in \mathcal{P}^*$.

The lemma has now been proved for a bounded adapted measurable process X.

Given an arbitrary adapted measurable process X, let $X^{(k)} = (X \wedge k) \vee (-k)$. By the previous part, $X^{(k)}$ is $\mathcal{P}^*$-measurable, and there exist $\mathcal{P}$-measurable processes $\bar{X}_k$ such that $X^{(k)} = \bar{X}_k$ $m \otimes P$-almost everywhere. Define (again) $\bar{X} = \limsup_{k\to\infty} \bar{X}_k$. Since $X^{(k)} \to X$ pointwise, X is also $\mathcal{P}^*$-measurable, and $X = \bar{X}$ $m \otimes P$-almost everywhere. □

A consequence of the lemma is that it makes sense to talk about the μ_M^*-measure of any event involving measurable, adapted processes.

Definition 5.53. Assume M is a local L^2-martingale whose Doléans measure μ_M satisfies $\mu_M \ll m \otimes P$ on $\mathcal{P}$. Define the extension μ_M^* of μ_M to $\mathcal{P}^*$ as in Definition 5.51.

Let $\mathcal{L}(M, \mathcal{P}^*)$ be the class of measurable adapted processes X for which there exists a nondecreasing sequence of stopping times ρ_k such that $\rho_k \nearrow \infty$ almost surely, and for each k,

$$\int_{[0,T]\times\Omega} \mathbf{1}_{[0,\rho_k]} |X|^2 \, d\mu_M^* < \infty \quad \text{for all } T < \infty.$$

For $X \in \mathcal{L}(M, \mathcal{P}^*)$, the stochastic integral $X \cdot M$ is the local L^2-martingale given by $\bar{X} \cdot M$, where $\bar{X}$ is the $\mathcal{P}$-measurable process that is μ_M^*-equivalent to X in the sense (5.59). This $\bar{X}$ will lie in $\mathcal{L}(M, \mathcal{P})$ and so the process $\bar{X} \cdot M$ exists. The process $X \cdot M$ thus defined is unique up to indistinguishability.

Justification of the definition. Since $\bar{X} = X$ μ_M^*-almost everywhere, their μ_M^*-integrals agree, and in particular

$$\int_{[0,T]\times\Omega} \mathbf{1}_{[0,\rho_k]} |\bar{X}|^2 \, d\mu_M = \int_{[0,T]\times\Omega} \mathbf{1}_{[0,\rho_k]} |X|^2 \, d\mu_M^* < \infty \quad \text{for all } T < \infty.$$

By Lemma 5.34, this is just another way of expressing $\bar{X} \in \mathcal{L}(M, \mathcal{P})$. It follows that the integral $\bar{X} \cdot M$ exists and is a member of $\mathcal{M}_{2,\mathrm{loc}}$. In particular, we can define $X \cdot M = \bar{X} \cdot M$ as an element of $\mathcal{M}_{2,\mathrm{loc}}$.

If we choose another $\mathcal{P}$-measurable process $\bar{Y}$ that is μ_M^*-equivalent to X, then $\bar{X}$ and $\bar{Y}$ are μ_M-equivalent, and the integrals $\bar{X} \cdot M$ and $\bar{Y} \cdot M$ are indistinguishable by Exercise 5.10. □

If M is an L^2-martingale to begin with, such as Brownian motion or the compensated Poisson process, we naturally define $\mathcal{L}_2(M, \mathcal{P}^*)$ as the space of measurable adapted processes X such that

$$\int_{[0,T]\times\Omega} |X|^2 \, d\mu_M^* < \infty \quad \text{for all } T < \infty.$$

The integral extended to $\mathcal{L}(M, \mathcal{P}^*)$ or $\mathcal{L}_2(M, \mathcal{P}^*)$ enjoys all the properties derived before, because the class of processes that appear as stochastic integrals has not been expanded.

A technical point worth noting is that once the integrand is not predictable, the stochastic integral does not necessarily coincide with the Lebesgue-Stieltjes integral even if the integrator has paths of bounded variation. This is perfectly illustrated by the Poisson process in Exercise 5.12. While this may seem unnatural, we must remember that the theory of Itô stochastic integrals is so successful because integrals are martingales.

5.6. Quadratic variation

This final section of the chapter extends our earlier treatment of quadratic variation and covariation, partly as preparation for Itô's formula. The section contains three main points. In Theorem 5.63 we reach the formula for quadratic covariation of stochastic integrals: $[\int G\,dY, \int H\,dZ] = \int GH\,d[Y,Z]$. Corollary 5.60 gives the substitution rule $G \cdot (H \cdot Y) = (GH) \cdot Y$. A stochastic integration by parts formula which is a special case of Itô's formula arrives in Theorem 5.61. As a consequence in Proposition 5.64 we prove a Riemann sum approximation for $d[Y,Z]$-integrals that will be used in the proof of Itô's formula. The reader who wishes to get quickly to the proof of Itô's formula can glance at Proposition 5.64 and move on to Section 6.1.

Recall from Section 2.2 that the quadratic variation $[X]$ of a process X, when it exists, is by definition a nondecreasing process with $[X]_0 = 0$ whose value at time t is determined, up to a null event, by the limit in probability

$$[X]_t = \lim_{\mathrm{mesh}(\pi)\to 0} \sum_i (X_{t_{i+1}} - X_{t_i})^2. \tag{5.61}$$

Here $\pi = \{0 = t_0 < t_1 < \cdots < t_{m(\pi)} = t\}$ is a partition of $[0,t]$. Quadratic covariation $[X,Y]$ of two processes X and Y was defined as the FV process

$$[X,Y] = [(X+Y)/2] - [(X-Y)/2], \tag{5.62}$$

assuming the processes on the right exist. Again there is a limit in probability:

$$\lim_{\text{mesh}(\pi)\to 0} \sum_i (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) = [X,Y]_t. \tag{5.63}$$

Identity $[X] = [X,X]$ holds. For cadlag semimartingales X and Y, $[X]$ and $[X,Y]$ exist and have cadlag versions (Corollary 3.32).

Limit (5.63) shows that, for processes X, Y and Z and reals α and β,

$$[\alpha X + \beta Y, Z] = \alpha[X,Z] + \beta[Y,Z] \tag{5.64}$$

provided these processes exist. The equality can be taken in the sense of indistinguishability for cadlag versions, provided such exist. Consequently $[\,\cdot\,,\,\cdot\,]$ operates somewhat in the manner of an inner product.

Lemma 5.54. *Suppose M_n, M, N_n and N are L^2-martingales. Fix $0 \le T < \infty$. Assume $M_n(T) \to M(T)$ and $N_n(T) \to N(T)$ in L^2 as $n \to \infty$. Then*

$$E\Big\{ \sup_{0\le t\le T} \big|\, [M_n,N_n]_t - [M,N]_t \,\big| \Big\} \to 0 \quad \text{as } n \to \infty.$$

Proof. It suffices to consider the case where $M_n = N_n$ and $M = N$ because the general case then follows from (5.62). Apply inequality (2.20) and note that $[X]_t$ is nondecreasing in t.

$$\begin{aligned} \big|\, [M_n]_t - [M]_t \,\big| &\le [M_n - M]_t + 2[M_n - M]_t^{1/2}[M]_t^{1/2} \\ &\le [M_n - M]_T + 2[M_n - M]_T^{1/2}[M]_T^{1/2}. \end{aligned}$$

Take expectations, apply Schwarz and recall (3.15).

$$\begin{aligned} E\Big[\sup_{0\le t\le T} \big|\, [M_n]_t - [M]_t \,\big| \Big] &\le \|M_n(T) - M(T)\|^2_{L^2(P)} \\ &\quad +\ 2\|M_n(T) - M(T)\|_{L^2(P)} \cdot \|M(T)\|_{L^2(P)}. \end{aligned}$$

Letting $n \to \infty$ completes the proof. □

Proposition 5.55. *Let $M, N \in \mathcal{M}_{2,\text{loc}}$, $G \in \mathcal{L}(M,\mathcal{P})$, and $H \in \mathcal{L}(N,\mathcal{P})$. Then*

$$[G\cdot M, H\cdot N]_t = \int_{(0,t]} G_s H_s \, d[M,N]_s.$$

Proof. It suffices to show

$$[G \cdot M, L]_t = \int_{(0,t]} G_s \, d[M, L]_s \tag{5.65}$$

for an arbitrary $L \in \mathcal{M}_{2,\text{loc}}$. This is because then (5.65) applied to $L = H \cdot N$ gives

$$[M, L]_t = \int_{(0,t]} H_s \, d[M, N]_s$$

so the Lebesgue-Stieltjes measures satisfy $d[M, L]_t = H_t \, d[M, N]_t$. Substitute this back into (5.65) to get the desired equality

$$[G \cdot M, L]_t = \int_{(0,t]} G_s \, d[M, L]_s = \int_{(0,t]} G_s H_s \, d[M, N]_s.$$

Step 1. Assume $L, M \in \mathcal{M}_2$. First consider $G = \xi \mathbf{1}_{(u,v]}$ for a bounded $\mathcal{F}_u$-measurable random variable ξ. Then $G \cdot M = \xi(M^v - M^u)$. By the bilinearity (5.64) of quadratic covariation,

$$\begin{aligned}[G \cdot M, L]_t &= \xi\big([M^v, L]_t - [M^u, L]_t\big) = \xi\big([M, L]_{v \wedge t} - [M, L]_{u \wedge t}\big) \\ &= \int_{(0,t]} \xi \mathbf{1}_{(u,v]}(s) \, d[M, L]_s = \int_{(0,t]} G_s \, d[M, L]_s.\end{aligned}$$

The second equality above used Lemma 3.30. ξ moves freely in and out of the integrals because they are path-by-path Lebesgue-Stieltjes integrals. By additivity of the covariation conclusion (5.65) follows for G that are simple predictable processes of the type (5.6).

Now take a general $G \in \mathcal{L}_2(M, \mathcal{P})$. Pick simple predictable processes G_n such that $G_n \to G$ in $\mathcal{L}_2(M, \mathcal{P})$. Then $(G_n \cdot M)_t \to (G \cdot M)_t$ in $L^2(P)$. By Lemma 5.54 $[G_n \cdot M, L]_t \to [G \cdot M, L]_t$ in $L^1(P)$. On the other hand the previous lines showed

$$[G_n \cdot M, L]_t = \int_{(0,t]} G_n(s) \, d[M, L]_s.$$

The desired equality

$$[G \cdot M, L]_t = \int_{(0,t]} G(s) \, d[M, L]_s$$

follows if we can show the $L^1(P)$ convergence

$$\int_{(0,t]} G_n(s) \, d[M, L]_s \longrightarrow \int_{(0,t]} G(s) \, d[M, L]_s.$$

The first step below is a combination of the Kunita-Watanabe inequality (2.22) and the Schwarz inequality. Next, the assumption $G_n \to G$ in

$\mathcal{L}_2(M,\mathcal{P})$ implies $G_n \to G$ in $L^2([0,t]\times\Omega,\mu_M)$.

$$\begin{aligned}
&E\Big\{\Big|\int_{(0,t]}\big(G_n(s)-G(s)\big)\,d[M,L]_s\Big|\Big\}\\
&\quad\le\Big(E\int_{(0,t]}|G_n(s)-G(s)|^2\,d[M]_s\Big)^{1/2}\big(E\{[L]_t\}\big)^{1/2}\\
&\quad=\Big(\int_{(0,t]\times\Omega}|G_n-G|^2\,d\mu_M\Big)^{1/2}\big(E\{L_t^2-L_0^2\}\big)^{1/2}\longrightarrow 0\quad\text{as } n\to\infty.
\end{aligned}$$

We have shown (5.65) for the case $L,M\in\mathcal{M}_2$ and $G\in\mathcal{L}_2(M,\mathcal{P})$.

Step 2. Now the case $L,M\in\mathcal{M}_{2,\mathrm{loc}}$ and $G\in\mathcal{L}(M,\mathcal{P})$. Pick stopping times $\{\tau_k\}$ that localize both L and (G,M). Abbreviate $G^k=\mathbf{1}_{(0,\tau_k]}G$. Then if $\tau_k(\omega)\ge t$,

$$\begin{aligned}
[G\cdot M,L]_t&=[G\cdot M,L]_{\tau_k\wedge t}=[(G\cdot M)^{\tau_k},L^{\tau_k}]_t=[(G^k\cdot M^{\tau_k}),L^{\tau_k}]_t\\
&=\int_{(0,t]}G_s^k\,d[M^{\tau_k},L^{\tau_k}]_s=\int_{(0,t]}G_s\,d[M,L]_s.
\end{aligned}$$

We used Lemma 3.30 to move the τ_k superscript in and out of covariations, and the definition (5.27) of the integral $G\cdot M$. Since $\tau_k(\omega)\ge t$, we have $G^k=G$ on $[0,t]$, and consequently G^k can be replaced by G on the last line. This completes the proof. □

We can complement part (c) of Proposition 5.14 with this result.

Corollary 5.56. *Suppose $M,N\in\mathcal{M}_2$, $G\in\mathcal{L}_2(M,\mathcal{P})$, and $H\in\mathcal{L}_2(N,\mathcal{P})$. Then*

$$(G\cdot M)_t(H\cdot N)_t-\int_{(0,t]}G_sH_s\,d[M,N]_s$$

is a martingale. If we weaken the hypotheses to $M,N\in\mathcal{M}_{2,\mathrm{loc}}$, $G\in\mathcal{L}(M,\mathcal{P})$, and $H\in\mathcal{L}(N,\mathcal{P})$, then the process above is a local martingale.

Proof. Follows from Proposition 5.55 above, and a general property of $[M,N]$ for (local) L^2-martingales M and N, stated as Theorem 3.31 in Section 3.4. □

Since quadratic variation functions like an inner product for local L^2-martingales, we can use the previous result as a characterization of the stochastic integral. According to this characterization, to verify that a given process is the stochastic integral, it suffices to check that it behaves the right way in the quadratic covariation.

Lemma 5.57. *Let $M \in \mathcal{M}_{2,\text{loc}}$ and assume $M_0 = 0$.*

(a) *For any $0 \le t < \infty$, $[M]_t = 0$ almost surely iff $\sup_{0\le s\le t}|M_s| = 0$ almost surely.*

(b) *Let $G \in \mathcal{L}(M, \mathcal{P})$. The stochastic integral $G\cdot M$ is the unique process $Y \in \mathcal{M}_{2,\text{loc}}$ that satisfies $Y_0 = 0$ and*

$$[Y, L] = \int_{(0,t]} G_s \, d[M, L]_s \quad \textit{almost surely}$$

for each $0 \le t < \infty$ and each $L \in \mathcal{M}_{2,\text{loc}}$.

Proof. Part (a). Fix t. Let $\{\tau_k\}$ be a localizing sequence for M. Then for $s \le t$, $[M^{\tau_k}]_s \le [M^{\tau_k}]_t = [M]_{\tau_k\wedge t} \le [M]_t = 0$ almost surely by Lemma 3.28 and the t-monotonicity of $[M]$. Consequently $E\{(M_s^{\tau_k})^2\} = E\{[M^{\tau_k}]_s\} = 0$, from which $M_s^{\tau_k} = 0$ almost surely. Taking k large enough so that $\tau_k(\omega) \ge s$, $M_s = M_s^{\tau_k} = 0$ almost surely. Since M has cadlag paths, there is a single event Ω_0 such that $P(\Omega_0) = 1$ and $M_s(\omega) = 0$ for all $\omega \in \Omega_0$ and $s \in [0, t]$.

Conversely, if M vanishes on $[0, t]$ then so does $[M]$ by its definition.

Part (b). We checked that $G \cdot M$ satisfies (5.65) which is the property required here. Conversely, suppose $Y \in \mathcal{M}_{2,\text{loc}}$ satisfies the property. Then by the additivity,

$$[Y - G \cdot M, L]_t = [Y, L] - [G \cdot M, L] = 0$$

for any $L \in \mathcal{M}_{2,\text{loc}}$. Taking $L = Y - G\cdot M$ gives $[Y - G\cdot M, Y - G\cdot M] = 0$, and then by part (a) $Y = G \cdot M$. □

Remark 5.58. Part (a) of Lemma 5.57 cannot hold for semimartingales. Any continuous BV function will have a vanishing quadratic variation.

The next change-of-integrator or substitution property could have been proved in Chapter 5 with a case-by-case argument, from simple predictable integrands to localized integrals. At this point we can give a quick proof, since we already did the tedious work in the previous proofs.

Proposition 5.59. *Let $M \in \mathcal{M}_{2,\text{loc}}$, $G \in \mathcal{L}(M, \mathcal{P})$, and let $N = G \cdot M$, also a member of $\mathcal{M}_{2,\text{loc}}$. Suppose $H \in \mathcal{L}(N, \mathcal{P})$. Then $HG \in \mathcal{L}(M, \mathcal{P})$ and $H \cdot N = (HG) \cdot M$.*

Proof. Let $\{\tau_k\}$ be a localizing sequence for (G, M) and (H, N). By part (b) of Proposition 5.31 $N^{\tau_k} = (G \cdot M)^{\tau_k} = G \cdot M^{\tau_k}$, and so Proposition 5.55 gives the equality of Lebesgue-Stieltjes measures $d[N^{\tau_k}]_s = G_s^2 \, d[M^{\tau_k}]_s$. Then for any $T < \infty$,

$$E \int_{[0,T]} \mathbf{1}_{[0,\tau_k]}(t) H_t^2 G_t^2 \, d[M^{\tau_k}]_t = E \int_{[0,T]} \mathbf{1}_{[0,\tau_k]}(t) H_t^2 \, d[N^{\tau_k}]_t < \infty$$

because $\{\tau_k\}$ is assumed to localize (H, N). This checks that $\{\tau_k\}$ localizes (HG, M) and so $HG \in \mathcal{L}(M, \mathcal{P})$.

Let $L \in \mathcal{M}_{2,\text{loc}}$. Equation (5.65) gives $G_s\, d[M, L]_s = d[N, L]_s$, and so

$$[(HG) \cdot M, L]_t = \int_{(0,t]} H_s G_s\, d[M, L]_s = \int_{(0,t]} H_s\, d[N, L]_s = [H \cdot N, L]_t.$$

By Lemma 5.57(b), $(HG) \cdot M$ must coincide with $H \cdot N$. ☐

We proceed to extend some of these results to semimartingales, beginning with the change of integrator. Recall some notation and terminology. The integral $\int G\, dY$ with respect to a cadlag semimartingale Y was defined in Section 5.3 for predictable integrands G that satisfy this condition:

(5.66) there exist stopping times $\{\sigma_n\}$ such that $\sigma_n \nearrow \infty$ almost surely and $\mathbf{1}_{(0,\sigma_n]} G$ is a bounded process for each n.

If X is a cadlag process, we defined the caglad process X_- by $X_-(0) = X(0)$ and $X_-(t) = X(t-)$ for $t > 0$. The notion of uniform convergence on compact time intervals in probability first appeared in our discussion of martingales (Lemma 3.42) and then with integrals (Propositions 5.32 and 5.37).

Corollary 5.60. *Let Y be a cadlag semimartingale, G and H predictable processes that satisfy* (5.66), *and $X = \int H\, dY$, also a cadlag semimartingale. Then*

$$\int G\, dX = \int GH\, dY.$$

Proof. Let $Y = Y_0 + M + U$ be a decomposition of Y into a local L^2-martingale M and an FV process U. Let $V_t = \int_{(0,t]} H_s\, dU_s$, another FV process. This definition entails that, for a fixed ω, H_s is the Radon-Nikodym derivative $d\Lambda_V / d\Lambda_U$ of the Lebesgue-Stieltjes measures on the time line (Lemma 1.15). $X = H \cdot M + V$ is a semimartingale decomposition of X. By definition of the integral $\int G\, dX$ and Proposition 5.59,

$$\begin{aligned}
\int_{(0,t]} G_s\, dX_s &= \big(G \cdot (H \cdot M)\big)_t + \int_{(0,t]} G_s\, dV_s \\
&= \big((GH) \cdot M\big)_t + \int_{(0,t]} G_s H_s\, dU_s \\
&= \int_{(0,t]} G_s H_s\, dY_s.
\end{aligned}$$

The last equality is the definition of the semimartingale integral $\int GH\, dY$. ☐

In terms of "stochastic differentials" we can express the conclusion as $dX = H\,dY$. These stochastic differentials do not exist as mathematical objects, but the rule $dX = H\,dY$ tells us that dX can be replaced by $H\,dY$ in the stochastic integral.

Theorem 5.61. *Let Y and Z be cadlag semimartingales. Then $[Y,Z]$ exists as a cadlag, adapted FV process and satisfies*

$$[Y,Z]_t = Y_tZ_t - Y_0Z_0 - \int_{(0,t]} Y_{s-}\,dZ_s - \int_{(0,t]} Z_{s-}\,dY_s. \tag{5.67}$$

The product YZ is a semimartingale, and for a predictable process H that satisfies (5.66),

$$\begin{aligned}\int_{(0,t]} H_s\,d(YZ)_s = \int_{(0,t]} H_sY_{s-}\,dZ_s + \int_{(0,t]} H_sZ_{s-}\,dY_s \\ + \int_{(0,t]} H_s\,d[Y,Z]_s.\end{aligned} \tag{5.68}$$

Proof. We can give here a proof of the existence $[Y,Z]$ independent of Corollary 3.32. Take a countably infinite partition $\pi = \{0 = t_1 < t_2 < t_3 < \cdots < t_i < \cdots\}$ of $\mathbf{R}_+$ such that $t_i \nearrow \infty$, and in fact take a sequence of such partitions with $\mathrm{mesh}(\pi) \to 0$. (We omit the index for the sequence of partitions.) For $t \in \mathbf{R}_+$ consider

$$\begin{aligned}S_\pi(t) &= \sum (Y_{t_{i+1}\wedge t} - Y_{t_i\wedge t})(Z_{t_{i+1}\wedge t} - Z_{t_i\wedge t}) \\ &= \Big\{ \sum (Y_{t_{i+1}\wedge t}Z_{t_{i+1}\wedge t} - Y_{t_i\wedge t}Z_{t_i\wedge t}) \\ &\qquad - \sum Y_{t_i}(Z_{t_{i+1}\wedge t} - Z_{t_i\wedge t}) - \sum Z_{t_i}(Y_{t_{i+1}\wedge t} - Y_{t_i\wedge t}) \Big\} \\ &\underset{\mathrm{mesh}(\pi)\to 0}{\longrightarrow} Y_tZ_t - Y_0Z_0 - \int_{(0,t]} Y_{s-}\,dZ_s - \int_{(0,t]} Z_{s-}\,dY_s.\end{aligned} \tag{5.69}$$

Proposition 5.37 implies that the convergence takes place in probability, uniformly on any compact time interval $[0,T]$, and also gives the limit integrals. By the usual Borel-Cantelli argument we then get the limit almost surely, uniformly on $[0,T]$, along some subsequence of the original sequence of partitions. The limit process is cadlag.

From this limit we get all the conclusions. Take first $Y = Z$. Suppose $s < t$. Once the mesh is smaller than $t - s$ we have indices $k < \ell$ such that

$t_k < s \le t_{k+1}$ and $t_\ell < t \le t_{\ell+1}$. Then

$$\begin{aligned} S_\pi(t) - S_\pi(s) &= (Y_t - Y_{t_\ell})^2 + \sum_{i=k}^{\ell-1}(Y_{t_{i+1}} - Y_{t_i})^2 - (Y_s - Y_{t_k})^2 \\ &\ge (Y_{t_{k+1}} - Y_{t_k})^2 - (Y_s - Y_{t_k})^2 \\ &\longrightarrow (\Delta Y_s)^2 - (\Delta Y_s)^2 = 0. \end{aligned}$$

Since this limit holds almost surely simultaneously for all $s < t$ in $[0, T]$, we can conclude that the limit process is nondecreasing. We have satisfied Definition 2.14 and now know that every cadlag semimartingale Y has a nondecreasing, cadlag quadratic variation $[Y]$.

Next Definition 2.15 gives the existence of a cadlag, FV process $[Y, Z]$. By the limits (2.13), $[Y, Z]$ coincides with the cadlag limit process in (5.69) almost surely at each fixed time, and consequently these processes are indistinguishable. We have verified (5.67).

We can turn (5.67) around to say

$$YZ = Y_0Z_0 + \int Y_-\,dZ + \int Z_-\,dY + [Y, Z]$$

which represents YZ as a sum of semimartingales, and thereby YZ itself is a semimartingale.

Equality (5.68) follows from the additivity of integrals, and the change-of-integrator formula applied to the semimartingales $\int Y_-\,dZ$ and $\int Z_-\,dY$. □

Remark 5.62. Identity (5.67) gives an *integration by parts* rule for stochastic integrals. It generalizes the integration by parts rule of Lebesgue-Stieltjes integrals that is part of standard real analysis [8, Section 3.5].

Next we extend Proposition 5.55 to semimartingales.

Theorem 5.63. *Let Y and Z be cadlag semimartingales and G and H predictable processes that satisfy* (5.66). *Then*

$$[G \cdot Y, H \cdot Z]_t = \int_{(0,t]} G_s H_s \, d[Y, Z]_s.$$

Proof. For the same reason as in the proof of Proposition 5.55, it suffices to show

$$[G \cdot Y, Z]_t = \int_{(0,t]} G_s \, d[Y, Z]_s.$$

Let $Y = Y_0 + M + A$ and $Z = Z_0 + N + B$ be decompositions into local L^2-martingales M and N and FV processes A and B. By the linearity of quadratic covariation,

$$[G \cdot Y, Z] = [G \cdot M, N] + [G \cdot M, B] + [G \cdot A, Z].$$

By Proposition 5.55 $[G \cdot M, N] = \int G \, d[M, N]$. To handle the other two terms, fix an ω such that the paths of the integrals and the semimartingales are cadlag, and the characterization of jumps given in Theorem 5.41 is valid for each integral that appears here. Then since B and $G \cdot A$ are FV processes, Lemma A.10 applies. Combining Lemma A.10, Theorem 5.41, and the definition of a Lebesgue-Stieltjes integral with respect to a step function gives

$$\begin{aligned}
[G \cdot M, B]_t + [G \cdot A, Z]_t &= \sum_{s \in (0,t]} \Delta(G \cdot M)_s \Delta B_s + \sum_{s \in (0,t]} \Delta(G \cdot A)_s \Delta Z_s \\
&= \sum_{s \in (0,t]} G_s \Delta M_s \Delta B_s + \sum_{s \in (0,t]} G_s \Delta A_s \Delta Z_s \\
&= \int_{(0,t]} G_s \, d[M, B]_s + \int_{(0,t]} G_s \, d[A, Z]_s.
\end{aligned}$$

Combining terms gives the desired equation. □

As the last result we generalize the limit taken in (5.69) by adding a coefficient to the sum. This is the technical lemma from this section that we need for the proof of Itô's formula.

Proposition 5.64. *Let Y and Z be cadlag semimartingales, and G an adapted cadlag process. Given a partition $\pi = \{0 = t_1 < t_2 < t_3 < \cdots < t_i \nearrow \infty\}$ of $[0, \infty)$, define*

$$R_t(\pi) = \sum_{i=1}^{\infty} G_{t_i} (Y_{t_{i+1} \wedge t} - Y_{t_i \wedge t})(Z_{t_{i+1} \wedge t} - Z_{t_i \wedge t}). \tag{5.70}$$

Then as $\text{mesh}(\pi) \to 0$, *$R(\pi)$ converges to $\int G_- \, d[Y, Z]$ in probability, uniformly on compact time intervals.*

Proof. Algebra gives

$$\begin{aligned}
R_t(\pi) = &\sum G_{t_i} (Y_{t_{i+1} \wedge t} Z_{t_{i+1} \wedge t} - Y_{t_i \wedge t} Z_{t_i \wedge t}) \\
&- \sum G_{t_i} Y_{t_i} (Z_{t_{i+1} \wedge t} - Z_{t_i \wedge t}) - \sum G_{t_i} Z_{t_i} (Y_{t_{i+1} \wedge t} - Y_{t_i \wedge t}).
\end{aligned}$$

We know from Theorem 5.61 that YZ is a semimartingale. Applying Proposition 5.37 to each sum gives the claimed type of convergence to the limit

$$\int_{(0,t]} G_{s-} d(YZ)_s - \int_{(0,t]} G_{s-} Y_{s-} \, dZ_s - \int_{(0,t]} G_{s-} Z_{s-} \, dY_s$$

which by (5.68) equals $\int_{(0,t]} G_{s-}\, d[Y,Z]_s$. $\square$

Exercises

Exercise 5.1. Recall the definition (2.6) of $\mathcal{F}_{t-}$. Show that if $X : \mathbf{R}_+ \times \Omega \to \mathbf{R}$ is $\mathcal{P}$-measurable, then $X_t(\omega) = X(t, \omega)$ is a process adapted to the filtration $\{\mathcal{F}_{t-}\}$.

Hint: Given $B \in \mathcal{B}_{\mathbf{R}}$, let $A = \{(t, \omega) : X(t, \omega) \in B\} \in \mathcal{P}$. The event $\{X_t \in B\}$ equals the t-section $A_t = \{\omega : (t, \omega) \in A\}$, so it suffices to show that an arbitrary $A \in \mathcal{P}$ satisfies $A_t \in \mathcal{F}_{t-}$ for all $t \in \mathbf{R}_+$. This follows from checking that predictable rectangles have this property, and that the collection of sets in $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{F}$ with this property form a sub-σ-field.

Exercise 5.2. (a) Show that for any Borel function $h : \mathbf{R}_+ \to \mathbf{R}$, the deterministic process $X(t, \omega) = h(t)$ is predictable. *Hint:* Intervals of the type $(a, b]$ generate the Borel σ-field of $\mathbf{R}_+$.

(b) Suppose X is an $\mathbf{R}^m$-valued predictable process and $g : \mathbf{R}^m \to \mathbf{R}^d$ a Borel function. Show that the process $Z_t = g(X_t)$ is predictable.

(c) Suppose X is an $\mathbf{R}^m$-valued predictable process and $f : \mathbf{R}_+ \times \mathbf{R}^m \to \mathbf{R}^d$ a Borel function. Show that the process $W_t = f(t, X_t)$ is predictable.

Hint. Parts (a) and (b) imply that $h(t)g(X_t)$ is predictable. Now appeal to results around the π-λ Theorem.

Exercise 5.3. Fill in the missing details of the proof that $\mathcal{P}$ is generated by continuous processes (see Lemma 5.1).

Exercise 5.4. Let N_t be a rate α Poisson process on $\mathbf{R}_+$ and $M_t = N_t - \alpha t$. In Example 5.3 we checked that $\mu_M = \alpha m \otimes P$ on $\mathcal{P}$. Here we use this result to show that the process N is not $\mathcal{P}$-measurable.

(a) Evaluate the integral

$$\int_{[0,T]\times\Omega} N_s(\omega)\,(\alpha m \otimes P)(ds, d\omega).$$

(b) Evaluate

$$E \int_{[0,T]} N_s \, d[M]_s$$

where the inner integral is the pathwise Lebsgue-Stieltjes integral, in accordance with the interpretation of definition (5.1) of μ_M. Conclude that N cannot be $\mathcal{P}$-measurable.

(c) For comparison, evaluate explicitly integrals

$$E \int_{(0,T]} N_{s-} \, dN_s \quad \text{and} \quad \int_{[0,T]\times\Omega} N_{s-}(\omega)\,(\alpha m \otimes P)(ds, d\omega).$$

Here $N_{s-}(\omega) = \lim_{u \nearrow s} N_u(\omega)$ is the left limit. Explain why we know without calculation that these integrals must agree.

Exercise 5.5. Let $\{\mathcal{F}_t\}$ be a filtration, $\mathcal{P}$ the corresponding predictable σ-field, $\{\mathcal{F}_{t+}\}$ defined as in (2.5), and $\mathcal{P}_+$ the predictable σ-field that corresponds to $\{\mathcal{F}_{t+}\}$. In other words, $\mathcal{P}_+$ is the σ-field generated by the sets $\{0\} \times F_0$ for $F_0 \in \mathcal{F}_{0+}$ and $(s,t] \times F$ for $F \in \mathcal{F}_{s+}$.

(a) Show that each $(s,t] \times F$ for $F \in \mathcal{F}_{s+}$ lies in $\mathcal{P}$.

(b) Show that a set of the type $\{0\} \times F_0$ for $F \in \mathcal{F}_{0+} \setminus \mathcal{F}_0$ cannot lie in $\mathcal{P}$.

(c) Show that the σ-fields $\mathcal{P}$ and $\mathcal{P}_+$ coincide on the subspace $(0,\infty) \times \Omega$ of $\mathbf{R}_+ \times \Omega$. *Hint:* Apply part (d) of Exercise 1.8.

(d) Suppose X is a $\mathcal{P}_+$-measurable process. Show that the process $Y(t,\omega) = X(t,\omega)\mathbf{1}_{(0,\infty)}(t)$ is $\mathcal{P}$-measurable.

Exercise 5.6. Suppose M is a continuous L^2-martingale and $X \in \mathcal{L}_2(M,\mathcal{P})$. Show that $(\mathbf{1}_{(a,b)}X) \cdot M = (\mathbf{1}_{[a,b]}X) \cdot M$. *Hint:* $[M]$ is continuous.

Exercise 5.7. Let X be a predictable process. Show that

$$\int_0^T X(s,\omega)^2\,ds < \infty \quad \text{for all } T < \infty, \text{ for } P\text{-almost every } \omega,$$

if and only if there exist stopping times $\tau_k \nearrow \infty$ such that

$$E\int_0^{\tau_k(\omega)\wedge T} X(s,\omega)^2\,ds < \infty \quad \text{for all } T < \infty.$$

From this argue the characterization of $\mathcal{L}(M,\mathcal{P})$ for Brownian motion and the compensated Poisson process claimed in Example 5.26.

Exercise 5.8. Finish the proof of Proposition 5.31.

Exercise 5.9. Let f be a cadlag function on $\mathbf{R}_+$ and define f_- by $f_-(0) = f(0)$ and $f_-(t) = f(t-)$ for $t > 0$. Show that f_- is a caglad function. Observe in passing that the right limits of f_- recover f.

Exercise 5.10. Let M be a local L^2 martingale. Show that if $X, Y \in \mathcal{L}(M,\mathcal{P})$ are μ_M-equivalent, namely

$$\mu_M\{(t,\omega) : X(t,\omega) \neq Y(t,\omega)\} = 0,$$

then $X \cdot M$ and $Y \cdot M$ are indistinguishable.

Exercise 5.11. (Computations with the Poisson process.) Let N be a homogeneous rate α Poisson process, and $M_t = N_t - \alpha t$ the compensated Poisson process which is an L^2-martingale. Let $0 < \tau_1 < \tau_2 < \ldots < \tau_{N(t)}$ denote the jump times of N in $(0,t]$.

(a) Show that

$$E\Big[\sum_{i=1}^{N(t)} \tau_i\Big] = \frac{\alpha t^2}{2}.$$

Hint: For a homogeneous Poisson process, given that there are n jumps in an interval I, the locations of the jumps are n i.i.d. uniform random variables in I.

(b) Compute the integral

$$\int_{(0,t]} N(s-)\,dM(s).$$

"Compute" means to find a reasonably simple formula in terms of N_t and the τ_i's. One way is to justify the evaluation of this integral as a Lebesgue-Stieltjes integral.

(c) Use the formula you obtained in part (b) to check that the process $\int N(s-)\,dM(s)$ is a martingale. (Of course, this conclusion is part of the theory but the point here is to obtain it through hands-on computation. Part (a) and Exercise 2.28 take care of parts of the work.)

(d) Suppose N were predictable. Then the stochastic integral $\int N\,dM$ would exist and be a martingale. Show that this is not true and conclude that N cannot be predictable.

Hints: It might be easiest to find

$$\int_{(0,t]} N(s)\,dM(s) - \int_{(0,t]} N(s-)\,dM(s) = \int_{(0,t]} \big(N(s) - N(s-)\big)\,dM(s)$$

and use the fact that the integral of $N(s-)$ *is* a martingale.

Exercise 5.12. (Extended stochastic integral of the Poisson process.) Let N_t be a rate α Poisson process, $M_t = N_t - \alpha t$ and $N_-(t) = N(t-)$. Show that N_- is a modification of N, and

$$\mu_M^*\{(t,\omega) : N_t(\omega) \neq N_{t-}(\omega)\} = 0.$$

Thus the stochastic integral $N \cdot M$ can be defined according to the extension in Section 5.5 and this $N \cdot M$ must agree with $N_- \cdot M$.

Exercise 5.13. (Riemann sum approximation in $\mathcal{M}_2$.) Let M be an L^2 martingale, $X \in \mathcal{L}_2(M, \mathcal{P})$, and assume X also satisfies the hypotheses of Proposition 5.32. Let $\pi^m = \{0 = t_1^m < t_2^m < t_3^m < \cdots\}$ be partitions such that $\operatorname{mesh} \pi^m \to 0$. Let $\xi_i^{k,m} = (X_{t_i^m} \wedge k) \vee (-k)$, and define the simple predictable processes

$$W_t^{k,m,n} = \sum_{i=1}^{n} \xi_i^{m,k} \mathbf{1}_{(t_i^m, t_{i+1}^m]}(t).$$

Then there exists a subsequences $\{m(k)\}$ and $\{n(k)\}$ such that

$$\lim_{k\to\infty} \|X \cdot M - W^{k,m(k),n(k)} \cdot M\|_{\mathcal{M}_2} = 0.$$

Exercise 5.14. Let $0 < a < b < \infty$ be constants and $M \in \mathcal{M}_{2,\text{loc}}$. Find the stochastic integral

$$\int_{(0,t]} \mathbf{1}_{[a,b)}(s)\, dM_s.$$

Hint: Check that if $M \in \mathcal{M}_2$ then $\mathbf{1}_{(a-1/n,b-1/n]}$ converges to $\mathbf{1}_{[a,b)}$ in $\mathcal{L}_2(M, \mathcal{P})$.

Exercise 5.15. Does part (a) of Lemma 5.57 extend to semimartingales?

Exercise 5.16. (Space-time Poisson random measure.) In some applications the random occurrences modeled by a Poisson process also come with a spatial structure. Let U be a subset of $\mathbf{R}^d$ and ν a finite Borel measure on U. Let ξ be a Poisson random measure on $U \times \mathbf{R}_+$ with mean measure $\nu \otimes m$. Define the filtration $\mathcal{F}_t = \sigma\{\xi(B) : B \in \mathcal{B}_{U\times[0,t]}\}$. Fix a bounded Borel function h on $U \times \mathbf{R}_+$. Define

$$X_t = \int_{U\times[0,t]} h(x,s)\, \xi(dx,ds)\,, \quad H_t = \int_0^t \int_U h(x,s)\, \nu(dx)\, ds$$

and $M_t = X_t - H_t$.

(a) Show that these are cadlag processes, M is a martingale, H is an FV process and X is semimartingale.

(b) Show that for an adapted cadlag process Z,

$$\int_{(0,t]} Z(s-)\, dX_s = \int_{U\times[0,t]} Z(s-)h(x,s)\, \xi(dx,ds).$$

Hint. The Riemann sums in Proposition 5.37 would work.

(c) What moment assumption on an adapted cadlag process Z guarantees that $Z_- \in \mathcal{L}_2(M, \mathcal{P})$? Show that for these processes Z,

$$E\int_{(0,t]} Z(s-)\, dX_s = \int_0^t \int_U E[Z(s)]h(x,s)\, \nu(dx)\, ds.$$

Hint. Recall that stochastic integrals of $\mathcal{L}_2$ integrands are themselves mean zero martingales.

Chapter 6

Itô's Formula

Itô's formula works as a fundamental theorem of calculus in stochastic analysis. We are familiar from calculus with the fact that if $x(t)$ is a continuously differentiable function of time, then

$$f(x(t)) = f(x(0)) + \int_0^t f'(x(s))\, x'(s)\, ds.$$

However, the exact analogue for Brownian motion

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\, dB_s$$

is *not* true. The wild oscillation of a Brownian path brings a second order term into the identity that involves quadratic variation $d[B] = dt$:

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\, dB_s + \tfrac{1}{2}\int_0^t f''(B_s)\, ds.$$

This is a special case of Itô's formula. When the process in question possesses jumps, even further terms need to be added on the right-hand side.

Right-continuity of the filtration $\{\mathcal{F}_t\}$ is not needed for the proofs of this chapter. This property might be needed for defining the stochastic integral with respect to a semimartingale one wants to work with. As explained in the beginning of Section 5.3, under this assumption we can apply Theorem 3.21 (fundamental theorem of local martingales) to guarantee that every semimartingale has a decomposition whose local martingale part is a local L^2-martingale. This L^2 property was used for the construction of the stochastic integral. The problem can arise only when the local martingale part can have unbounded jumps. When the local martingale is continuous or its jumps have a uniform bound, it can be localized into an L^2 martingale and the fundamental theorem is not needed.

6.1. Itô's formula: proofs and special cases

We prove Itô's formula in two main stages, first for real-valued semimartingales and then for vector-valued semimartingales. Additionally we state several simplifications that result if the cadlag semimartingale specializes to an FV process, a continuous semimartingale, or Brownian motion.

For an open set $D \subseteq \mathbf{R}$, $C^2(D)$ is the space of functions $f : D \to \mathbf{R}$ such that the derivatives f' and f'' exist everywhere on D and are continuous functions. For a real or vector-valued cadlag process X, the jump at time s is denoted by

$$\Delta X_s = X_s - X_{s-}$$

as before. Recall also the notion of the closure of the path over a time interval, for a cadlag process given by

$$\overline{X[0,t]} = \{X(s) : 0 \le s \le t\} \cup \{X(s-) : 0 < s \le t\}.$$

Itô's formula contains a term which is a sum over the jumps of the process. This sum has at most countably many terms because a cadlag path has at most countably many discontinuities (Lemma A.7). It is also possible to define rigorously what is meant by a convergent sum of *uncountably* many terms, and arrive at the same value. See the discussion around (A.5) in the appendix.

Theorem 6.1. *Fix $0 < T < \infty$. Let D be an open subset of $\mathbf{R}$ and $f \in C^2(D)$. Let Y be a cadlag semimartingale with quadratic variation process $[Y]$. Assume that for all ω outside some event of probability zero, $\overline{Y[0,T]} \subseteq D$. Then*

$$\begin{aligned} f(Y_t) = f(Y_0) &+ \int_{(0,t]} f'(Y_{s-})\,dY_s + \tfrac{1}{2}\int_{(0,t]} f''(Y_{s-})\,d[Y]_s \\ &+ \sum_{s\in(0,t]} \Big\{ f(Y_s) - f(Y_{s-}) - f'(Y_{s-})\Delta Y_s - \tfrac{1}{2} f''(Y_{s-})(\Delta Y_s)^2 \Big\}. \end{aligned} \tag{6.1}$$

Part of the conclusion is that the last sum over $s \in (0,t]$ converges absolutely for almost every ω. Both sides of the equality above are cadlag processes, and the meaning of the equality is that these processes are indistinguishable on $[0,T]$. In other words, there exists an event Ω_0 of full probability such that for $\omega \in \Omega_0$, (6.1) holds for all $0 \le t \le T$.

Proof. The proof starts by Taylor expanding f. We formulate this in the following way. Define the function γ on $D \times D$ by $\gamma(x,x) = 0$ for all $x \in D$, and for $x \neq y$

$$\gamma(x,y) = \frac{1}{(y-x)^2}\Big\{ f(y) - f(x) - f'(x)(y-x) - \tfrac{1}{2}f''(x)(y-x)^2 \Big\}. \tag{6.2}$$

On the set $\{(x,y) \in D \times D : x \neq y\}$ γ is continuous as a quotient of two continuous functions. That γ is continuous also at diagonal points (z,z) follows from Taylor's theorem (Theorem A.14 in the appendix). Given $z \in D$, pick $r > 0$ small enough so that $G = (z-r, z+r) \subseteq D$. Then for $x, y \in G$, $x \neq y$, there exists a point $\theta_{x,y}$ between x and y such that

$$f(y) = f(x) + f'(x)(y-x) + \tfrac{1}{2}f''(\theta_{x,y})(y-x)^2.$$

So for these (x,y)

$$\gamma(x,y) = \tfrac{1}{2}f''(\theta_{x,y}) - \tfrac{1}{2}f''(x).$$

As (x,y) converges to (z,z), $\theta_{x,y}$ converges to z, and so by the assumed continuity of f''

$$\gamma(x,y) \longrightarrow \tfrac{1}{2}f''(z) - \tfrac{1}{2}f''(z) = 0 = \gamma(z,z).$$

We have verified that γ is continuous on $D \times D$.

Write

$$f(y) - f(x) = f'(x)(y-x) + \tfrac{1}{2}f''(x)(y-x)^2 + \gamma(x,y)(y-x)^2.$$

Given a partition $\pi = \{t_i\}$ of $[0,\infty)$, apply the above identity to each partition interval to write

$$\begin{aligned} f(Y_t) &= f(Y_0) + \sum_i \big\{ f(Y_{t\wedge t_{i+1}}) - f(Y_{t\wedge t_i}) \big\} \\ &= f(Y_0) + \sum_i f'(Y_{t\wedge t_i})(Y_{t\wedge t_{i+1}} - Y_{t\wedge t_i}) \qquad (6.3) \\ &\quad + \tfrac{1}{2}\sum_i f''(Y_{t\wedge t_i})(Y_{t\wedge t_{i+1}} - Y_{t\wedge t_i})^2 \qquad (6.4) \\ &\quad + \sum_i \gamma(Y_{t\wedge t_i}, Y_{t\wedge t_{i+1}})(Y_{t\wedge t_{i+1}} - Y_{t\wedge t_i})^2. \qquad (6.5) \end{aligned}$$

By Propositions 5.37 and 5.64 we can fix a sequence of partitions π^ℓ such that $\mathrm{mesh}(\pi^\ell) \to 0$, and so that the following limits happen almost surely, uniformly for $t \in [0,T]$, as $\ell \to \infty$.

(i) The sum on line (6.3) converges to

$$\int_{(0,t]} f'(Y_{s-})\,dY_s.$$

(ii) The term on line (6.4) converges to

$$\tfrac{1}{2}\int_{(0,t]} f''(Y_{s-})\,d[Y]_s.$$

(iii) The limit

$$\sum_i (Y_{t\wedge t_{i+1}} - Y_{t\wedge t_i})^2 \longrightarrow [Y]_t$$

happens.

Fix ω so that the limits in items (i)–(iii) above happen. We apply the scalar case of Lemma A.13, but simplified so that ϕ and γ in (A.8) have no time variables, to the cadlag function $s \to Y_s(\omega)$ on $[0,t]$ and the sequence of partitions π^ℓ chosen above. For the closed set K in Lemma A.13 take $K = \overline{Y[0,T]}$. For the continuous function ϕ in Lemma A.13 take $\phi(x,y) = \gamma(x,y)(y-x)^2$. By hypothesis, K is a subset of D. Consequently, as verified above, the function

$$\gamma(x,y) = \begin{cases} (x-y)^{-2}\phi(x,y), & x \neq y \\ 0, & x = y \end{cases}$$

is continuous on $K\times K$. Assumption (A.9) of Lemma A.13 holds by item (iii) above. The hypotheses of Lemma A.13 have been verified. The conclusion is that for this fixed ω and each $t \in [0,T]$, the sum on line (6.5) converges to

$$\begin{aligned}\sum_{s\in(0,t]} \phi(Y_{s-},Y_s) &= \sum_{s\in(0,t]} \gamma(Y_{s-},Y_s)\bigl(Y_s - Y_{s-}\bigr)^2 \\ &= \sum_{s\in(0,t]} \Bigl\{ f(Y_s) - f(Y_{s-}) - f'(Y_{s-})\Delta Y_s - \tfrac{1}{2} f''(Y_{s-})(\Delta Y_s)^2 \Bigr\}.\end{aligned}$$

Lemma A.13 also contains the conclusion that this last sum is absolutely convergent.

To summarize, given $0 < T < \infty$, we have shown that for almost every ω, (6.1) holds for all $0 \le t \le T$. □

Remark 6.2. (Semimartingale property) A corollary of Theorem 6.1 is that under the conditions of the theorem $f(Y)$ is a semimartingale. Equation (6.1) expresses $f(Y)$ as a sum of semimartingales. The integral $f''(Y_-)\,d[Y]$ is an FV process. To see that the sum over jump times $s \in (0,t]$ produces also an FV process, fix ω such that $Y_s(\omega)$ is a cadlag function and (6.1) holds. Let $\{s_i\}$ denote the (at most countably many) jumps of $s \mapsto Y_s(\omega)$ in $[0,T]$. The theorem gives the absolute convergence

$$\sum_i \Bigl| \Delta f(Y_{s_i}(\omega)) - f'(Y_{s_i-}(\omega))\Delta Y_{s_i}(\omega) - \tfrac{1}{2} f''(Y_{s_i-}(\omega))(\Delta Y_{s_i}(\omega))^2 \Bigr| < \infty.$$

Consequently for this fixed ω the sum in (6.1) defines a function in $BV[0,T]$, as in Example 1.13.

Let us state some simplifications of Itô's formula.

Corollary 6.3. *Under the hypotheses of Theorem 6.1 we have the following special cases.*

(a) *If Y is continuous on $[0,T]$, then*

$$f(Y_t) = f(Y_0) + \int_0^t f'(Y_s)\,dY_s + \tfrac{1}{2}\int_0^t f''(Y_s)\,d[Y]_s. \tag{6.6}$$

(b) *If Y has bounded variation on $[0,T]$, then*

$$\begin{aligned} f(Y_t) = f(Y_0) &+ \int_{(0,t]} f'(Y_{s-})\,dY_s \\ &+ \sum_{s\in(0,t]} \Big\{ f(Y_s) - f(Y_{s-}) - f'(Y_{s-})\Delta Y_s \Big\} \end{aligned} \tag{6.7}$$

(c) *If $Y_t = Y_0 + B_t$, where B is a standard Brownian motion independent of Y_0, then*

$$f(B_t) = f(Y_0) + \int_0^t f'(Y_0 + B_s)\,dB_s + \tfrac{1}{2}\int_0^t f''(Y_0 + B_s)\,ds. \tag{6.8}$$

Proof. Part (a). Continuity eliminates the sum over jumps, and renders endpoints of intervals irrelevant for integration.

Part (b). By Corollary A.11 the quadratic variation of a cadlag BV path consists exactly of the squares of the jumps. Consequently

$$\tfrac{1}{2}\int_{(0,t]} f''(Y_{s-})\,d[Y]_s = \sum_{s\in(0,t]} \tfrac{1}{2} f''(Y_{s-})(\Delta Y_s)^2$$

and we get cancellation in the formula (6.1).

Part (c). Specialize part (a) to $[B]_t = t$. □

The open set D in the hypotheses of Itô's formula does not have to be an interval, so it can be disconnected.

The important hypothesis $\overline{Y[0,T]} \subseteq D$ prevents the process from reaching the boundary. Precisely speaking, the hypothesis implies that for some $\delta > 0$, $\operatorname{dist}(Y(s), D^c) \ge \delta$ for all $s \in [0,T]$. To prove this, assume the contrary, namely the existence of $s_i \in [0,T]$ such that $\operatorname{dist}(Y(s_i), D^c) \to 0$. Since $[0,T]$ is compact, we may pass to a convergent subsequence $s_i \to s$. And then by the cadlag property, $Y(s_i)$ converges to some point y. Since $\operatorname{dist}(y, D^c) = 0$ and D^c is a closed set, $y \in D^c$. But $y \in \overline{Y[0,T]}$, and we have contradicted $\overline{Y[0,T]} \subseteq D$.

But note that the δ (the distance to the boundary of D) can depend on ω. So the hypothesis $\overline{Y[0,T]} \subseteq D$ does *not* require that there exists a fixed closed subset H of D such that $P\{Y(t) \in H \text{ for all } t \in [0,T]\} = 1$.

Hypothesis $\overline{Y[0,T]} \subseteq D$ is needed because otherwise a "blow-up" at the boundary can cause problems. The next example illustrates why we need to assume the containment in D of the *closure* $\overline{Y[0,T]}$, and not merely $Y[0,T] \subseteq D$.

Example 6.4. Let $D = (-\infty, 1) \cup (\frac{3}{2}, \infty)$, and define

$$f(x) = \begin{cases} \sqrt{1-x}, & x < 1 \\ 0, & x > \frac{3}{2}, \end{cases}$$

a C^2-function on D. Define the deterministic process

$$Y_t = \begin{cases} t, & 0 \le t < 1 \\ 1+t, & t \ge 1. \end{cases}$$

$Y_t \in D$ for all $t \ge 0$. However, if $t > 1$,

$$\begin{aligned} \int_{(0,t]} f'(Y_{s-})\, dY_s &= \int_{(0,1)} f'(s)\, ds + f'(Y_{1-}) + \int_{(1,t]} f'(s)\, ds \\ &= -1 + (-\infty) + 0. \end{aligned}$$

As the calculation shows, the integral is not finite. The problem is that the closure $\overline{Y[0,t]}$ contains the point 1 which lies at the boundary of D, and the derivative f' blows up there.

We extend Itô's formula to vector-valued semimartingales. For purposes of matrix multiplication we think of points $\mathbf{x} \in \mathbf{R}^d$ as column vectors, so with T denoting transposition,

$$\mathbf{x} = [x_1, x_2, \dots, x_d]^T.$$

Let $Y_1(t), Y_2(t), \dots, Y_d(t)$ be cadlag semimartingales with respect to a common filtration $\{\mathcal{F}_t\}$. We write $Y(t) = [Y_1(t), \dots, Y_d(t)]^T$ for the column vector with coordinates $Y_1(t), \dots, Y_d(t)$, and call Y an $\mathbf{R}^d$-valued semimartingale. Its jump is the vector of jumps in the coordinates:

$$\Delta Y(t) = [\Delta Y_1(t), \Delta Y_2(t), \dots, \Delta Y_d(t)]^T.$$

For $0 < T < \infty$ and an open subset D of $\mathbf{R}^d$, $C^{1,2}([0,T] \times D)$ is the space of continuous functions $f : [0,T] \times D \to \mathbf{R}$ whose partial derivatives f_t, f_{x_i}, and f_{x_i,x_j} exist and are continuous in the interior $(0,T) \times D$, and extend as continuous functions to $[0,T] \times D$. So the superscript in $C^{1,2}$ stands for one continuous time derivative and two continuous space derivatives. For $f \in C^{1,2}([0,T] \times D)$ and $(t, \mathbf{x}) \in [0,T] \times D$, the spatial gradient

$$Df(t, \mathbf{x}) = \big[f_{x_1}(t, \mathbf{x}), f_{x_2}(t, \mathbf{x}), \dots, f_{x_d}(t, \mathbf{x})\big]^T$$

is the column vector of first-order partial derivatives in the space variables, and the Hessian matrix $D^2 f(t,\mathbf{x})$ is the $d \times d$ matrix of second-order spatial partial derivatives:

$$D^2 f(t,\mathbf{x}) = \begin{bmatrix} f_{x_1,x_1}(t,\mathbf{x}) & f_{x_1,x_2}(t,\mathbf{x}) & \cdots & f_{x_1,x_d}(t,\mathbf{x}) \\ f_{x_2,x_1}(t,\mathbf{x}) & f_{x_2,x_2}(t,\mathbf{x}) & \cdots & f_{x_2,x_d}(t,\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_d,x_1}(t,\mathbf{x}) & f_{x_d,x_2}(t,\mathbf{x}) & \cdots & f_{x_d,x_d}(t,\mathbf{x}) \end{bmatrix}.$$

Theorem 6.5. *Fix $d \geq 2$ and $0 < T < \infty$. Let D be an open subset of $\mathbf{R}^d$ and $f \in C^{1,2}([0,T] \times D)$. Let Y be an $\mathbf{R}^d$-valued cadlag semimartingale such that outside some event of probability zero, $\overline{Y[0,T]} \subseteq D$. Then*

$$\begin{aligned} f(t,Y(t)) = f(0,Y(0)) &+ \int_0^t f_t(s,Y(s))\,ds \\ &+ \sum_{j=1}^d \int_{(0,t]} f_{x_j}(s,Y(s-))\,dY_j(s) \\ &+ \tfrac{1}{2} \sum_{1\leq j,k\leq d} \int_{(0,t]} f_{x_j,x_k}(s,Y(s-))\,d[Y_j,Y_k](s) \\ + \sum_{s\in(0,t]} \Big\{ &f(s,Y(s)) - f(s,Y(s-)) \\ - Df(s,Y(s-))^T \Delta Y(s) &- \tfrac{1}{2}\Delta Y(s)^T D^2 f(s,Y(s-))\Delta Y(s) \Big\} \end{aligned} \tag{6.9}$$

Proof. Let us write $Y^k_t = Y_k(t)$ in the proof. The pattern is the same as in the scalar case. Define a function ϕ on $[0,T]^2 \times D^2$ by the equality

$$\begin{aligned} f(t,\mathbf{y}) - f(s,\mathbf{x}) = f_t(s,\mathbf{x})(t-s) + Df(s,\mathbf{x})^T(\mathbf{y}-\mathbf{x}) \\ + \tfrac{1}{2}(\mathbf{y}-\mathbf{x})^T D^2 f(s,\mathbf{x})(\mathbf{y}-\mathbf{x}) + \phi(s,t,\mathbf{x},\mathbf{y}). \end{aligned} \tag{6.10}$$

Apply this to partition intervals to write

$$\begin{aligned} f(t,Y_t) = f(0,Y_0) &+ \sum_i \big\{ f(t\wedge t_{i+1}, Y_{t\wedge t_{i+1}}) - f(t\wedge t_i, Y_{t\wedge t_i}) \big\} \\ = f(0,Y_0)& \end{aligned}$$

$$+ \sum_i f_t(t\wedge t_i, Y_{t\wedge t_i})\big((t\wedge t_{i+1}) - (t\wedge t_i)\big) \tag{6.11}$$

$$+ \sum_{k=1}^d \sum_i f_{x_k}(t\wedge t_i, Y_{t\wedge t_i})\big(Y^k_{t\wedge t_{i+1}} - Y^k_{t\wedge t_i}\big) \tag{6.12}$$

$$+ \tfrac{1}{2} \sum_{1\leq j,k\leq d} \sum_i f_{x_j,x_k}(t\wedge t_i, Y_{t\wedge t_i})\big(Y^j_{t\wedge t_{i+1}} - Y^j_{t\wedge t_i}\big)\big(Y^k_{t\wedge t_{i+1}} - Y^k_{t\wedge t_i}\big) \tag{6.13}$$

$$+\sum_i \phi(t\wedge t_i, t\wedge t_{i+1}, Y_{t\wedge t_i}, Y_{t\wedge t_{i+1}}). \tag{6.14}$$

By Propositions 5.37 and 5.64 we can fix a sequence of partitions π^ℓ such that $\mathrm{mesh}(\pi^\ell)\to 0$, and so that the following limits happen almost surely, uniformly for $t\in[0,T]$, as $\ell\to\infty$.

(i) Line (6.11) converges to

$$\int_{(0,t]} f_t(s, Y_s)\,ds.$$

(ii) Line (6.12) converges to

$$\sum_{k=1}^d \int_{(0,t]} f_{x_k}(s, Y_{s-})\,dY_s^k.$$

(iii) Line (6.13) converges to

$$\tfrac{1}{2}\sum_{1\le j,k\le d}\int_{(0,t]} f_{x_j,x_k}(s, Y_{s-})\,d[Y^j, Y^k]_s.$$

(iv) The limit

$$\sum_i \bigl(Y^k_{t\wedge t_{i+1}} - Y^k_{t\wedge t_i}\bigr)^2 \longrightarrow [Y^k]_t$$

happens for $1\le k\le d$.

In (i)–(iii) the integrand is the left limit process, for example in (ii)

$$\lim_{r\searrow s} f_{x_k}(r, Y_r) = f_{x_k}\bigl(\lim_{r\searrow s}(r, Y_r)\bigr) = f_{x_k}(s, Y_{s-}).$$

However in (i) the ds integral does not distinguish between Y_{s-} and Y_s because a cadlag path has at most countably many jumps.

Fix ω such that $\overline{Y[0,T]}\subseteq D$ and the limits in items (i)–(iv) hold. By the above paragraph and by hypothesis these conditions hold for almost every ω.

To treat the sum on line (6.14), we apply Lemma A.13 to the $\mathbf{R}^d$-valued cadlag function $s\mapsto Y_s(\omega)$ on $[0,T]$, with the function ϕ defined by (6.10), the closed set $K=\overline{Y[0,T]}$, and the sequence of partitions π^ℓ chosen above. We need to check that ϕ and the set K satisfy the hypotheses of Lemma A.13. Continuity of ϕ follows from the definition (6.10). Next we argue that if $(s_n, t_n, \mathbf{x}_n, \mathbf{y}_n)\to(u, u, \mathbf{z}, \mathbf{z})$ in $[0,T]^2\times K^2$ while for each n, either $s_n\neq t_n$ or $\mathbf{x}_n\neq\mathbf{y}_n$, then

$$\frac{\phi(s_n, t_n, \mathbf{x}_n, \mathbf{y}_n)}{|t_n - s_n| + |\mathbf{y}_n - \mathbf{x}_n|^2}\to 0. \tag{6.15}$$

Given $\varepsilon > 0$, let I be an interval around u in $[0,T]$ and let B be an open ball centered at $\mathbf{z}$ and contained in D such that

$$|f_t(v,\mathbf{w}) - f_t(u,\mathbf{z})| + |D^2 f(v,\mathbf{w}) - D^2 f(u,\mathbf{z})| \le \varepsilon$$

for all $v \in I, \mathbf{w} \in B$. Such an interval I and ball B exist by the openness of D and by the assumption of continuity of derivatives of f in $[0,T] \times D$. For large enough n, we have $s_n, t_n \in I$ and $\mathbf{x}_n, \mathbf{y}_n \in B$. Since a ball is convex, by Taylor's formula (A.16) we can write

$$\begin{aligned}\phi(s_n,t_n,\mathbf{x}_n,\mathbf{y}_n) &= \big(f_t(\tau_n,\mathbf{y}_n) - f_t(s_n,\mathbf{x}_n)\big)(t_n - s_n)\\ &\quad + \tfrac{1}{2}(\mathbf{y}_n - \mathbf{x}_n)^T\big(D^2 f(s_n,\xi_n) - D^2 f(s_n,\mathbf{x}_n)\big)(\mathbf{y}_n - \mathbf{x}_n),\end{aligned}$$

where τ_n lies between s_n and t_n, and ξ_n is a point on the line segment connecting $\mathbf{x}_n$ and $\mathbf{y}_n$. Hence $\tau_n \in I$ and $\xi_n \in B$, and by Schwarz inequality in the form (A.7),

$$\begin{aligned}|\phi(s_n,t_n,\mathbf{x}_n,\mathbf{y}_n)| &\le \big|f_t(\tau_n,\mathbf{y}_n) - f_t(s_n,\mathbf{x}_n)\big| \cdot |t_n - s_n|\\ &\quad + \big|D^2 f(s_n,\xi_n) - D^2 f(t_n,\mathbf{x}_n)\big| \cdot |\mathbf{y}_n - \mathbf{x}_n|^2\\ &\le 2\varepsilon \cdot \big(|t_n - s_n| + |\mathbf{y}_n - \mathbf{x}_n|^2\big).\end{aligned}$$

Thus

$$\frac{\phi(s_n,t_n,\mathbf{x}_n,\mathbf{y}_n)}{|t_n - s_n| + |\mathbf{y}_n - \mathbf{x}_n|^2} \le 2\varepsilon$$

for large enough n, and we have verified (6.15).

The function

$$(s,t,\mathbf{x},\mathbf{y}) \mapsto \frac{\phi(s,t,\mathbf{x},\mathbf{y})}{|t-s| + |\mathbf{y}-\mathbf{x}|^2}$$

is continuous at points where either $s \ne t$ or $\mathbf{x} \ne \mathbf{y}$, as a quotient of two continuous functions. Consequently the function γ defined by (A.8) is continuous on $[0,T]^2 \times K^2$.

Hypothesis (A.9) of Lemma A.13 is a consequence of the limit in item (iv) above.

The hypotheses of Lemma A.13 have been verified. By this lemma, for this fixed ω and each $t \in [0,T]$, the sum on line (6.14) converges to

$$\begin{aligned}\sum_{s\in(0,t]} \phi(s,s,Y_{s-},Y_s) &= \sum_{s\in(0,t]} \Big\{f(s,Y_s) - f(s,Y_{s-})\\ &\quad - Df(s,Y_{s-})\Delta Y_s - \tfrac{1}{2}\Delta Y_s^T D^2 f(s,Y_{s-})\Delta Y_s\Big\}.\end{aligned}$$

This completes the proof of Theorem 6.5. □

Remark 6.6. (Notation) Often Itô's formula is expressed in terms of differential notation which is more economical than the integral notation. As an example, if Y is a continuous $\mathbf{R}^d$-valued semimartingale, equation (6.9) can be written as

$$
\begin{aligned}
df(t, Y(t)) &= f_t(t, Y(t))\, dt + \sum_{j=1}^{d} f_{x_j}(t, Y(t-))\, dY_j(t) \\
&\quad + \tfrac{1}{2} \sum_{1 \le j,k \le d} f_{x_j,x_k}(t, Y(t-))\, d[Y_j, Y_k](t).
\end{aligned} \tag{6.16}
$$

As mentioned already, these "stochastic differentials" have no rigorous meaning. The formula above is to be regarded only as an abbreviation of the integral formula (6.9).

We state the Brownian motion case as a corollary. For $f \in C^2(D)$ in $\mathbf{R}^d$, the *Laplace* operator Δ is defined by

$$\Delta f = f_{x_1,x_1} + \cdots + f_{x_d,x_d}.$$

A function f is *harmonic* in D if $\Delta f = 0$ on D.

Corollary 6.7. *Let $B(t) = (B_1(t), \dots, B_d(t))$ be Brownian motion in $\mathbf{R}^d$, with random initial point $B(0)$, and $f \in C^2(\mathbf{R}^d)$. Then*

$$f(B(t)) = f(B(0)) + \int_0^t Df(B(s))^T\, dB(s) + \tfrac{1}{2} \int_0^t \Delta f(B(s))\, ds. \tag{6.17}$$

Suppose f is harmonic in an open set $D \subseteq \mathbf{R}^d$. Let D_1 be an open subset of D such that $\operatorname{dist}(D_1, D^c) > 0$. Assume initially $B(0) = \mathbf{z}$ for some point $\mathbf{z} \in D_1$, and let

$$\tau = \inf\{t \ge 0 : B(t) \in D_1^c\} \tag{6.18}$$

be the exit time for Brownian motion from D_1. Then $f(B^\tau(t))$ is a local L^2-martingale.

Proof. Formula (6.17) comes directly from Itô's formula, because $[B_i, B_j] = \delta_{i,j} t$.

The process B^τ is a (vector) L^2 martingale that satisfies $\overline{B^\tau[0,T]} \subseteq D$ for all $T < \infty$. Thus Itô's formula applies. Note that $[B_i^\tau, B_j^\tau]_t = [B_i, B_j]_t^\tau = \delta_{i,j}(t \wedge \tau)$. Hence $\Delta f = 0$ in D eliminates the second-order term, and the formula simplifies to

$$f(B^\tau(t)) = f(\mathbf{z}) + \int_0^t Df(B^\tau(s))^T\, dB^\tau(s)$$

which shows that $f(B^\tau(t))$ is a local L^2-martingale. ∎

6.2. Applications of Itô's formula

One can use Itô's formula to find pathwise expressions for stochastic integrals.

Example 6.8. To evaluate $\int_0^t B_s^k \, dB_s$ for standard Brownian motion and $k \geq 1$, take $f(x) = (k+1)^{-1}x^{k+1}$, so that $f'(x) = x^k$ and $f''(x) = kx^{k-1}$. Itô's formula gives

$$\int_0^t B_s^k \, dB_s = (k+1)^{-1} B_t^{k+1} - \frac{k}{2} \int_0^t B_s^{k-1} \, ds.$$

The integral on the right is a familiar Riemann integral of the continuous function $s \mapsto B_s^{k-1}$.

Often Itô's formula is used to find martingales, which can be useful for calculations. Here is a systematic way to do this for Brownian motion.

Lemma 6.9. *Suppose $f \in C^{1,2}(\mathbf{R}_+ \times \mathbf{R})$ and $f_t + \frac{1}{2} f_{xx} = 0$. Let B_t be one-dimensional standard Brownian motion. Then $f(t, B_t)$ is a local L^2-martingale. If*

$$\int_0^T E\big[f_x(t, B_t)^2\big] \, dt < \infty, \tag{6.19}$$

then $f(t, B_t)$ is an L^2-martingale on $[0, T]$.

Proof. Since $[B]_t = t$, (6.9) specializes to

$$\begin{aligned} f(t, B_t) &= f(0,0) + \int_0^t f_x(s, B_s) \, dB_s + \int_0^t \big(f_t(s, B_s) + \tfrac{1}{2} f_{xx}(s, B_s)\big) \, ds \\ &= f(0,0) + \int_0^t f_x(s, B_s) \, dB_s \end{aligned}$$

where the last line is a local L^2-martingale. (The integrand $f_x(s, B_s)$ is a continuous process, hence predictable, and satisfies the local boundedness condition (5.29).)

The integrability condition (6.19) guarantees that $f_x(s, B_s)$ lies in the space $\mathcal{L}_2(B, \mathcal{P})$ of integrands on the interval $[0, T]$. In our earlier development of stochastic integration we always considered processes defined for all time. To get an integrand process on the entire time line $[0, \infty)$, one can extend $f_x(s, B_s)$ by declaring it identically zero on (T, ∞). This does not change the integral on $[0, T]$. ∎

Example 6.10. Let $\mu \in \mathbf{R}$ and $\sigma \neq 0$ be constants. Let $a < 0 < b$. Let B_t be one-dimensional standard Brownian motion, and $X_t = \mu t + \sigma B_t$ a Brownian motion with drift. Question: What is the probability that X_t exits the interval (a, b) through the point b?

Define the stopping time

$$\tau = \inf\{t > 0 : X_t = a \text{ or } X_t = b\}.$$

First we need to check that $\tau < \infty$ almost surely. For example, the events

$$\{\sigma B_{n+1} - \sigma B_n > b - a + |\mu| + 1\},\ n = 0, 1, 2, \dots$$

are independent and have a common positive probability. Hence one of them happens almost surely. Consequently X_n cannot remain in (a, b) for all n.

We seek a function h such that $h(X_t)$ is a martingale. Then, if we could justify $Eh(X_\tau) = h(0)$, we could compute the desired probability $P(X_\tau = b)$ from

$$h(0) = Eh(X_\tau) = h(a)P(X_\tau = a) + h(b)P(X_\tau = b).$$

To utilize Lemma 6.9, let $f(t, x) = h(\mu t + \sigma x)$. The condition $f_t + \frac{1}{2}f_{xx} = 0$ becomes

$$\mu h' + \tfrac{1}{2}\sigma^2 h'' = 0.$$

At this point we need to decide whether $\mu = 0$ or not. Let us work the case $\mu \neq 0$. Solving for h gives

$$h(x) = C_1 \exp\bigl\{-2\mu\sigma^{-2}x\bigr\} + C_2$$

for two constants of integration C_1, C_2. To check (6.19), from $f(t, x) = h(\mu t + \sigma x)$ derive

$$f_x(t, x) = -2C_2\mu\sigma^{-1} \exp\bigl\{-2\mu\sigma^{-2}(\mu t + \sigma x)\bigr\}.$$

Since B_t is a mean zero normal with variance t, one can verify that (6.19) holds for all $T < \infty$.

Now

$$M_t = h(X_t) = C_1 e^{-2\frac{\mu}{\sigma}B_t - 2(\frac{\mu}{\sigma})^2 t} + C_2 \tag{6.20}$$

is a martingale. By optional stopping, $M_{\tau\wedge t}$ is also a martingale, and so $EM_{\tau\wedge t} = EM_0 = h(0)$. By path continuity and $\tau < \infty$, $M_{\tau\wedge t} \to M_\tau$ almost surely as $t \to \infty$. Furthermore, the process $M_{\tau\wedge t}$ is bounded, because up to time τ process X_t remains in $[a, b]$, and so $|M_{\tau\wedge t}| \le C \equiv \sup_{a\le x\le b}|h(x)|$. Dominated convergence gives $EM_{\tau\wedge t} \to EM_\tau$ as $t \to \infty$. We have verified that $Eh(X_\tau) = h(0)$.

Finally, we can choose the constants C_1 and C_2 so that $h(b) = 1$ and $h(a) = 0$. After some details,

$$P(X_\tau = b) = h(0) = \frac{e^{-2\mu a/\sigma^2} - 1}{e^{-2\mu a/\sigma^2} - e^{-2\mu b/\sigma^2}}. \tag{6.21}$$

Can you explain what you see as you let either $a \to -\infty$ or $b \to \infty$? (Decide first whether μ is positive or negative.)

We leave the case $\mu = 0$ as an exercise. You should get $P(X_\tau = b) = (-a)/(b - a)$.

Martingale (6.20) found above is a special case of the general exponential martingale $M_t = e^{\alpha B_t - \alpha^2 t/2}$ for one-dimensional Brownian motion and real α. This is turn is a special case of the following: $M_t = e^{X_t - \frac{1}{2}[X]_t}$ is a continuous local L^2 martingale whenever X is. This is revealed by a quick Itô computation:

$$dM = M\,dX - \tfrac{1}{2}M\,d[X] + \tfrac{1}{2}M\,d[X] = M\,dX,$$

where $[X,[X]] = [[X]] = 0$ due to continuity and bounded variation (Lemma A.10). The computation reveals that $M_t = M_0 + \int_0^t M_s\,dX_s$, which is a local L^2 martingale by the construction of the stochastic integral on the right-hand side.

Next we use Itô's formula to investigate recurrence and transience of Brownian motion, and whether Brownian motion ever hits a point. Let us first settle these questions in one dimension.

Proposition 6.11. *Let B_t be Brownian motion in $\mathbf{R}$. Then $\overline{\lim}_{t\to\infty} B_t = \infty$ and $\underline{\lim}_{t\to\infty} B_t = -\infty$, almost surely. Consequently almost every Brownian path visits every point infinitely often.*

Proof. Let $\tau_0 = 0$ and $\tau(k+1) = \inf\{t > \tau(k) : |B_t - B_{\tau(k)}| = 4^{k+1}\}$. By the strong Markov property of Brownian motion, for each k the restarted process $\{B_{\tau(k)+s} - B_{\tau(k)} : s \geq\}$ is a standard Brownian motion, independent of $\mathcal{F}_{\tau(k)}$. By symmetry, or by the case $\mu = 0$ of Example 6.10,

$$P\big[B_{\tau(k+1)} - B_{\tau(k)} = 4^{k+1}\big] = P\big[B_{\tau(k+1)} - B_{\tau(k)} = -4^{k+1}\big] = \tfrac{1}{2}.$$

By the strong Markov property these random variables indexed by k are independent. Thus, for almost every ω, there are arbitrarily large j and k such that $B_{\tau(j+1)} - B_{\tau(j)} = 4^{j+1}$ and $B_{\tau(k+1)} - B_{\tau(k)} = -4^{k+1}$. But then since

$$|B_{\tau(j)}| \leq \sum_{i=1}^{j} 4^i = \frac{4^{j+1}-1}{4-1} \leq \frac{4^{j+1}}{2},$$

$B_{\tau(j+1)} \geq 4^j$, and by the same argument $B_{\tau(k+1)} \leq -4^k$. Thus $\overline{\lim}_{t\to\infty} B_t = \infty$ and $\underline{\lim}_{t\to\infty} B_t = -\infty$ almost surely.

Almost every Brownian path visits every point infinitely often due to a special property of one dimension: it is impossible to go "around" a point. □

Proposition 6.12. *Let B_t be Brownian motion in $\mathbf{R}^d$, and let $P^{\mathbf{z}}$ denote the probability measure when the process B_t is started at point $\mathbf{z} \in \mathbf{R}^d$. Let*

$$\tau_r = \inf\{t \geq 0 : |B_t| \leq r\}$$

be the first time Brownian motion hits the ball of radius r around the origin.

(a) *If $d = 2$, $P^{\mathbf{z}}(\tau_r < \infty) = 1$ for all $r > 0$ and $z \in \mathbf{R}^d$.*

(b) *If $d \geq 3$, then for $\mathbf{z}$ outside the ball of radius r,*

$$P^{\mathbf{z}}(\tau_r < \infty) = \left(\frac{r}{|\mathbf{z}|}\right)^{d-2}.$$

There will be an almost surely finite time T such that $|B_t| > r$ for all $t \geq T$.

(c) *For $d \geq 2$ and any $\mathbf{z}, \mathbf{y} \in \mathbf{R}^d$,*

$$P^{\mathbf{z}}[\, B_t \neq \mathbf{y} \text{ for all } 0 < t < \infty] = 1.$$

Note that $\mathbf{z} = \mathbf{y}$ is allowed. That is why $t = 0$ is not included in the event.

Proof. Observe that for (c) it suffices to consider $\mathbf{y} = 0$, because

$$P^{\mathbf{z}}[\, B_t \neq \mathbf{y} \text{ for all } 0 < t < \infty] = P^{\mathbf{z}-\mathbf{y}}[\, B_t \neq 0 \text{ for all } 0 < t < \infty].$$

Then, it suffices to consider $\mathbf{z} \neq 0$, because if we have the result for all $\mathbf{z} \neq 0$, then we can use the Markov property to restart the Brownian motion after a small time s:

$$\begin{aligned} &P^0[\, B_t \neq 0 \text{ for all } 0 < t < \infty] \\ &\quad = \lim_{s \searrow 0} P^0[\, B_t \neq 0 \text{ for all } s < t < \infty] \\ &\quad = \lim_{s \searrow 0} E^0\big[P^{B(s)}\{\, B_t \neq 0 \text{ for all } 0 < t < \infty\}\big] = 1. \end{aligned}$$

The last equality is true because $B(s) \neq 0$ with probability one.

Now we turn to the actual proofs, and we can assume $\mathbf{z} \neq 0$ and the small radius $r < |\mathbf{z}|$.

We start with dimension $d = 2$. The function $g(\mathbf{x}) = \log|\mathbf{x}|$ is harmonic in $D = \mathbf{R}^2 \setminus \{0\}$. (Only in $d = 2$, check.) Let

$$\sigma_R = \inf\{t \geq 0 : |B_t| \geq R\}$$

be the first time to exit the open ball of radius R. Pick $r < |\mathbf{z}| < R$, and define the annulus $A = \{\mathbf{x} : r < |\mathbf{x}| < R\}$. The time to exit the annulus is $\zeta = \tau_r \wedge \sigma_R$. Since any coordinate of B_t has limsup ∞ almost surely, or by the independent increments argument used in Example 6.10, the exit time σ_R and hence also ζ is finite almost surely.

Apply Corollary 6.7 to the harmonic function

$$f(\mathbf{x}) = \frac{\log R - \log|\mathbf{x}|}{\log R - \log r}$$

and the annulus A. We get that $f(B_{\zeta\wedge t})$ is a local L^2-martingale, and since f is bounded on the closure of A, it follows that $f(B_{\zeta\wedge t})$ is an L^2-martingale. The optional stopping argument used in Example 6.10, in conjunction with letting $t \to \infty$, gives again $f(\mathbf{z}) = E^{\mathbf{z}} f(B_0) = E^{\mathbf{z}} f(B_\zeta)$. Since $f = 1$ on the boundary of radius r but vanishes at radius R,

$$P^{\mathbf{z}}(\tau_r < \sigma_R) = P^{\mathbf{z}}(\,|B_\zeta| = r) = E^{\mathbf{z}} f(B_\zeta) = f(\mathbf{z}),$$

and so

$$P^{\mathbf{z}}(\tau_r < \sigma_R) = \frac{\log R - \log|\mathbf{z}|}{\log R - \log r}. \tag{6.22}$$

From this we can extract both part (a) and part (c) for $d = 2$. First, $\sigma_R \nearrow \infty$ as $R \nearrow \infty$ because a fixed path of Brownian motion is bounded on bounded time intervals. Consequently

$$\begin{aligned} P^{\mathbf{z}}(\tau_r < \infty) &= \lim_{R\to\infty} P^{\mathbf{z}}(\tau_r < \sigma_R) \\ &= \lim_{R\to\infty} \frac{\log R - \log|\mathbf{z}|}{\log R - \log r} = 1. \end{aligned}$$

Secondly, consider $r = r(k) = (1/k)^k$ and $R = R(k) = k$. Then we get

$$P^{\mathbf{z}}(\tau_{r(k)} < \sigma_{R(k)}) = \frac{\log k - \log|\mathbf{z}|}{(k+1)\log k}$$

which vanishes as $k \to \infty$. Let $\tau = \inf\{t \ge 0 : B_t = 0\}$ be the first hitting time of 0. For $0 < r < |\mathbf{z}|$, $\tau_r \le \tau$ because B_t cannot hit zero without first entering the ball of radius r. Again since $\sigma_{R(k)} \nearrow \infty$ as $k \nearrow \infty$,

$$\begin{aligned} P^{\mathbf{z}}(\tau < \infty) &= \lim_{k\to\infty} P^{\mathbf{z}}(\tau < \sigma_{R(k)}) \\ &\le \lim_{k\to\infty} P^{\mathbf{z}}(\tau_{r(k)} < \sigma_{R(k)}) = 0. \end{aligned}$$

For dimension $d \ge 3$ we use the harmonic function $g(\mathbf{x}) = |\mathbf{x}|^{2-d}$, and apply Itô's formula to the function

$$f(\mathbf{x}) = \frac{R^{2-d} - |\mathbf{x}|^{2-d}}{R^{2-d} - r^{2-d}}.$$

The annulus A and stopping times σ_R and ζ are defined as above. The same reasoning now leads to

$$P^{\mathbf{z}}(\tau_r < \sigma_R) = \frac{R^{2-d} - |\mathbf{z}|^{2-d}}{R^{2-d} - r^{2-d}}. \tag{6.23}$$

Letting $R \to \infty$ gives

$$P^{\mathbf{z}}(\tau_r < \infty) = \frac{|\mathbf{z}|^{2-d}}{r^{2-d}} = \left(\frac{r}{|\mathbf{z}|}\right)^{d-2}$$

as claimed. Part (c) follows now because the quantity above tends to zero as $r \to 0$.

It remains to show that after some finite time, the ball of radius r is no longer visited. Let $r < R$. Define $\sigma_R^1 = \sigma_R$, and for $n \geq 2$,

$$\tau_r^n = \inf\{t > \sigma_R^{n-1} : |B_t| \leq r\}$$

and

$$\sigma_R^n = \inf\{t > \tau_r^n : |B_t| \geq R\}.$$

In other words, $\sigma_R^1 < \tau_r^2 < \sigma_R^2 < \tau_r^3 < \cdots$ are the successive visits to radius R and back to radius r. Let $\alpha = (r/R)^{d-2} < 1$. We claim that for $n \geq 2$,

$$P^{\mathbf{z}}(\tau_r^n < \infty) = \alpha^{n-1}.$$

For $n = 2$, since $\sigma_R^1 < \tau_r^2$, use the strong Markov property to restart the Brownian motion at time σ_R^1.

$$\begin{aligned} P^{\mathbf{z}}(\tau_r^2 < \infty) &= P^{\mathbf{z}}(\sigma_R^1 < \tau_r^2 < \infty) \\ &= E^{\mathbf{z}}\big[P^{B(\sigma_R^1)}\{\tau_r < \infty\}\big] = \alpha. \end{aligned}$$

Then by induction.

$$\begin{aligned} P^{\mathbf{z}}(\tau_r^n < \infty) &= P^{\mathbf{z}}(\tau_r^{n-1} < \sigma_R^{n-1} < \tau_r^n < \infty) \\ &= E^{\mathbf{z}}\big[\mathbf{1}\{\tau_r^{n-1} < \sigma_R^{n-1} < \infty\}P^{B(\sigma_R^{n-1})}\{\tau_r < \infty\}\big] \\ &= P^{\mathbf{z}}(\tau_r^{n-1} < \infty) \cdot \alpha = \alpha^{n-1}. \end{aligned}$$

Above we used the fact that if $\tau_r^{n-1} < \infty$, then necessarily $\sigma_R^{n-1} < \infty$ because each coordinate of B_t has limsup ∞.

The claim implies

$$\sum_n P^{\mathbf{z}}(\tau_r^n < \infty) < \infty.$$

By Borel-Cantelli, $\tau_r^n < \infty$ can happen only finitely many times, almost surely. □

Remark 6.13. We see here an example of an uncountable family of events with probability one whose intersection must vanish. Namely, the intersection of the events that Brownian motion does not hit a particular point would be the event that Brownian motion never hits any point. Yet Brownian motion must reside somewhere.

Theorem 6.14. (Lévy's Characterization of Brownian Motion.) *Let* $M = [M_1, \dots, M_d]^T$ *be a continuous* $\mathbf{R}^d$*-valued local martingale and* $X(t) = M(t) - M(0)$. *Then* X *is a standard Brownian motion relative to* $\{\mathcal{F}_t\}$ *iff* $[X_i, X_j]_t = \delta_{i,j}t$. *In particular, in this case process* X *is independent of* $\mathcal{F}_0$.

Proof. We already know that a d-dimensional standard Brownian motion satisfies $[B_i, B_j]_t = \delta_{i,j}t$.

What we need to show is that X with the assumed covariance is Brownian motion. As a continuous local martingale, X is also a local L^2-martingale. Fix a vector $\theta = (\theta_1, \dots, \theta_d)^T \in \mathbf{R}^d$ and

$$f(t, \mathbf{x}) = \exp(i\theta^T \mathbf{x} + \tfrac{1}{2}|\theta|^2 t).$$

Let $Z_t = f(t, X(t))$. Itô's formula applies equally well to complex-valued functions, and so

$$\begin{aligned} Z_t &= 1 + \frac{|\theta|^2}{2} \int_0^t Z_s\, ds + \sum_{j=1}^d i\theta_j \int_0^t Z_s\, dX_j(s) \\ &\qquad - \frac{1}{2} \sum_{j=1}^d \theta_j^2 \int_0^t Z_s\, ds \\ &= 1 + \sum_{j=1}^d i\theta_j \int_0^t Z_s\, dX_j(s). \end{aligned}$$

This shows that Z is a local L^2-martingale. On any bounded time interval Z is bounded because the random factor $\exp(i\theta^T X(t))$ has absolute value one. Consequently Z is an L^2-martingale. For $s < t$, $E[Z_t|\mathcal{F}_s] = Z_s$ can be rewritten as

$$E\big[\exp\{i\theta^T(X(t) - X(s))\} \bigm| \mathcal{F}_s\big] = \exp\{-\tfrac{1}{2}|\theta|^2(t-s)\}. \tag{6.24}$$

By Lemma B.18 in the appendix, conditioned on $\mathcal{F}_s$, the increment $X(t) - X(s)$ has normal distribution with mean zero and covariance matrix identity. In particular, $X(t) - X(s)$ is independent of $\mathcal{F}_s$. Thus X has all the properties of Brownian motion relative to $\{\mathcal{F}_t\}$. □

Here is an application of Lévy's criterion.

Example 6.15. (Bessel processes.) Let $d \geq 2$ and $B(t) = [B_1(t), B_2(t), \dots, B_d(t)]^T$ a d-dimensional Brownian motion. Set

$$R_t = |B(t)| = \big(B_1(t)^2 + B_2(t)^2 + \cdots + B_d(t)^2\big)^{1/2}. \tag{6.25}$$

We find the semimartingale decomposition of R_t. Start $B(t)$ at a point $\mathbf{z} \neq 0$ so that $R_0 = |\mathbf{z}| > 0$. Let $D = \mathbf{R}^d \smallsetminus \{0\}$ and $f(\mathbf{x}) = |\mathbf{x}|$. Then $f \in C^2(D)$ with $f_{x_i} = x_i|\mathbf{x}|^{-1}$ and $\Delta f = (d-1)|\mathbf{x}|^{-1}$. By Proposition 6.12(c) with probability 1 the path $B[0,T]$ is a closed subset of D, for any $T < \infty$. Thus Itô's formula applies and gives

$$R_t = |\mathbf{z}| + \sum_{i=1}^d \int_0^t \frac{B_i(s)}{|B(s)|}\, dB_i(s) + \frac{d-1}{2} \int_0^t R_s^{-1} ds. \tag{6.26}$$

It turns out that the stochastic integral term is simply a 1-dimensional Brownian motion. Let

$$W_t = \sum_{i=1}^{d} \int_0^t \frac{B_i(s)}{|B(s)|}\, dB_i(s).$$

W_t is a continuous local L^2 martingale with quadratic variation

$$\begin{aligned}[W]_t &= \Big[\sum_{i=1}^{d} \int \frac{B_i}{|B|}\, dB_i \,,\, \sum_{j=1}^{d} \int \frac{B_j}{|B|}\, dB_j \Big]_t \\ &= \sum_{i,j} \int_0^t \frac{B_i B_j}{|B|^2}\, d[B_i, B_j] = \sum_i \int_0^t \frac{B_i(s)^2}{|B(s)|^2}\, ds = \int_0^t 1\, ds = t.\end{aligned}$$

Thus by Lévy's criterion W_t is a standard Brownian motion. We can rewrite (6.26) in the form

$$R_t = |\mathbf{z}| + \frac{d-1}{2} \int_0^t R_s^{-1} ds + W_t. \tag{6.27}$$

Process R_t is the *Bessel process* with dimension d, or with parameter $\frac{d-1}{2}$. By Proposition 6.12 we know that if $d = 2$ R_t returns infinitely often to any neighborhood of the origin, while if $d \geq 3$ then R_t drifts off to $+\infty$ as $t \to \infty$.

Next we prove a useful moment inequality. It is one case of the Burkholder-Davis-Gundy inequalities. Recall the notation $M_t^* = \sup_{0 \leq s \leq t} |M_s|$.

Proposition 6.16. *Let $p \in [2, \infty)$ and $C_p = (p(p-1)e)^{\frac{p}{2}}$. Then for all continuous local martingales M with $M_0 = 0$ and all $0 < t < \infty$,*

$$E[(M_t^*)^p] \leq C_p E\big([M]_t^{p/2}\big). \tag{6.28}$$

Proof. Check that $f(x) = |x|^p$ is a C^2 function on all of $\mathbf{R}$, with derivatives $f'(x) = \operatorname{sign}(x) p|x|^{p-1}$ and $f''(x) = p(p-1)|x|^{p-2}$. (The origin needs a separate check. The sign function is $\operatorname{sign}(x) = x/|x|$ for nonzero x, and here the convention at $x = 0$ is immaterial.) Version (6.6) of Itô's formula gives

$$|M_t|^p = \int_0^t \operatorname{sign}(M_s) p|M_s|^{p-1}\, dM_s + \tfrac{1}{2} p(p-1) \int_0^t |M_s|^{p-2}\, d[M]_s.$$

Assume M is bounded. Then M is an L^2 martingale and $M \in \mathcal{L}_2(M, \mathcal{P})$. Consequently the term $\int_0^t \operatorname{sign}(M_s)p|M_s|^{p-1}\,dM_s$ is a mean zero L^2 martingale. Take expectations and apply Hölder's inequality:

$$\begin{aligned}
E\bigl(|M_t|^p\bigr) &= \tfrac{1}{2}p(p-1)E\int_0^t |M_s|^{p-2}\,d[M]_s \\
&\le \tfrac{1}{2}p(p-1)E\bigl((M_t^*)^{p-2}[M]_t\bigr) \\
&\le \tfrac{1}{2}p(p-1)\Bigl\{E\bigl((M_t^*)^p\bigr)\Bigr\}^{1-\frac{2}{p}}\Bigl\{E\bigl([M]_t^{p/2}\bigr)\Bigr\}^{\frac{2}{p}}.
\end{aligned}$$

Combining the above with Doob's inequality (3.11) gives

$$E[(M_t^*)^p] \le 2eE\bigl(|M_t|^p\bigr) \le ep(p-1)\Bigl\{E\bigl((M_t^*)^p\bigr)\Bigr\}^{1-\frac{2}{p}}\Bigl\{E\bigl([M]_t^{p/2}\bigr)\Bigr\}^{\frac{2}{p}}.$$

Rearranging the above inequality gives the conclusion for a bounded martingale.

The general case comes by localization. Let $\tau_k = \inf\{t \ge 0 : |M_t| \ge k\}$, a stopping time by Corollary 2.10. By continuity M^{τ_k} is bounded, so we can apply the already proved result to claim

$$E[(M^*_{\tau_k \wedge t})^p] \le C_p E\bigl([M]^{p/2}_{\tau_k \wedge t}\bigr).$$

On both sides monotone convergence leads to (6.28) as $k \nearrow \infty$. $\square$

Exercises

Exercise 6.1. (a) Let Y be a cadlag semimartingale with $Y_0 = 0$. Check that Itô's formula gives $2\int Y_- dY = Y^2 - [Y]$, as already follows from the intregration by parts formula (5.67).

(b) Let N be a rate α Poisson process and $M_t = N_t - \alpha t$. Apply part (a) to show that $M_t^2 - N_t$ is a martingale.

Exercise 6.2. This exercise gives an instance of an important connection between stochastic analysis and partial differential equations known as the *Feynman-Kac formula*. Let h and V be bounded continuous functions on $\mathbf{R}^d$. Suppose u is continuous on $\mathbf{R}_+ \times \mathbf{R}^d$, bounded on $[0,T] \times \mathbf{R}^d$ for each $T < \infty$, and $u \in C^{1,2}((0,\infty) \times \mathbf{R}^d)$. Let u satisfy the initial value problem

$$\begin{aligned}
&\frac{\partial}{\partial t}u(t,x) = \tfrac{1}{2}\Delta u(t,x) + V(x)u(t,x) \quad \text{on } (0,\infty)\times\mathbf{R}^d \\
&u(0,x) = h(x) \quad x \in \mathbf{R}^d.
\end{aligned} \tag{6.29}$$

Show that

$$u(t,x) = E^x\bigl[h(B(t))\,e^{\int_0^t V(B(s))\,ds}\bigr] \tag{6.30}$$

where $B(\cdot)$ is Brownian motion on $\mathbf{R}^d$ and E^x is expectation under the path measure of $B(\cdot)$ started at $B(0) = x$. *Hint.* Consider the process $Z_t = u(t_0 - t, B(t))e^{\int_0^t V(B(s))\,ds}$ for $t \in [0, t_0 - \varepsilon]$, $\varepsilon > 0$.

Exercise 6.3. Let $a < 0 < b$, $\theta \in \mathbf{R}$, and

$$X_t = f(t)\cosh\bigl(\theta B_t - \theta \tfrac{a+b}{2}\bigr)$$

where B_t is a standard Brownian motion on $\mathbf{R}$ and $f(t)$ is a deterministic function.

(a) Use Itô's formula to find $f(t)$ such that X_t is a local L^2 martingale. Show that X_t is actually a martingale.

(b) Find the Laplace transform of the hitting time $\tau = \inf\{t \geq 0 : B_t = a \text{ or } B_t = b\}$. Namely,

$$E(e^{-\lambda\tau}) = \frac{\cosh(-\sqrt{2\lambda}\frac{a+b}{2})}{\cosh(\sqrt{2\lambda}\frac{b-a}{2})} \quad \text{for } \lambda \geq 0.$$

Exercise 6.4. Let B_t be standard Brownian motion and $M_t = B_t^2 - t$.

(a) Justify the identity $[M] = [B^2]$. (Use the linearity (5.64).)

(b) Apply Itô's formula to B_t^2 and from that find a representation for $[B^2]$ in terms of a single ds-integal.

(c) Using Itô's formula and your answer to (b), check that $M^2 - [B^2]$ is a martingale, as it should. (Of course without appealing to the fact that $M^2 - [B^2] = M^2 - [M]$ is a martingale.)

(d) Use integration by parts (5.67) to find another representation of $[B^2]$. Then use Itô's formula to show that this representation agrees with the answer in (b).

Exercise 6.5. (a) Check that for a rate α Poisson process N Itô's formula reduces to an obvious identity, namely a telescoping sum over jumps that can be written as

$$f(N_t) = f(0) + \int_{(0,t]} \bigl(f(N_s) - f(N_{s-})\bigr)\,dN_s. \tag{6.31}$$

(b) Suppose $E\int_0^T |f(N_s)|^2\,ds < \infty$. Show that for $t \in [0, T]$

$$E[f(N_t)] = f(0) + E\int_0^t \bigl(f(N_s + 1) - f(N_s)\bigr)\,\alpha\,ds. \tag{6.32}$$

Warning. Do not attempt to apply stochastic calculus to non-predictable integrands.

Exercise 6.6. Suppose X is a nondecreasing cadlag process such that $X(0) = 0$, all jumps are of size 1 (that is, $X_s - X_{s-} = 0$ or 1), between jumps X is constant, and $M_t = X_t - \alpha t$ is a martingale. Show that X is a rate α Poisson process. *Hint.* Imitate the proof of Theorem 6.14.

Exercise 6.7. Suppose X and Y are adapted cadlag processes, X is $\mathbf{Z}_+$-valued and its jumps are of size 1, and $X(0) = 0$. (Such an X is called a counting process.) Suppose

$$M_t = X_t - \int_0^t Y_s \, ds$$

is a martingale. Show that

$$M_t^2 - \int_0^t Y_s \, ds$$

is also a martingale. That is, $\langle M \rangle_t = \int_0^t Y_s \, ds$. *Hint.* Apply Itô's formula to M_t^2.

Exercise 6.8. Let B be standard one-dimensional Brownian motion. Show that for a continuously differentiable nonrandom function ϕ,

$$\int_0^t \phi(s) \, dB_s = \phi(t) B_t - \int_0^t B_s \phi'(s) \, ds.$$

Exercise 6.9. Define

$$V_1(t) = \int_0^t B_s \, ds, \quad V_2(t) = \int_0^t V_1(s) \, ds,$$

and generally

$$V_n(t) = \int_0^t V_{n-1}(s) \, ds = \int_0^t ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \, B_{s_n}.$$

V_n is known as n times integrated Brownian motion, and appears in applications in statistics. Show that

$$V_n(t) = \frac{1}{n!} \int_0^t (t-s)^n \, dB_s.$$

Then show that the process

$$M_n(t) = V_n(t) - \sum_{j=1}^{n} \frac{t^j}{j!(n-j)!} \int_0^t (-s)^{n-j} \, dB_s$$

is a martingale.

Exercise 6.10. Let X and Y be independent rate α Poisson processes.

(a) Show that

$$P\{X \text{ and } Y \text{ have no jumps in common}\} = 1.$$

Find the covariation $[X, Y]$.

(b) Find a process U such that $X_tY_t - U_t$ is a martingale.

Exercise 6.11. Let $D = \mathbf{R}^3 \smallsetminus \{0\}$, the complement of the origin in $\mathbf{R}^3$. Define $f \in C^2(D)$ by

$$f(\mathbf{x}) = \frac{1}{|\mathbf{x}|} = \frac{1}{(x_1^2 + x_2^2 + x_3^2)^{1/2}}.$$

Let $B(t) = (B_1(t), B_2(t), B_3(t))$ be standard Brownian motion in $\mathbf{R}^3$. Show that $X_t = f(B_{1+t})$ is an L^2-bounded, hence uniformly integrable, local L^2 martingale, but not a martingale. L^2-bounded means that $\sup_{t\ge 0} E[X_t^2] < \infty$. *Hint.* To show that $X[0,T]$ is a closed subset of D use $P^{\mathbf{z}}\{B_t \ne 0 \ \forall t\} = 1$ from Proposition 6.12. Itô's formula can be applied. If X_t were a martingale, EX_t would have to be constant in time.

Exercise 6.12. Let B_t be Brownian motion in $\mathbf{R}^k$ started at point $\mathbf{z} \in \mathbf{R}^k$. As before, let

$$\sigma_R = \inf\{t \ge 0 : |B_t| \ge R\}$$

be the first time the Brownian motion leaves the ball of radius R. Compute the expectation $E^{\mathbf{z}}[\sigma_R]$ as a function of $\mathbf{z}$, k and R.

Hint. Start by applying Itô's formula to $f(\mathbf{x}) = x_1^2 + \cdots + x_k^2$.

Exercise 6.13 (Hölder continuity of a stochastic integral)**.** Let X be an adapted measurable process, $T < \infty$, and assume that

$$\sup_{t\in[0,T]} E(|X_t|^p) < \infty \tag{6.33}$$

for some $2 < p < \infty$. How much Hölder continuity can you get for the paths of the process $M_t = \int_0^t X_s\, dB_s$ from a combination of inequality (6.28) and Theorem B.20? What if (6.33) holds for all $p < \infty$?

Exercise 6.14 (Fisk-Stratonovich integral)**.** Let $X = X_0 + M + U$ and $Y = Y_0 + N + V$ be two semimartingales with continuous local martingale parts M and N and continuous finite variation parts U and V. The *Fisk-Stratonovich* integral is defined by

$$\int_0^t Y_s \circ dX_s = \int_0^t Y_s\, dX_s + \tfrac{1}{2}[X,Y]_t \tag{6.34}$$

where the first integral on the right-hand side is the Itô integral from Definition 5.35. Since

$$\begin{aligned}\sum_{i=1}^n &\tfrac{Y_{t_i}+Y_{t_{i+1}}}{2}(X_{t_{i+1}} - X_{t_i})\\ &= \sum_{i=1}^n Y_{t_i}(X_{t_{i+1}} - X_{t_i}) + \tfrac{1}{2}\sum_{i=1}^n (Y_{t_{i+1}} - Y_{t_i})(X_{t_{i+1}} - X_{t_i}),\end{aligned}$$

Propositions 5.37 and 5.64 give the limit

$$\int_0^t Y_s \circ dX_s = \lim_{\text{mesh}(\pi)\to 0} \sum_{i=1}^{n} \tfrac{Y_{t_i}+Y_{t_{i+1}}}{2}(X_{t_{i+1}} - X_{t_i}) \tag{6.35}$$

in probability, uniformly over compact time intervals.

Let $X(t) = (X_1(t), \dots, X_d(t))$ be a vector-valued semimartingale, with continuous local martingale and finite variation parts, and $f \in C^3(\mathbf{R}^d)$. Show that

$$f(X(t)) = f(X(0)) + \sum_{i=1}^{d} \int_0^t f_{x_i}(X(s)) \circ dX_i(s). \tag{6.36}$$

In other words, the Fisk-Stratonovich integral obeys the normal rules of calculus, without the second-order correction term needed for Itô's formula. *Hint.* Do not attempt to prove this from first principles. Use Itô's formula and our knowledge of quadratic variation.

Chapter 7

Stochastic Differential Equations

In this chapter we study equations of the type

$$X(t,\omega) = H(t,\omega) + \int_{(0,t]} F(s,\omega,X(\omega))\,dY(s,\omega) \tag{7.1}$$

for an unknown $\mathbf{R}^d$-valued process X. In the equation, Y is a given $\mathbf{R}^m$-valued cadlag semimartingale and H is a given $\mathbf{R}^d$-valued adapted cadlag process. The coefficient $F(t,\omega,\eta)$ is a $d \times m$-matrix valued function of the time variable t, the sample point ω, and a cadlag path η. The integral $\int F\,dY$ with a matrix-valued integrand and vector-valued integrator has a natural componentwise interpretation, as explained in Remark 5.38.

Underlying the equation is a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$, on which the ingredients H, Y and F are defined. A *solution* X is an $\mathbf{R}^d$-valued cadlag process that is defined on this given probability space $(\Omega, \mathcal{F}, P)$ and adapted to $\{\mathcal{F}_t\}$, and that satisfies the equation in the sense that the two sides of (7.1) are indistinguishable. This notion is the so-called *strong solution* which means that the solution process can be constructed on the given probability space, adapted to the given filtration. Later we shall also address *weak solutions* where the probability space and the filtration are part of the solution, not prescribed ahead of time.

In order for the integral $\int F\,dY$ to be sensible, $F(t,\omega,\eta)$ has to have the property that, whenever an adapted cadlag process X is substituted for η, the resulting process $(t,\omega) \mapsto F(t,\omega,X(\omega))$ is predictable and locally bounded in the sense (5.36). We list precise assumptions on F when we state and prove an existence and uniqueness theorem.

Since the integral term vanishes at time zero, equation (7.1) contains the initial value $X(0) = H(0)$. The integral equation (7.1) can be written in the differential form

$$dX(t) = dH(t) + F(t, X)\,dY(t), \quad X(0) = H(0) \tag{7.2}$$

where the initial value must then be displayed explicitly. The notation can be further simplified by dropping the superfluous time variables:

$$dX = dH + F(t, X)\,dY, \quad X(0) = H(0). \tag{7.3}$$

Equations (7.2) and (7.3) have no other interpretation except as abbreviations for (7.1). Equations such as (7.3) are known as SDEs (stochastic differential equations) even though precisely speaking they are integral equations.

7.1. Examples of stochastic equations and solutions

7.1.1. Itô equations. Let B_t be a standard Brownian motion in $\mathbf{R}^m$ with respect to a filtration $\{\mathcal{F}_t\}$. Let ξ be an $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable. Often ξ is a nonrandom point x_0. The assumption of $\mathcal{F}_0$-measurability implies that ξ is independent of the Brownian motion. An Itô equation has the form

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \quad X_0 = \xi \tag{7.4}$$

or in integral form

$$X_t = \xi + \int_0^t b(s, X_s)\,ds + \int_0^t \sigma(s, X_s)\,dB_s. \tag{7.5}$$

The coefficients $b(t, x)$ and $\sigma(t, x)$ are Borel measurable functions of $(t, x) \in [0, \infty) \times \mathbf{R}^d$. The *drift vector* $b(t, x)$ is $\mathbf{R}^d$-valued, and the *dispersion matrix* $\sigma(t, x)$ is $d \times m$-matrix valued. The $d \times d$ matrix $a(t, x) = \sigma(t, x)\sigma(t, x)^T$ is called the *diffusion matrix*. X is the unknown $\mathbf{R}^d$-valued process.

Linear Itô equations can be explicitly solved by a technique from basic theory of ODEs (ordinary differential equations) known as the *integrating factor*. See Exercise 7.2 for the general case. Here in the text we go through important special cases that describe commonly studied processes. We begin with the nonrandom case from elementary calculus.

Example 7.1. (Integrating factor for linear ODEs.) Let $a(t)$ and $g(t)$ be given functions. Suppose we wish to solve

$$x' + a(t)x = g(t) \tag{7.6}$$

for an unknown function $x = x(t)$. The trick is to multiply through the equation by the integrating factor

$$z(t) = e^{\int_0^t a(s)\,ds} \tag{7.7}$$

and then identify the left-hand side $zx' + azx$ as the derivative $(zx)'$. The equation becomes

$$\frac{d}{dt}\left(x(t)e^{\int_0^t a(s)\,ds}\right) = g(t)e^{\int_0^t a(s)\,ds}.$$

Integrating from 0 to t gives

$$x(t)e^{\int_0^t a(s)\,ds} - x(0) = \int_0^t g(s)e^{\int_0^s a(u)\,du}\,ds$$

which rearranges into

$$x(t) = x(0)e^{-\int_0^t a(s)\,ds} + e^{-\int_0^t a(s)\,ds}\int_0^t g(s)e^{\int_0^s a(u)\,du}\,ds.$$

Now one can check by differentiation that this formula gives a solution.

We apply this idea to the most basic Itô equation.

Example 7.2. (Ornstein-Uhlenbeck process.) Let $\alpha > 0$ and $0 \le \sigma < \infty$ be constants, and consider the SDE

$$dX_t = -\alpha X_t\,dt + \sigma\,dB_t \tag{7.8}$$

with a given initial value X_0 independent of the one-dimensional Brownian motion B_t. Formal similarity with the linear ODE (7.6) suggests we try the integrating factor $Z_t = e^{\alpha t}$. For the ODE the key was to take advantage of the formula $d(zx) = x\,dz + z\,dx$ for the derivative of a product. Here we attempt the same, but differentiation rules from calculus have to be replaced by Itô's formula. The integration by parts rule which is a special case of Itô's formula gives

$$d(ZX) = Z\,dX + X\,dZ + d[Z, X].$$

The definition of Z gives $dZ = \alpha Z\,dt$. Since Z is a continuous FV process, $[Z, X] = 0$. *Assuming* X satisfies (7.8), we get

$$d(ZX) = -\alpha ZX\,dt + \sigma Z\,dB + \alpha ZX\,dt = \sigma Z\,dB$$

which in integrated form becomes (note $Z_0 = 1$)

$$Z_t X_t = X_0 + \sigma\int_0^t e^{\alpha s}\,dB_s$$

and we can solve for X to get

$$X_t = X_0 e^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}\,dB_s. \tag{7.9}$$

Since we arrived at this formula by assuming the solution of the equation, now we turn around and verify that (7.9) defines a solution of (7.8). This is a straightforward Itô calculation. It is conveniently carried out in terms of formal stochastic differentials (but could just as well be recorded in terms of integrals):

$$\begin{aligned} dX &= -\alpha X_0 e^{-\alpha t} dt - \alpha\sigma e^{-\alpha t}\biggl(\int_0^t e^{\alpha s}\,dB\biggr)\,dt + \sigma e^{-\alpha t} e^{\alpha t}\,dB \\ &\qquad + d\Bigl[\alpha\sigma e^{-\alpha t}, \int e^{\alpha s}\,dB\Bigr] \\ &= -\alpha\biggl(X_0 e^{-\alpha t} + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}\,dB\biggr)\,dt + \sigma\,dB \\ &= -\alpha X\,dt + \sigma\,dB. \end{aligned}$$

The quadratic covariation vanishes by Lemma A.10 because $\alpha\sigma e^{-\alpha t}$ is a continuous BV function. Notice that we engaged in slight abuse of notation by putting $\alpha\sigma e^{-\alpha t}$ inside the brackets, though of course we do not mean the point value but rather the process, or function of t. By the same token, $\int e^{\alpha s}\,dB$ stands for the process $\int_0^t e^{\alpha s}\,dB$.

The process defined by the SDE (7.8) or by the formula (7.9) is known as the Ornstein-Uhlenbeck process.

If the equation had no noise ($\sigma = 0$) the solution would be $X_t = X_0 e^{-\alpha t}$ which simply decays to 0 as $t \to \infty$. Let us observe that in contrast to this, when the noise is turned on ($\sigma > 0$), the process X_t is not driven to zero but instead to a nontrivial statistical steady state.

Consider first the random variable $Y_t = \sigma e^{-\alpha t}\int_0^t e^{\alpha s}\,dB_s$ that captures the dynamical noise part in the solution (7.9). Y_t is a mean zero martingale. From the construction of stochastic integrals we know that Y_t is a limit (in probability, or almost sure along a particular sequence) of sums:

$$Y_t = \lim_{\mathrm{mesh}(\pi)\to 0} \sigma e^{-\alpha t} \sum_{k=0}^{m(\pi)-1} e^{\alpha s_k}(B_{s_{k+1}} - B_{s_k}).$$

From this we can find the distribution of Y_t by computing its characteristic function. By the independence of Brownian increments, for $\theta \in \mathbf{R}$,

$$\begin{aligned} E(e^{i\theta Y_t}) &= \lim_{\mathrm{mesh}(\pi)\to 0} E\Bigl[\exp\Bigl\{i\theta\sigma e^{-\alpha t}\sum_{k=0}^{m(\pi)-1} e^{\alpha s_k}(B_{s_{k+1}} - B_{s_k})\Bigr\}\Bigr] \\ &= \lim_{\mathrm{mesh}(\pi)\to 0} \exp\Bigl\{-\tfrac{1}{2}\theta^2\sigma^2 e^{-2\alpha t}\sum_{k=0}^{m(\pi)-1} e^{2\alpha s_k}(s_{k+1} - s_k)\Bigr\} \\ &= \exp\Bigl(-\tfrac{1}{2}\theta^2\sigma^2 e^{-2\alpha t}\int_0^t e^{2\alpha s}\,ds\Bigr) = \exp\Bigl(-\frac{\theta^2\sigma^2}{4\alpha}(1 - e^{-2\alpha t})\Bigr). \end{aligned}$$

This says that Y_t is a mean zero Gaussian with variance $\frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t})$. In the $t \to \infty$ limit Y_t converges weakly to the $\mathcal{N}(0, \frac{\sigma^2}{2\alpha})$ distribution. Then from (7.9) we see that this distribution is the weak limit of X_t as $t \to \infty$.

Suppose we take the initial point $X_0 \sim \mathcal{N}(0, \frac{\sigma^2}{2\alpha})$, independent of B. Then (7.9) represents X_t as the sum of two independent mean zero Gaussians, and we can add the variances to see that for each $t \in \mathbf{R}_+$, $X_t \sim \mathcal{N}(0, \frac{\sigma^2}{2\alpha})$.

X_t is in fact a Markov process. This can be seen from

$$X_{t+s} = e^{-\alpha s} X_t + \sigma e^{-\alpha(t+s)} \int_t^{t+s} e^{\alpha u}\, dB_u.$$

Our computations above show that as $t \to \infty$, for every choice of the initial state X_0, X_t converges weakly to its invariant distribution $\mathcal{N}(0, \frac{\sigma^2}{2\alpha})$.

We continue with another example of a linear equation where the integrating factor needs to take into consideration the stochastic part.

Example 7.3. (Geometric Brownian motion.) Let μ, σ be constants, B one-dimensional Brownian motion, and consider the SDE

$$dX = \mu X\, dt + \sigma X\, dB \tag{7.10}$$

with a given initial value X_0 independent of the Brownian motion. Since the equation rewrites as

$$dX - X(\mu\, dt + \sigma\, dB) = 0$$

a direct adaptation of (7.7) would suggest multiplication by $\exp(-\mu t - \sigma B_t)$. The reader is invited to try this. Some trial and error with Itô's formula reveals that the quadratic variation must be taken into account, and the useful integrating factor turns out to be

$$Z_t = \exp\bigl(-\mu t - \sigma B_t + \tfrac{1}{2}\sigma^2 t\bigr).$$

Itô's formula gives first

$$dZ = (-\mu + \sigma^2) Z\, dt - \sigma Z\, dB. \tag{7.11}$$

(Note that a second $\sigma^2/2$ factor comes from the quadratic variation term.) Then, assuming X satisfies (7.10), one can check that $d(ZX) = 0$. For this calculation, note that the only nonzero contribution to the covariation of X and Z comes from the dB-terms of their semimartingale representations (7.11) and (7.10):

$$\begin{aligned} [Z, X]_t &= \Bigl[-\sigma \int Z\, dB\,,\, \sigma \int X\, dB\Bigr]_t \\ &= -\sigma^2 \int_0^t X_s Z_s\, d[B, B]_s = -\sigma^2 \int_0^t X_s Z_s\, ds. \end{aligned}$$

Above we used Proposition 5.55. In differential form the above is simply $d[Z,X] = -\sigma^2 ZX\,dt$.

Continuing with the solution of equation (7.10), $d(ZX) = 0$ and $Z_0 = 1$ give $Z_t X_t = Z_0 X_0 = X_0$, from which $X_t = X_0 Z_t^{-1}$. More explicitly, our tentative solution is

$$X_t = X_0 \exp\bigl\{(\mu - \tfrac{1}{2}\sigma^2)t + \sigma B_t\bigr\}$$

whose correctness can again be verified with Itô's formula. This process is called geometric Brownian motion, although some texts reserve the term for the case $\mu = 0$.

Solutions of the previous two equations are defined for all time. In the next example this is not the case.

Example 7.4. (Brownian bridge.) Fix $0 < T < 1$. The SDE is now

$$dX = -\frac{X}{1-t}dt + dB, \quad \text{for } 0 \le t \le T, \text{ with } X_0 = 0. \tag{7.12}$$

The integrating factor

$$Z_t = \exp\Bigl(\int_0^t \frac{ds}{1-s}\Bigr) = \frac{1}{1-t}$$

works, and we arrive at the solution

$$X_t = (1-t)\int_0^t \frac{1}{1-s}\,dB_s. \tag{7.13}$$

To check that this solves (7.12), apply the product formula $d(UV) = U\,dV + V\,dU + d[U,V]$ with $U = 1-t$ and $V = \int (1-s)^{-1}\,dB_s = (1-t)^{-1}X$. With $dU = -dt$, $dV = (1-t)^{-1}\,dB$ and $[U,V] = 0$ this gives

$$\begin{aligned} dX = d(UV) &= (1-t)(1-t)^{-1}\,dB - (1-t)^{-1}X\,dt \\ &= dB - (1-t)^{-1}X\,dt \end{aligned}$$

which is exactly (7.12).

The solution (7.13) is defined for $0 \le t < 1$. One can show that it converges to 0 as $t \nearrow 1$, along almost every path of B_s (Exercise 7.1).

7.1.2. Stochastic exponential. The exponential function $g(t) = e^{ct}$ can be characterized as a the unique function g that satisfies the equation

$$g(t) = 1 + c\int_0^t g(s)\,ds.$$

The stochastic exponential generalizes the dt-integral to a semimartingale integral.

Theorem 7.5. *Let Y be a real-valued cadlag semimartingale such that $Y_0 = 0$. Define*

$$Z_t = \exp\{Y_t - \tfrac{1}{2}[Y]_t\} \prod_{s\in(0,t]} (1 + \Delta Y_s) \exp\{-\Delta Y_s + \tfrac{1}{2}(\Delta Y_s)^2\}. \tag{7.14}$$

Then the process Z is a cadlag semimartingale, and it is the unique cadlag semimartingale Z that satisfies the equation

$$Z_t = 1 + \int_{(0,t]} Z_{s-}\, dY_s. \tag{7.15}$$

Proof. The uniqueness of the solution of (7.15) will follow from the general uniqueness theorem for solutions of semimartingale equations.

We start by showing that Z defined by (7.14) is a semimartingale. The definition can be reorganized as

$$Z_t = \exp\biggl\{Y_t - \tfrac{1}{2}[Y]_t + \tfrac{1}{2}\sum_{s\in(0,t]} (\Delta Y_s)^2\biggr\} \prod_{s\in(0,t]} (1 + \Delta Y_s) \exp\{-\Delta Y_s\}.$$

The continuous part

$$[Y]^c_t = [Y]_t - \sum_{s\in(0,t]} (\Delta Y_s)^2$$

of the quadratic variation is an increasing process (cadlag, nondecreasing), hence an FV process. Since e^x is a C^2 function, the factor

$$\exp(W_t) \equiv \exp\biggl\{Y_t - \tfrac{1}{2}[Y]_t + \tfrac{1}{2}\sum_{s\in(0,t]} (\Delta Y_s)^2\biggr\} \tag{7.16}$$

is a semimartingale by Itô's formula. It remains to show that for a fixed ω the product

$$U_t \equiv \prod_{s\in(0,t]} (1 + \Delta Y_s) \exp\{-\Delta Y_s\} \tag{7.17}$$

converges and has paths of bounded variation. The part

$$\prod_{s\in(0,t]:|\Delta Y(s)|\ge 1/2} (1 + \Delta Y_s) \exp\{-\Delta Y_s\}$$

has only finitely many factors because a cadlag path has only finitely many jumps exceeding a given size in a bounded time interval. Hence this part is piecewise constant, cadlag and in particular FV. Let $\xi_s = \Delta Y_s \mathbf{1}_{\{|\Delta Y(s)|<1/2\}}$ denote a jump of magnitude below $\frac{1}{2}$. It remains to show that

$$H_t = \prod_{s\in(0,t]} (1 + \xi_s) \exp\{-\xi_s\} = \exp\biggl\{\sum_{s\in(0,t]} \bigl(\log(1 + \xi_s) - \xi_s\bigr)\biggr\}$$

is an FV process. For $|x| < 1/2$,

$$\bigl|\log(1+x) - x\bigr| \le x^2.$$

Hence the sum above is absolutely convergent:

$$\sum_{s\in(0,t]} \bigl|\log(1+\xi_s) - \xi_s\bigr| \le \sum_{s\in(0,t]} \xi_s^2 \le [Y]_t < \infty.$$

It follows that $\log H_t$ is a cadlag FV process (see Example 1.13 in this context). Since the exponential function is locally Lipschitz, $H_t = \exp(\log H_t)$ is also a cadlag FV process (Lemma A.6).

To summarize, we have shown that e^{W_t} in (7.16) is a semimartingale and U_t in (7.17) is a well-defined real-valued FV process. Consequently $Z_t = e^{W_t} U_t$ is a semimartingale.

The second part of the proof is to show that Z satisfies equation (7.15). Let $f(w,u) = e^w u$, and find

$$f_w = f,\ f_u = e^w,\ f_{uu} = 0,\ f_{uw} = e^w \text{ and } f_{ww} = f.$$

Note that $\Delta W_s = \Delta Y_s$ because the jump in $[Y]$ at s equals exactly $(\Delta Y_s)^2$. A straight-forward application of Itô gives

$$\begin{aligned} Z_t = 1 + \int_{(0,t]} e^{W(s-)}\, dU_s &+ \int_{(0,t]} Z_{s-}\, dW_s \\ &+ \tfrac{1}{2}\int_{(0,t]} Z_{s-}\, d[W]_s + \int_{(0,t]} e^{W(s-)}\, d[W,U]_s \\ + \sum_{s\in(0,t]} \Bigl\{ \Delta Z_s - Z_{s-}\Delta Y_s &- e^{W(s-)}\Delta U_s - \tfrac{1}{2} Z_{s-}(\Delta Y_s)^2 - e^{W(s-)}\Delta Y_s \Delta U_s \Bigr\}. \end{aligned}$$

Since

$$W_t = Y_t - \tfrac{1}{2}\Bigl([Y]_t - \sum_{s\in(0,t]} (\Delta Y_s)^2\Bigr)$$

where the part in parentheses is FV and continuous, linearity of covariation and Lemma A.10 imply $[W] = [Y]$ and $[W,U] = [Y,U] = \sum \Delta Y \Delta U$.

Now match up and cancel terms. First,

$$\int_{(0,t]} e^{W(s-)}\, dU_s - \sum_{s\in(0,t]} e^{W(s-)} \Delta U_s = 0$$

because U is an FV process whose paths are step functions so the integral reduces to a sum over jumps. Second,

$$\int_{(0,t]} Z_{s-}\,dW_s + \tfrac{1}{2}\int_{(0,t]} Z_{s-}\,d[W]_s - \sum_{s\in(0,t]} Z_{s-}(\Delta Y_s)^2$$
$$= \int_{(0,t]} Z_{s-}\,dY_s$$

by the definition of W and the observation $[W]=[Y]$ from above. Then

$$\int_{(0,t]} e^{W(s-)}\,d[W,U]_s - \sum_{s\in(0,t]} e^{W(s-)}\Delta Y_s \Delta U_s = 0$$

by the observation $[W,U]=\sum \Delta Y \Delta U$ from above. Lastly, $\Delta Z_s - Z_{s-}\Delta Y_s = 0$ directly from the definition of Z.

After the cancelling we are left with

$$Z_t = 1 + \int_{(0,t]} Z_{s-}\,dY_s,$$

the desired equation. □

The semimartingale Z defined in the theorem is called the *stochastic exponential* of Y, and denoted by $\mathcal{E}(Y)$.

Example 7.6. Let $Y_t = \lambda B_t$ where B is Brownian motion. The stochastic exponential $Z = \mathcal{E}(\lambda B)$, given by $Z_t = \exp(\lambda B_t - \frac{1}{2}\lambda^2 t)$, another instance of geometric Brownian motion. The equation

$$Z_t = 1 + \lambda \int_0^t Z_s\,dB_s$$

and moment bounds show that geometric Brownian motion is a continuous L^2-martingale.

7.2. Itô equations

We present the existence and uniqueness results for SDEs in two stages. This section covers the standard strong existence and uniqueness result for Itô equations contained in every stochastic analysis book. In addition we prove weak uniqueness and the strong Markov property of the solution. In the subsequent two sections we undertake a much more technical existence and uniqueness proof for a cadlag semimartingale equation. On $\mathbf{R}^d$ we use the Euclidean norm $|x| = (\sum |x_i|^2)^{1/2}$, and similarly for a matrix $A = (a_{i,j})$ the norm is $|A| = (\sum |a_{i,j}|^2)^{1/2}$.

Let B_t be a standard $\mathbf{R}^m$-valued Brownian motion with respect to the complete filtration $\{\mathcal{F}_t\}$ on the complete probability space $(\Omega, \mathcal{F}, P)$. The given data are coefficient functions $b : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ and $\sigma : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^{d\times m}$ and an $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable ξ that gives the initial position. The assumptions imply that the initial position ξ is independent of the driving Brownian motion B.

The goal is to prove that on $(\Omega, \mathcal{F}, P)$ there exists an $\mathbf{R}^d$-valued process X that is adapted to $\{\mathcal{F}_t\}$ and satisfies the integral equation

$$X_t = \xi + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{7.18}$$

in the sense that the processes on the right and left of the equality sign are indistinguishable. Part of the requirement is that the integrals are well-defined, for which we require that

$$P\Big\{\forall\, T < \infty : \int_0^T |b(s, X_s)|\, ds + \int_0^T |\sigma(s, X_s)|^2\, ds < \infty\Big\} = 1. \tag{7.19}$$

Such a process X on $(\Omega, \mathcal{F}, P)$ is called a *strong solution* because it exists on the given probability space and is adapted to the given filtration. Strong uniqueness then means that if Y is another process adapted to $\{\mathcal{F}_t\}$ that satisfies (7.18)– (7.19), then X and Y are indistinguishable.

Next we state the standard Lipschitz assumption on the coefficient functions and then the strong existence and uniqueness theorem.

Assumption 7.7. Assume the Borel functions $b : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ and $\sigma : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^{d\times m}$ satisfy the Lipschitz condition

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \le L|x - y| \tag{7.20}$$

and the bound

$$|b(t, x)| + |\sigma(t, x)| \le L(1 + |x|) \tag{7.21}$$

for a constant $L < \infty$ and all $t \in \mathbf{R}_+$ and $x, y \in \mathbf{R}^d$.

Theorem 7.8. *Let Assumption 7.7 be in force and let ξ be an $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable. Then there exists a continuous process X on $(\Omega, \mathcal{F}, P)$ adapted to $\{\mathcal{F}_t\}$ that satisfies integral equation (7.18) and property (7.19). The process X with these properties is unique up to indistinguishability.*

Before turning to the proofs we present simple examples from ODE theory that illustrate the loss of existence and uniqueness when the Lipschitz assumption on the coefficient is weakened.

Example 7.9. (a) Consider the equation

$$x(t) = \int_0^t 2\sqrt{x(s)}\, ds.$$

The function $b(x) = \sqrt{x}$ is not Lipschitz on $[0,1]$ because $b'(x)$ blows up at the origin. The equation has infinitely many solutions. Two of them are $x(t) = 0$ and $x(t) = t^2$.

(b) The equation

$$x(t) = 1 + \int_0^t x^2(s)\,ds$$

does not have a solution for all time. The unique solution starting at $t = 0$ is $x(t) = (1-t)^{-1}$ which exists only for $0 \le t < 1$. The function $b(x) = x^2$ is locally Lipschitz. This is enough for uniqueness, as determined in Exercise 7.4.

We prove Theorem 7.8 in stages, and begin with existence under an L^2 assumption on the initial state. Part (b) of the next theorem will be used later to establish regularity of solutions as functions of the initial state.

Theorem 7.10. *Let Assumption 7.7 be in force and the $\mathcal{F}_0$-measurable initial point satisfy*

$$E(|\xi|^2) < \infty.$$

(a) *There exists a continuous process X adapted to $\{\mathcal{F}_t\}$ that satisfies integral equation (7.18) and this moment bound: for each $T < \infty$ there exists a constant $C = C(T, L) < \infty$ such that*

$$E\big(\sup_{t\in[0,T]} |X_t|^2\big) \le C\big(1 + E(|\xi|^2)\big) \qquad \text{for } t \in [0,T]. \tag{7.22}$$

In particular, the integrand $\sigma(s, X_s)$ is a member of $\mathcal{L}_2(B)$ and

$$E\int_0^T |b(s,X_s)|^2\,ds < \infty \qquad \text{for each } T < \infty.$$

(b) *Let $\tilde{\xi} \in L^2(P)$ be also $\mathcal{F}_0$-measurable, and let $\widetilde{X}$ be the solution constructed in part* (a) *with initial condition $\tilde{\xi}$. Then there exists a constant $C < \infty$ such that for all $t < \infty$,*

$$E\big[\sup_{0\le s\le t} |X_s - \widetilde{X}_s|^2\,\big] \le 9e^{Ct^2} E[\,|\xi - \tilde{\xi}|^2\,]. \tag{7.23}$$

Proof of Theorem 7.10. (a) The existence proof uses classic Picard iteration. We define a sequence of processes X_n indexed by $n \in \mathbf{Z}_+$ by $X_0(t) = \xi$ and then for $n \ge 0$

$$X_{n+1}(t) = \xi + \int_0^t b(s, X_n(s))\,ds + \int_0^t \sigma(s, X_n(s))\,dB_s. \tag{7.24}$$

Step 1. We show that for each n the process X_n is well-defined, continuous, and satisfies

$$E\big(\sup_{s\in[0,t]}|X_n(s)|^2\big) < \infty \quad \text{for all } t \in \mathbf{R}_+. \tag{7.25}$$

Bound (7.25) is clear in the case $n = 0$ by the definition $X_0(t) = \xi$ and assumption $\xi \in L^2(P)$.

Assume (7.25) for n. That process X_{n+1} is well-defined by the integrals in (7.24) follows from

$$\begin{aligned} &E\bigg[\int_0^t |b(s, X_n(s))|^2\,ds + \int_0^t |\sigma(s, X_n(s))|^2\,ds\bigg] \\ &\qquad \le 4L^2 t\big(1 + \sup_{s\in[0,t]} E|X_n(s)|^2\big) < \infty \end{aligned} \tag{7.26}$$

where we used assumption (7.21) and then (7.25) for n.

Property (7.26) implies that the first (Lebesgue) integral on the right of (7.24) is a continuous L^2 process and that the second (stochastic) integral is a continuous L^2 martingale because the integrand lies in $\mathcal{L}_2(B)$.

Now we know that $X_{n+1}(t)$ is a well-defined continuous square-integrable process. Take supremum over time in equation (7.24). To bound the second moment, use first $(a+b+c)^2 \le 9(a^2+b^2+c^2)$, then Schwarz inequality to the first integral, and apply Doob's inequality (Theorem (3.12)) followed by the isometry of stochastic integration to the second integral.

$$\begin{aligned} E\big(\sup_{s\in[0,t]}|X_{n+1}(s)|^2\big) &\le 9E(|\xi|^2) + 9E\bigg[\sup_{s\in[0,t]}\bigg|\int_0^s b(u, X_n(u))\,du\bigg|^2\bigg] \\ &\qquad + 9E\bigg[\sup_{s\in[0,t]}\bigg|\int_0^s \sigma(u, X_n(u))\,dB_u\bigg|^2\bigg] \\ &\le 9E(|\xi|^2) + 9tE\int_0^t |b(s, X_n(s))|^2\,ds + 36E\int_0^t |\sigma(s, X_n(s))|^2\,ds \\ &\le 9E(|\xi|^2) + 90L^2(1+t)t + 90L^2(1+t)\int_0^t E|X_n(s)|^2\,ds. \end{aligned} \tag{7.27}$$

In the last step we applied assumption (7.21). Now it is clear that (7.25) is passed from n to $n+1$.

Step 2. For each $T < \infty$ there exists a constant $A = A(T, L) < \infty$ such that

$$E\big(\sup_{s\in[0,t]}|X_n(s)|^2\big) \le A\big(1 + E(|\xi|^2)\big) \qquad \text{for all } n \text{ and } t \in [0, T]. \tag{7.28}$$

To prove Step 2, restrict t to $[0,T]$ and introduce a constant $C = C(T,L) \in [1,\infty)$ to rewrite the outcome of (7.27) as

$$E\big(\sup_{s\in[0,t]} |X_{n+1}(s)|^2\big) \le C\big(1+E(|\xi|^2)\big) + C\int_0^t E\big(\sup_{u\in[0,s]} |X_n(u)|^2\big)\,ds.$$

Thus with $y_n(t) = E\big(\sup_{s\in[0,t]}|X_n(s)|^2\big)$ and $B = C\big(1+E(|\xi|^2)\big)$ we have this situation:

$$y_0(t) \le B \quad \text{and} \quad y_{n+1}(t) \le B + C\int_0^t y_n(s)\,ds. \tag{7.29}$$

This assumption gives inductively

$$y_n(t) \le B\sum_{k=0}^{n} \frac{C^k t^k}{k!} \le Be^{Ct}. \tag{7.30}$$

This translates into (7.28).

Step 3. There exists a continuous, adapted process X such that $X_n \to X$ uniformly on compact time intervals both almost surely and in this L^2 sense: for each $T < \infty$

$$\lim_{n\to\infty} E\big[\sup_{0\le s\le T} |X(s) - X_n(s)|^2\big] = 0. \tag{7.31}$$

Furthermore, X satisfies moment bound (7.22).

We isolate a part of this step as a lemma for later use.

Lemma 7.11. *Let Assumption 7.7 be in force. Let X and Y be continuous, adapted L^2 processes that satisfy estimate (7.25), ξ and η $L^2(P)$ random variables, and processes $\bar{X}$ and $\bar{Y}$ defined by*

$$\bar{X}_t = \xi + \int_0^t b(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s$$

and

$$\bar{Y}_t = \eta + \int_0^t b(s,Y_s)\,ds + \int_0^t \sigma(s,Y_s)\,dB_s.$$

Then there exists a constant $C = C(L) < \infty$ such that for all $0 \le t \le T < \infty$

$$\begin{aligned} &E\big[\sup_{0\le s\le t} |\bar{X}_s - \bar{Y}_s|^2\big] \\ &\qquad \le 9E|\xi-\eta|^2 + CT\int_0^t E\big[\sup_{0\le u\le s} |X_u - Y_u|^2\big]\,ds. \end{aligned} \tag{7.32}$$

Proof. Assuming (7.25) for X and Y ensures that all integrals are well-defined.

$$\begin{aligned}\bar X_t - \bar Y_t &= \xi - \eta + \int_0^t \big[b(s,X_s) - b(s,Y_s)\big]\,ds \\ &\qquad + \int_0^t \big[\sigma(s,X_s) - \sigma(s,Y_s)\big]\,dB_s \\ &\equiv \xi - \eta + G(t) + M(t)\end{aligned} \tag{7.33}$$

where G is an FV process and M an L^2 martingale. Schwarz inequality and the Lipschitz assumption (7.20) give

$$\begin{aligned}E\big[\sup_{0\le s\le t}|G(s)|^2\big] &\le t\int_0^t E\big|b(s,X_s) - b(s,Y_s)\big|^2\,ds \\ \le L^2 t\int_0^t E|X_s - Y_s|^2\,ds &\le L^2 t\int_0^t E\big[\sup_{0\le u\le s}|X_u - Y_u|^2\big]\,ds\end{aligned} \tag{7.34}$$

The isometry of stochastic integration and Doob's inequality (Theorem 3.12) give

$$\begin{aligned}&E\big[\sup_{0\le s\le t}|M(s)|^2\big] \le 4E\big[\,|M(t)|^2\big] \\ &= 4\int_0^t E\big|\sigma(s,X_s) - \sigma(s,Y_s)\big|^2\,ds \\ &\le 4L^2\int_0^t E|X_s - Y_s|^2\,ds \le 4L^2\int_0^t E\big[\sup_{0\le u\le s}|X_u - Y_u|^2\big]\,ds.\end{aligned} \tag{7.35}$$

Collect the estimates. □

Return to the proof of Step 3. Restrict t to $[0,T]$. Lemma 7.11 applied to $\bar X = X_{n+1}$, $\bar Y = X = X_n$ and $Y = X_{n-1}$ gives

$$\begin{aligned}&E\big[\sup_{0\le s\le t}|X_{n+1}(s) - X_n(s)|^2\big] \\ &\qquad \le C\int_0^t E\big[\sup_{0\le u\le s}|X_n(u)) - X_{n-1}(u)|^2\big]\,ds.\end{aligned} \tag{7.36}$$

For $y_n(t) = E\big[\sup_{0\le s\le t}|X_{n+1}(s) - X_n(s)|^2\big]$ we have constants B, C such that these inequalities hold for $t \in [0,T]$:

$$y_0(t) \le B \quad \text{and} \quad y_{n+1}(t) \le C\int_0^t y_n(s)\,ds. \tag{7.37}$$

The inequality for $y_0(t)$ is checked separately:

$$\begin{aligned} y_0(t) &= E\big[\sup_{0\le s\le t}|X_1(s)-X_0(s)|^2\big] \\ &= E\Big[\sup_{0\le s\le t}\Big|\int_0^s b(u,\xi)\,du + \int_0^s \sigma(u,\xi)\,dB_u\Big|^2\Big] \end{aligned}$$

and reason as above once again.

Bounds (7.37) develop inductively into

$$y_n(t) \le B\frac{C^n t^n}{n!}. \tag{7.38}$$

An application of Chebyshev's inequality gives us a summable bound that we can feed into the Borel-Cantelli lemma:

$$\begin{aligned} P\big\{\sup_{0\le s\le t}|X_{n+1}(s)-X_n(s)| \ge 2^{-n}\big\} &\le 4^n E\big[\sup_{0\le s\le t}|X_{n+1}(s)-X_n(s)|^2\big] \\ &\le B\frac{4^n C^n t^n}{n!}. \end{aligned}$$

Consider now $T \in \mathbf{N}$. From

$$\sum_n P\big\{\sup_{0\le s\le T}|X_{n+1}(s)-X_n(s)| \ge 2^{-n}\big\} < \infty$$

and the Borel-Cantelli lemma we conclude that for almost every ω and each $T \in \mathbf{N}$ there exists a finite $N_T(\omega) < \infty$ such that $n \ge N_T(\omega)$ implies

$$\sup_{0\le s\le T}|X_{n+1}(s)-X_n(s)| < 2^{-n}.$$

Thus for almost every ω the paths $\{X_n\}$ form a Cauchy sequence in the space $C[0,T]$ of continuous functions with supremum distance, for each $T \in \mathbf{N}$. This space is complete (Lemma A.5). Consequently there exists a continuous path $\{X_s(\omega) : s \in \mathbf{R}_+\}$, defined for almost every ω, such that

$$\sup_{0\le s\le T}|X(s)-X_n(s)| \to 0 \quad \text{as } n\to\infty \tag{7.39}$$

for all $T \in \mathbf{N}$ and almost every ω. This defines X and proves the almost sure convergence part of Step 3. Adaptedness of X comes from $X_n(t) \to X(t)$ a.s. and the completeness of the σ-algebra $\mathcal{F}_t$.

From the uniform convergence we get

$$\sup_{0\le s\le T}|X(s)-X_n(s)| = \lim_{m\to\infty}\sup_{0\le s\le T}|X_m(s)-X_n(s)|.$$

Abbreviate $\eta_n = \sup_{0\le s\le T}|X_{n+1}(s)-X_n(s)|$ and let $\|\eta_n\|_2 = E(\eta_n^2)^{1/2}$ denote L^2 norm. By Fatou's lemma and the triangle inequality for the L^2

norm,

$$\Big\| \sup_{0\le s\le T} |X(s)-X_n(s)| \Big\|_2 \le \varliminf_{m\to\infty} \Big\| \sup_{0\le s\le T} |X_m(s)-X_n(s)| \Big\|_2$$
$$\le \varliminf_{m\to\infty} \sum_{k=n}^{m-1} \|\eta_k\|_2 = \sum_{k=n}^{\infty} \|\eta_k\|_2.$$

By estimate (7.38) the last expression vanishes as $n\to\infty$. This gives (7.31). Finally, L^2 convergence (or Fatou's lemma) converts (7.28) into (7.22). This completes the proof of Step 3.

Step 4. Process X satisfies (7.18).

The integrals in (7.18) are well-defined continuous processes because we have established the L^2 bound on X that gives the L^2 bounds on the integrands as in the last step of (7.27). To prove Step 4 we show that the terms in (7.24) converge to those of (7.18) in L^2. We have derived the required estimates already. For example, by the Lipschitz assumption and (7.31), for every $T<\infty$,

$$E\int_0^T \big|\sigma(s,X_n(s))-\sigma(s,X(s))\big|^2\,ds \le L^2 E\int_0^T |X_n(s)-X(s)|^2\,ds \to 0.$$

This says that the integrand $\sigma(s,X_n(s))$ converges to $\sigma(s,X(s))$ in the space $\mathcal{L}_2(B)$. Consequently we have the L^2 convergence of stochastic integrals:

$$\int_0^t \sigma(s,X_n(s))\,dB_s \to \int_0^t \sigma(s,X(s))\,dB_s$$

which also holds uniformly on compact time intervals.

We leave the argument for the other integral as an exercise. Proof of part (a) of Theorem 7.10 is complete.

(b) Apply Lemma 7.11 to $\eta=\widetilde{\xi}$, $\bar X = X$ and $\bar Y = Y = \widetilde{X}$. The conclusion (7.32) justifies an application of Gronwall's inequality (Lemma A.20) which gives the conclusion. ☐

At this point we prove strong uniqueness. We prove a slightly more general result which we can use also to extend the existence result beyond $\xi\in L^2(P)$. Let η be another initial condition and Y a process that satisfies

$$Y_t = \eta + \int_0^t b(s,Y_s)\,ds + \int_0^t \sigma(s,Y_s)\,dB_s. \tag{7.40}$$

Theorem 7.12. *Let ξ and η be $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variables without any integrability assumptions. Assume that coefficients b and σ are Borel functions that satisfy Lipschitz condition (7.20). Let X and Y be two continuous processes on $(\Omega,\mathcal{F},P)$ adapted to $\{\mathcal{F}_t\}$. Assume X satisfies* *integral equation (7.18) and condition (7.19). Assume Y satisfies integral equation (7.40) and condition (7.19) with X replaced by Y. Then on the event $\{\xi = \eta\}$ processes X and Y are indistinguishable. That is, for almost every ω such that $\xi(\omega) = \eta(\omega)$, $X_t(\omega) = Y_t(\omega)$ for all $t \in \mathbf{R}_+$.*

In the case $\eta = \xi$ we get the strong uniqueness. The uniqueness statement is actually true with a weaker local Lipschitz condition (Exercise 7.4).

Corollary 7.13. *Let ξ be an $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable. Assume that b and σ are Borel functions that satisfy Lipschitz condition (7.20). Then up to indistinguishability there is at most one continuous process X on $(\Omega, \mathcal{F}, P)$ adapted to $\{\mathcal{F}_t\}$ that satisfies (7.18)–(7.19).*

Proof of Theorem 7.12. In order to use L^2 bounds we stop the processes before they get too large. Fix $n \in \mathbf{N}$ for the time being and define

$$\nu = \inf\{t \geq 0 : |X_t - \xi| \geq n \;\text{ or }\; |Y_t - \eta| \geq n\}.$$

This is a stopping time by Corollary 2.10. Processes $X^\nu - \xi$ and $Y^\nu - \eta$ are bounded.

The bounded $\mathcal{F}_0$-measurable random variable $\mathbf{1}\{\eta = \xi\}$ moves freely in and out of stochastic integrals. This can be seen for example from the constructions of stochastic integrals. By subtracting the equations for X_t and Y_t and then stopping we get this equation:

$$\begin{aligned}
&(X_t^\nu - Y_t^\nu) \cdot \mathbf{1}\{\eta = \xi\} \\
&= \int_0^{\nu\wedge t} \bigl[b(s, X_s) - b(s, Y_s)\bigr] \mathbf{1}\{\eta = \xi\}\, ds \\
&\qquad + \int_0^{\nu\wedge t} \bigl[\sigma(s, X_s) - \sigma(s, Y_s)\bigr] \mathbf{1}\{\eta = \xi\}\, dB_s \\
&= \int_0^{\nu\wedge t} \bigl[b(s, X_s^\nu) - b(s, Y_s^\nu)\bigr] \mathbf{1}\{\eta = \xi\}\, ds \\
&\qquad + \int_0^{\nu\wedge t} \bigl[\sigma(s, X_s^\nu) - \sigma(s, Y_s^\nu)\bigr] \mathbf{1}\{\eta = \xi\}\, dB_s \\
&\equiv G(t) + M(t).
\end{aligned} \tag{7.41}$$

The last equality defines the processes G and M. The next to last equality comes from the fact that if integrands agree up to a stopping time, then so do stochastic integrals (Proposition 4.10(c) for Brownian motion). We have the constant bound

$$\begin{aligned}
\bigl|\sigma(s, X_s^\nu) - \sigma(s, Y_s^\nu)\bigr| \cdot \mathbf{1}\{\eta = \xi\} &\leq L|X_s^\nu - Y_s^\nu| \cdot \mathbf{1}\{\eta = \xi\} \\
&\leq L|(X_s^\nu - \xi) - (Y_s^\nu - \eta)| \cdot \mathbf{1}\{\eta = \xi\} \leq 2Ln
\end{aligned}$$

and similarly for the other integrand. Consequently G is an L^2 FV process and M an L^2 martingale.

Schwarz inequality, bounding $\nu \wedge t$ above with t, and the Lipschitz assumption (7.20) give

$$E\big[\,|G(t)|^2\,\big] \le t \int_0^t E\big[\,|b(s,X_s^\nu) - b(s,Y_s^\nu)|^2\,\mathbf{1}\{\eta=\xi\}\big]\,ds$$
$$\le L^2 t \int_0^t E\big[\,|X_s^\nu - Y_s^\nu|^2\,\mathbf{1}\{\eta=\xi\}\big]\,ds.$$

The isometry of stochastic integration gives

$$\begin{aligned}
E\big[\,|M(t)|^2\,\big] &= E\bigg[\bigg(\int_0^{\nu\wedge t} \big[\sigma(s,X_s^\nu) - \sigma(s,Y_s^\nu)\big]\,\mathbf{1}\{\eta=\xi\}\,dB_s\bigg)^2\,\bigg] \\
&= E\bigg[\bigg(\int_0^t \mathbf{1}_{[0,\nu]}(t)\big[\sigma(s,X_s^\nu) - \sigma(s,Y_s^\nu)\big]\,\mathbf{1}\{\eta=\xi\}\,dB_s\bigg)^2\,\bigg] \\
&= \int_0^t E\big[\,\mathbf{1}_{[0,\nu]}(t)|\sigma(s,X_s^\nu) - \sigma(s,Y_s^\nu)|^2\,\mathbf{1}\{\eta=\xi\}\,\big]\,ds \\
&\le L^2 \int_0^t E\big[\,|X_s^\nu - Y_s^\nu|^2\,\mathbf{1}\{\eta=\xi\}\big]\,ds.
\end{aligned}$$

In the second equality above we moved the cut-off at ν from the upper limit into the integrand in accordance with (5.23) from Proposition 5.16. (Strictly speaking we have proved this only for predictable integrands but naturally the same property can be proved for the more general Brownian motion integral.)

Combine the above bounds into

$$E\big[\,|X_t^\nu - Y_t^\nu|^2\,\mathbf{1}\{\eta=\xi\}\big] \le 2L^2(t+1)\int_0^t E\big[\,|X_s^\nu - Y_s^\nu|^2\,\mathbf{1}\{\eta=\xi\}\big]\,ds.$$

If we restrict t to $[0,T]$ we can replace $2L^2(t+1)$ with the constant $B = 2L^2(T+1)$ above. Then Gronwall's inequality (Lemma A.20) implies

$$E\big[\,|X_t^\nu - Y_t^\nu|^2\,\mathbf{1}\{\eta=\xi\}\big] = 0 \qquad \text{for } t \in [0,T].$$

We can repeat this argument for each $T \in \mathbf{N}$ and so extend the conclusion to all $t \in \mathbf{R}$. Thus on the event $\eta = \xi$ the continuous processes X^ν and Y^ν agree almost surely at each fixed time. It follows that they are indistinguishable. Now we have this statement for each n that appears in the definition of ν. With continuous paths we have $\nu \nearrow \infty$ as we take n to infinity. In the end we can conclude that X and Y are indistinguishable on the event $\eta = \xi$. $\square$

Completion of the proof of Theorem 7.8. It remains to remove the assumption $\xi \in L^2(P)$ from the existence proof. Let $X_m(t)$ be the solution given by Theorem 7.10 for the initial point $\xi\mathbf{1}\{|\xi| \le m\}$. By Theorem

7.12, for $m < n$ processes X_m and X_n are indistinguishable on the event $\{|\xi| \le m\}$. Thus we can consistently define

$$X(t) = X_m(t) \text{ on the event } \{|\xi| \le m\}.$$

X has continuous paths because each X_m does, and adaptedness of X can be checked because $\{|\xi| \le m\} \in \mathcal{F}_0$. X satisfies (7.19) because each X_m does. Thus the integrals in (7.18) are well-defined for X.

To verify equation (7.18) we can again pass the $\mathcal{F}_0$-measurable indicator $\mathbf{1}\{|\xi| \le m\}$ in and out of stochastic integrals, and then use the equality of processes $b(s, X(s))\mathbf{1}\{|\xi| \le m\} = b(s, X_m(s))\mathbf{1}\{|\xi| \le m\}$ with the same property for σ:

$$\begin{aligned}
&X(t)\mathbf{1}\{|\xi| \le m\} = X_m(t)\mathbf{1}\{|\xi| \le m\} \\
&= \xi\mathbf{1}\{|\xi| \le m\} + \int_0^t b(s, X_m(s))\mathbf{1}\{|\xi| \le m\}\, ds \\
&\qquad + \int_0^t \sigma(s, X_m(s))\mathbf{1}\{|\xi| \le m\}\, dB_s \\
&= \xi\mathbf{1}\{|\xi| \le m\} + \int_0^t b(s, X(s))\mathbf{1}\{|\xi| \le m\}\, ds \\
&\qquad + \int_0^t \sigma(s, X(s))\mathbf{1}\{|\xi| \le m\}\, dB_s \\
&= \mathbf{1}\{|\xi| \le m\}\biggl(\xi + \int_0^t b(s, X(s))\, ds + \int_0^t \sigma(s, X(s))\, dB_s\biggr).
\end{aligned}$$

Since the union of the events $\{|\xi| \le m\}$ is the entire space (almost surely), we have the equation for X. □

Let us also address *weak uniqueness* or *uniqueness in distribution*. This concerns solutions on possibly different probability spaces. As above, assume given a Brownian motion B relative to a complete filtration $\{\mathcal{F}_t\}$ on a probability space $(\Omega, \mathcal{F}, P)$, and an $\mathcal{F}_0$-measurable random variable ξ. Let X be a continuous $\{\mathcal{F}_t\}$-adapted process that satisfies integral equation (7.18) and condition (7.19). Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$, $\{\widetilde{\mathcal{F}}_t\}$, $\widetilde{B}$, $\tilde{\xi}$ be another system with exactly the same properties, and let $\widetilde{X}$ be a continuous process on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ adapted to $\{\widetilde{\mathcal{F}}_t\}$ that satisfies the analogue of (7.19) and the equation

$$\widetilde{X}_t = \tilde{\xi} + \int_0^t b(s, \widetilde{X}_s)\, ds + \int_0^t \sigma(s, \widetilde{X}_s)\, d\widetilde{B}_s. \tag{7.42}$$

The point is that the coefficient functions b and σ are shared. The theorem says that the probability distribution of the solution process X is uniquely determined by the distribution of ξ and the functions b and σ.

Theorem 7.14. *Let Assumption 7.7 be in force and assume b and σ are continuous functions of (t,x). Assume $\xi \overset{d}{=} \tilde\xi$. Then processes X and $\widetilde X$ have the same probability distribution. That is, for any measurable subset A of $C_{\mathbf{R}^d}[0,\infty)$, $P(X \in A) = \widetilde P(\widetilde X \in A)$.*

Proof. Assume first that ξ and $\tilde\xi$ are L^2 variables. Follow the recipe (7.24) of the proof of Theorem 7.10 to construct the two sequences of processes, X_n on Ω and $\widetilde X_n$ on $\widetilde\Omega$, that converge in the sense of (7.31) and (7.39) to solution processes. By the strong uniqueness, $X_n \to X$ and $\widetilde X_n \to \widetilde X$ in the sense mentioned. Thus it suffices to show the distributional equality $X_n \overset{d}{=} \widetilde X_n$ of the approximating processes. We do this inductively.

$(X_0, B) \overset{d}{=} (\widetilde X_0, \widetilde B)$ follows from $\xi \overset{d}{=} \tilde\xi$ and the fact that both pairs (ξ, B) and $(\tilde\xi, \widetilde B)$ are independent.

Suppose $(X_n, B) \overset{d}{=} (\widetilde X_n, \widetilde B)$. Since the integrands are continuous and satisfy uniform estimates of the type (7.28), we can fix the partitions $s_i^k = i2^{-k}$ of the time axis and realize the integrals on the right-hand side of (7.24) as limits of integrals of simple functions:

$$X_{n+1}(t) = \xi + \lim_{k\to\infty} \biggl\{ \sum_{i\ge 0} b(s_i^k, X_n(s_i^k))(t \wedge s_{i+1}^k - t \wedge s_i^k) + \sum_{i\ge 0} \sigma(s_i^k, X_n(s_i^k))(B_{t\wedge s_{i+1}^k} - B_{t\wedge s_i^k}) \biggr\}$$

and

$$\widetilde X_{n+1}(t) = \tilde\xi + \lim_{k\to\infty} \biggl\{ \sum_{i\ge 0} b(s_i^k, \widetilde X_n(s_i^k))(t \wedge s_{i+1}^k - t \wedge s_i^k) + \sum_{i\ge 0} \sigma(s_i^k, \widetilde X_n(s_i^k))(\widetilde B_{t\wedge s_{i+1}^k} - \widetilde B_{t\wedge s_i^k}) \biggr\}.$$

These limits are uniform over bounded time intervals and L^2 or in probability over the probability space. By passing to a subsequence, the limit is also almost sure.

Fix time points $0 \le t_1 < \cdots < t_m$. The induction assumption $(X_n, B) \overset{d}{=} (\widetilde X_n, \widetilde B)$ and the limits above imply that the vectors

$$(X_{n+1}(t_1), \dots, X_{n+1}(t_m), B_{t_1}, \dots, B_{t_m})$$
$$\text{and } (\widetilde X_{n+1}(t_1), \dots, \widetilde X_{n+1}(t_m), \widetilde B_{t_1}, \dots, \widetilde B_{t_m})$$

have identical distributions. Finite-dimensional distributions determine the distribution of an entire process, and we may conclude that $(X_{n+1}, B) \overset{d}{=} (\widetilde X_{n+1}, \widetilde B)$.

This concludes the proof for the case of square integrable ξ and $\tilde{\xi}$. For the general case do an approximation through $\xi\mathbf{1}\{|\xi| \le m\}$ and $\tilde{\xi}\mathbf{1}\{|\tilde{\xi}| \le m\}$, as in the previous proof. □

For $x \in \mathbf{R}^d$, let P^x be the distribution on $C = C_{\mathbf{R}^d}[0,\infty)$ of the process X that solves the SDE (7.18) with deterministic initial point $\xi = x$. E^x denotes expectation under P^x. The family of probability measures $\{P^x\}_{x\in\mathbf{R}^d}$ is well defined on account of Theorem 7.14. Let $x_j \to x$ in $\mathbf{R}^d$. Then estimate (7.23) implies that $P^{x_j} \to P^x$ in the weak topology of the space $\mathcal{M}_1(C)$ of Borel probability measures on C. Continuous mappings are in general Borel measurable. Consequently $x \mapsto P^x$ is a measurable mapping from $\mathbf{R}^d$ into $\mathcal{M}_1(C)$. This implies that $x \mapsto E^x(G)$ is a Borel function for any bounded Borel function $G : C \to \mathbf{R}$.

Thus we can inquire about the Markov property of the solution of an Itô equation. Since we only discussed time-homogeneous Markov processes, let us assume that the coefficient functions b and σ do not depend explicitly on the time-coordinate. This is not really a restriction because the time coordinate can always be included in the state of the system; that is, X_t can be replaced by $\bar{X}_t = (t, X_t)$. So now the equation whose solution uniquely determines distribution P^x on C is

$$X_t = x + \int_0^t b(X_s)\,ds + \int_0^t \sigma(X_s)\,dB_s. \tag{7.43}$$

Theorem 7.15. *Let Assumption 7.7 be in force. Then the family $\{P^x\}_{x\in\mathbf{R}^d}$ of probability measures on C defined by the solutions of the Itô equations (7.43) forms a Markov process that also possesses the strong Markov property.*

Proof. The continuity of $x \mapsto P^x$ gives the Feller property, so by Theorem 2.23 all we need to verify is the Markov property itself: namely, for any bounded measurable function $G : C \to \mathbf{R}$ and $r \in [0,\infty)$,

$$E^x[G(X_{r+\cdot}) \,|\, \mathcal{F}_r] = E^{X_r}[G] \quad P^x\text{-a.s.}, \tag{7.44}$$

where (X_t) now denotes the coordinate process on C, $\mathcal{F}_t = \sigma\{X_s : 0 \le s \le t\}$ is the filtration on C generated by the coordinates, and $X_{r+\cdot} = (X_{r+t})_{t\in[0,\infty)}$ is the coordinate process from time r onwards. To check this we take the calculation back to a space where we solve the SDE.

Fix $x \in \mathbf{R}^d$ and $r \in [0,\infty)$. Let $(\Omega, \mathcal{H}, \mathbf{P})$ be a probability space with complete filtration $\{\mathcal{H}_t\}$, and B an $\mathbf{R}^m$-valued Brownian motion with respect to the filtration $\{\mathcal{H}_t\}$. Let Y denote the a.s. unique solution of the Itô equation (7.43) with deterministic initial point x and adapted to $\{\mathcal{H}_t\}$.

We condition on $\mathcal{H}_r$ on the space $(\Omega, \mathcal{H}, \mathbf{P})$. We can take Ω Polish with Borel σ-algebra $\mathcal{H}$, for example simply by $(\Omega, \mathcal{H}) = (C, \mathcal{B}_C)$. Conditional probability measures exist on Polish spaces. Consequently we have a measurable function $\omega \mapsto \mathbf{P}^\omega$ from Ω into $\mathcal{M}_1(\Omega)$ such that $\mathbf{P}^\omega$ is the conditional probability measure of $\mathbf{P}$, given $\mathcal{H}_r$. This means that the expectation $\mathbf{E}^\omega(f)$ of f under $\mathbf{P}^\omega$ is a version of the conditional expectation $\mathbf{E}[f \,|\, \mathcal{H}_r](\omega)$ for any bounded measurable $f : \Omega \to \mathbf{R}$.

Under $\mathbf{P}$ the process $\check{B}_t = B_{r+t} - B_r$ is a standard Brownian motion, independent of $\mathcal{H}_r$. Hence under $\mathbf{P}^\omega$ process $\check{B}$ is still a standard Brownian motion. Next, there is an event Ω_0 such that $\mathbf{P}(\Omega_0) = 1$ and for each $\omega \in \Omega_0$, the variable Y_r is $\mathbf{P}^\omega$-a.s. equal to the constant $Y_r(\omega)$. To see this, begin with the $\mathbf{P}$-a.s. identity

$$\mathbf{P}^\omega(Y_r \in U) = \mathbf{E}[\mathbf{1}_U(Y_r) | \, \mathcal{H}_r](\omega) = \mathbf{1}_U(Y_r(\omega))$$

that follows from the $\mathcal{H}_r$-measurability of Y_r. Now let Ω_0 be the set of ω for which the above identity holds for all sets U in a countable base for the topology of $\mathbf{R}^d$: for example, U ranges over balls with rational centers and radii.

Define the process $\check{Y}_t = Y_{r+t}$, adapted to the filtration $\check{\mathcal{H}}_t = \mathcal{H}_{r+t}$. Increments of the new Brownian motion are shifts of the increments of the original Brownian motion: $\check{B}_t - \check{B}_s = B_{r+t} - B_{r+s}$. Consequently we have equality of stochastic integrals $\int_0^t \sigma(\check{Y}_s)\, d\check{B}_s = \int_r^{r+t} \sigma(Y_s)\, dB_s$, and thereby the equation

$$\check{Y}_t = Y_r + \int_0^t b(\check{Y}_s)\, ds + \int_0^t \sigma(\check{Y}_s)\, d\check{B}_s.$$

The points above amount to the following. For $\omega \in \Omega_0$, on the probability space $(\Omega, \mathcal{H}, \mathbf{P}^\omega)$ process $\check{Y}$ is a solution of the SDE (7.43) with initial point $x = Y_r(\omega)$. Consequently the distribution of $\check{Y}$ under the measure $\mathbf{P}^\omega$ is $P^{Y_r(\omega)}$.

We can now verify (7.44). Let $A \in \mathcal{F}_r$ on the space C. Note that then $\mathbf{1}_A(Y)$ depends only on the segment $Y_{[0,r]} = (Y_s : s \in [0, r])$ and hence is $\mathcal{H}_r$-measurable. Under $\mathbf{P}$ process Y has distribution P^x, so we can compute as follows.

$$\begin{aligned} E^x[\mathbf{1}_A\, G(X_{r+\cdot})] &= \mathbf{E}[\mathbf{1}_A(Y)\, G(\check{Y})] = \mathbf{E}\big[\mathbf{1}_A(Y)\, \mathbf{E}^\omega(G(\check{Y}))\big] \\ &= \mathbf{E}\big[\mathbf{1}_A(Y)\, E^{Y_r}(G)\big] = E^x\big[\mathbf{1}_A\, E^{X_r}(G)\big]. \end{aligned}$$

The second equality followed from conditioning on $\mathcal{H}_r$ inside $\mathbf{E}[\cdots]$. This completes the proof of the strong Markov property. □

7.3. A semimartingale equation

Fix a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$. We consider the equation

$$X(t,\omega) = H(t,\omega) + \int_{(0,t]} F(s,\omega,X(\omega))\,dY(s,\omega) \tag{7.45}$$

where Y is a given $\mathbf{R}^m$-valued cadlag semimartingale, H is a given $\mathbf{R}^d$-valued adapted cadlag process, and X is the unknown $\mathbf{R}^d$-valued process. The coefficient F is a $d \times m$-matrix valued function of its arguments. (The componentwise interpretation of the stochastic integral $\int F\,dY$ appeared Remark 5.38.) For the coefficient F we make these assumptions.

Assumption 7.16. The coefficient function $F(s,\omega,\eta)$ in equation (7.45) is a measurable function from the space $\mathbf{R}_+ \times \Omega \times D_{\mathbf{R}^d}[0,\infty)$ into the space $\mathbf{R}^{d\times m}$ of $d \times m$ matrices, and satisfies the following requirements.

(i) F satisfies a spatial Lipschitz condition uniformly in the other variables: there exists a finite constant L such that this holds for all $(t,\omega) \in \mathbf{R}_+ \times \Omega$ and all $\eta, \zeta \in D_{\mathbf{R}^d}[0,\infty)$:

$$|F(t,\omega,\eta) - F(t,\omega,\zeta)| \le L \cdot \sup_{s\in[0,t)} |\eta(s) - \zeta(s)|. \tag{7.46}$$

(ii) Given any adapted $\mathbf{R}^d$-valued cadlag process X on Ω, the function $(t,\omega) \mapsto F(t,\omega,X(\omega))$ is a predictable process.

(iii) Given any adapted $\mathbf{R}^d$-valued cadlag process X on Ω, there exist stopping times $\nu_k \nearrow \infty$ such that $\mathbf{1}_{(0,\nu_k]}(t)F(t,X)$ is bounded for each k.

These conditions on F are rather technical. They are written to help prove the next theorem. The proof is lengthy and postponed to the next section. There are technical steps where we need right-continuity of the filtration, hence we include this requirement in the hypotheses.

Theorem 7.17. *Assume $\{\mathcal{F}_t\}$ is complete and right-continuous. Let H be an adapted $\mathbf{R}^d$-valued cadlag process and Y an $\mathbf{R}^m$-valued cadlag semimartingale. Assume F satisfies Assumption 7.16. Then there exists a cadlag process $\{X(t) : 0 \le t < \infty\}$ adapted to $\{\mathcal{F}_t\}$ that satisfies equation (7.45), and X is unique up to indistinguishability.*

In the remainder of this section we discuss the assumptions on F and state some consequences of the existence and uniqueness theorem.

Notice the exclusion of the endpoint t on the right-hand side of the Lipschitz condition (7.46). This implies that $F(t,\omega,\,\cdot\,)$ is a function of the stopped path

$$\eta^{t-}(s) = \begin{cases} \eta(0), & t = 0,\ 0 \le s < \infty \\ \eta(s), & 0 \le s < t \\ \eta(t-), & s \ge t > 0. \end{cases}$$

In other words, the function $F(t, \omega, \cdot\,)$ only depends on the path on the time interval $[0, t)$.

Parts (ii)–(iii) guarantees that the stochastic integral $\int F(s, X)\, dY(s)$ exists for an arbitrary adapted cadlag process X and semimartingale Y. The existence of the stopping times $\{\nu_k\}$ in part (iii) can be verified via this local boundedness condition.

Lemma 7.18. *Assume F satisfies parts* (i) *and* (ii) *of Assumpotion 7.16. Suppose there exists a path $\bar{\zeta} \in D_{\mathbf{R}^d}[0, \infty)$ such that for all $T < \infty$,*

$$c(T) = \sup_{t \in [0,T],\, \omega \in \Omega} |F(t, \omega, \bar{\zeta}\,)| < \infty. \tag{7.47}$$

Then for any adapted $\mathbf{R}^d$-valued cadlag process X there exist stopping times $\nu_k \nearrow \infty$ such that $\mathbf{1}_{(0,\nu_k]}(t) F(t, X)$ is bounded for each k.

Proof. Define

$$\nu_k = \inf\{t \ge 0 : |X(t)| \ge k\} \wedge \inf\{t > 0 : |X(t-)| \ge k\} \wedge k$$

These are bounded stopping times by Lemma 2.9. $|X(s)| \le k$ for $0 \le s < \nu_k$, but if $\nu_k = 0$ we cannot claim that $|X(0)| \le k$. The stopped process

$$X^{\nu_k-}(t) = \begin{cases} X(0), & \nu_k = 0 \\ X(t), & 0 \le t < \nu_k \\ X(\nu_k-), & t \ge \nu_k > 0 \end{cases}$$

is cadlag and adapted.

We show that $\mathbf{1}_{(0,\nu_k]}(s) F(s, X)$ is bounded. If $s = 0$ or $\nu_k = 0$ then this random variable vanishes. Suppose $0 < s \le \nu_k$. Since $[0, s) \subseteq [0, \nu_k)$, X agrees with X^{ν_k-} on $[0, s)$, and then $F(s, X) = F(s, X^{\nu_k-})$ by (7.46).

$$\begin{aligned} |F(s, X)| = |F(s, X^{\nu_k-})| &\le |F(s, \bar{\zeta}\,)| + |F(s, X^{\nu_k-}) - F(s, \bar{\zeta}\,)| \\ &\le c(k) + L \cdot \sup_{0 \le t < \nu_k} |X(t) - \bar{\zeta}(t)| \\ &\le c(k) + L\Big(k + \sup_{0 \le t \le k} |\bar{\zeta}(t)|\Big). \end{aligned}$$

The last line above is a finite quantity because $\bar{\zeta}$ is locally bounded, being a cadlag path. ∎

Here is how to apply the theorem to an equation that is not defined for all time.

Corollary 7.19. *Let $0 < T < \infty$. Assume $\{\mathcal{F}_t\}$ is right-continuous, Y is a cadlag semimartingale and H is an adapted cadlag process, all defined for $0 \le t \le T$. Let F satisfy Assumption 7.16 for $(t,\omega) \in [0,T] \times \Omega$. In particular, part* (ii) *takes this form: if X is a predictable process defined on $[0,T] \times \Omega$, then so is $F(t,X)$, and there is a nondecreasing sequence of stopping times $\{\sigma_k\}$ such that $\mathbf{1}_{(0,\sigma_k]}(t)F(t,X)$ is bounded for each k, and for almost every ω, $\sigma_k = T$ for all large enough k.*

Then there exists a unique solution X to equation (7.45) *on $[0,T]$.*

Proof. Extend the filtration, H, Y and F to all time in this manner: for $t \in (T,\infty)$ and $\omega \in \Omega$ define $\mathcal{F}_t = \mathcal{F}_T$, $H_t(\omega) = H_T(\omega)$, $Y_t(\omega) = Y_T(\omega)$, and $F(t,\omega,\eta) = 0$. Then the extended processes H and Y and the coefficient F satisfy all the original assumptions on all of $[0,\infty)$. Note in particular that if $G(t) = F(t,X)$ is a predictable process for $0 \le t \le T$, then extending it as a constant to (T,∞) produces a predictable process for $0 \le t < \infty$. And given a predictable process X on $[0,\infty)$, let $\{\sigma_k\}$ be the stopping times given by the assumption, and then define

$$\nu_k = \begin{cases} \sigma_k, & \sigma_k < T \\ \infty, & \sigma_k = T. \end{cases}$$

These stopping times satisfy part (iii) of Assumption 7.16 for $[0,\infty)$. Now Theorem 7.45 gives a solution X for all time $0 \le t < \infty$ for the extended equation, and on $[0,T]$ X solves the equation with the original H, Y and F.

For the uniqueness part, given a solution X of the equation on $[0,T]$, extend it to all time by defining $X_t = X_T$ for $t \in (T,\infty)$. Then we have a solution of the extended equation on $[0,\infty)$, and the uniqueness theorem applies to that. ∎

Another easy generalization is to the situation where the Lipschitz constant is an unbounded function of time.

Corollary 7.20. *Let the assumptions be as in Theorem 7.17, except that the Lipschitz assumption is weakened to this: for each $0 < T < \infty$ there exists a finite constant $L(T)$ such that this holds for all $(t,\omega) \in [0,T] \times \Omega$ and all $\eta, \zeta \in D_{\mathbf{R}^d}[0,\infty)$:*

$$|F(t,\omega,\eta) - F(t,\omega,\zeta)| \le L(T) \cdot \sup_{s\in[0,t)} |\eta(s) - \zeta(s)|. \tag{7.48}$$

Then equation (7.45) *has a unique solution X adapted to $\{\mathcal{F}_t\}$.*

Proof. For $k \in \mathbf{N}$, the function $\mathbf{1}_{\{0\le t\le k\}}F(t,\omega,\eta)$ satisfies the original hypotheses. By Theorem 7.17 there exists a process X_k that satisfies the equation

$$X_k(t) = H^k(t) + \int_{(0,t]} \mathbf{1}_{[0,k]}(s) F(s, X_k)\, dY^k(s). \tag{7.49}$$

The notation above is $H^k(t) = H(k \wedge t)$ for a stopped process as before. Let $k < m$. Stopping the equation

$$X_m(t) = H^m(t) + \int_{(0,t]} \mathbf{1}_{[0,m]}(s) F(s, X_m)\, dY^m(s)$$

at time k gives the equation

$$X_m^k(t) = H^k(t) + \int_{(0,t\wedge k]} \mathbf{1}_{[0,m]}(s) F(s, X_m)\, dY^m(s),$$

valid for all t. By Proposition 5.39 stopping a stochastic integral can be achieved by stopping the integrator or by cutting off the integrand with an indicator function. If we do both, we get the equation

$$X_m^k(t) = H^k(t) + \int_{(0,t]} \mathbf{1}_{[0,k]}(s) F(s, X_m^k)\, dY^k(s).$$

Thus X_k and X_m^k satisfy the same equation, so by the uniqueness theorem, $X_k = X_m^k$ for $k < m$. Thus we can unambiguously define a process X by setting $X = X_k$ on $[0, k]$. Then for $0 \le t \le k$ we can substitute X for X_k in equation (7.49) and get the equation

$$X(t) = H^k(t) + \int_{(0,k\wedge t]} F(s, X)\, dY(s), \quad 0 \le t \le k.$$

Since this holds for all k, X is a solution to the original SDE (7.45).

Uniqueness works similarly. If X and $\widetilde{X}$ solve (7.45), then $X(k \wedge t)$ and $\widetilde{X}(k \wedge t)$ solve (7.49). By the uniqueness theorem $X(k \wedge t) = \widetilde{X}(k \wedge t)$ for all t, and since k can be taken arbitrary, $X = \widetilde{X}$. □

Example 7.21. Here are ways by which a coefficient F satisfying the assumptions of Corollary 7.20 can arise.

(i) Let $f(t, \omega, x)$ be a $\mathcal{P} \otimes \mathcal{B}_{\mathbf{R}^d}$-measurable function from $(\mathbf{R}_+ \times \Omega) \times \mathbf{R}^d$ into $d \times m$-matrices. Assume f satisfies the Lipschitz condition

$$|f(t, \omega, x) - f(t, \omega, y)| \le L(T)|x - y|$$

for $(t, \omega) \in [0, T] \times \Omega$ and $x, y \in \mathbf{R}^d$, and the local boundedness condition

$$\sup\bigl\{|f(t, \omega, 0)| : 0 \le t \le T, \omega \in \Omega\bigr\} < \infty$$

for all $0 < T < \infty$. Then put $F(t, \omega, \eta) = f(t, \omega, \eta(t-))$.

Satisfaction of the conditions of Assumption 7.16 is straight-forward, except perhaps the predictability. Fix a cadlag process X and let $\mathcal{U}$ be the space of $\mathcal{P} \otimes \mathcal{B}_{\mathbf{R}^d}$-measurable f such that $(t, \omega) \mapsto f(t, \omega, X_{t-}(\omega))$ is

$\mathcal{P}$-measurable. This space is linear and closed under pointwise convergence. By Theorem B.4 it remains to check that $\mathcal{U}$ contains indicator functions $f = \mathbf{1}_{\Gamma \times B}$ of products of $\Gamma \in \mathcal{P}$ and $B \in \mathcal{B}_{\mathbf{R}^d}$. For such f, $f(t, \omega, X_{t-}(\omega)) = \mathbf{1}_\Gamma(t,\omega)\mathbf{1}_B(X_{t-}(\omega))$. The first factor $\mathbf{1}_\Gamma(t,\omega)$ is predictable by construction. The second factor $\mathbf{1}_B(X_{t-}(\omega))$ is predictable because X_{t-} is a predictable process, and because of the general fact that $g(Z_t)$ is predictable for any predictable process Z and any measurable function g on the state space of Z.

(ii) A special case of the preceding is a Borel measurable function $f(t,x)$ on $[0,\infty) \times \mathbf{R}^d$ with the required Lipschitz and boundedness conditions. In other words, no dependence on ω.

(iii) Continuing with non-random functions, suppose a function $G : D_{\mathbf{R}^d}[0,\infty) \to D_{\mathbf{R}^d}[0,\infty)$ satisfies a Lipschitz condition in this form:

$$G(\eta)_t - G(\zeta)_t \le L \cdot \sup_{s \in [0,t)} |\eta(s) - \zeta(s)|. \tag{7.50}$$

Then $F(t,\omega,\eta) = G(\eta)_{t-}$ satisfies Assumption 7.16.

Next, let us specialize to Itô equations to recover Theorem 7.8 proved in Section 7.2.

Corollary 7.22. *Let B_t be a standard Brownian motion in $\mathbf{R}^m$ with respect to a right-continuous filtration $\{\mathcal{F}_t\}$ and ξ an $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable. Fix $0 < T < \infty$. Assume the functions $b : [0,T] \times \mathbf{R}^d \to \mathbf{R}^d$ and $\sigma : [0,T] \times \mathbf{R}^d \to \mathbf{R}^{d\times m}$ satisfy the Lipschitz condition*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le L|x-y|$$

and the bound

$$|b(t,x)| + |\sigma(t,x)| \le L(1+|x|)$$

for a constant L and all $0 \le t \le T$, $x, y \in \mathbf{R}^d$.

Then there exists a unique continuous process X on $[0,T]$ that is adapted to $\{\mathcal{F}_t\}$ and satisfies

$$X_t = \xi + \int_0^t b(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dB_s \tag{7.51}$$

for $0 \le t \le T$.

Proof. To fit this into Theorem 7.17, let $Y(t) = [t, B_t]^T$, $H(t) = \xi$, and

$$F(t,\omega,\eta) = \big[b(t,\eta(t-)), \sigma(t,\eta(t-))\big].$$

We write $b(t,\eta(t-))$ to get a predictable coefficient. For a continuous path η this left limit is immaterial as $\eta(t-) = \eta(t)$. The hypotheses on b and σ establish the conditions required by the existence and uniqueness theorem.

Once there is a cadlag solution X from the theorem, continuity of X follows by observing that the right-hand side of (7.51) is a continuous process. $\square$

7.4. Existence and uniqueness for a semimartingale equation

This section proves Theorem 7.17. The key part of both the existence and uniqueness proof is a Gronwall-type estimate, which we derive after some technical preliminaries. Note that we can assume the Lipschitz constant $L > 0$. If $L = 0$ then $F(s, \omega, X)$ does not depend on X, equation (7.45) directly defines the process X and there is nothing to prove.

7.4.1. Technical preliminaries.

Lemma 7.23. *Let X be a cadlag process and Y a cadlag semimartingale. Fix $t > 0$. Let γ be a nondecreasing continuous process on $[0, t]$ such that $\gamma(0) = 0$ and $\gamma(u)$ is a stopping time for each u. Then*

$$\int_{(0,\gamma(t)]} X(s-)\, dY(s) = \int_{(0,t]} X \circ \gamma(s-)\, d(Y \circ \gamma)(s). \tag{7.52}$$

Remark 7.24. On the right the integrand is

$$X \circ \gamma(s-) = \lim_{u \nearrow s,\, u<s} X(\gamma(u))$$

which is not the same as $X(\gamma(s-))$ if the latter is interpreted as X evaluated at $\gamma(s-)$. To see this just think of $\gamma(u) = u$ and imagine X has a jump at s. Note that for the stochastic integral on the right in (7.52) the filtration changed to $\{\mathcal{F}_{\gamma(u)}\}$.

Proof. As $\{t_i^n\}$ goes through partitions of $[0, t]$ with mesh tending to zero, $\{\gamma(t_i^n)\}$ goes through partitions of $[0, \gamma(t)]$ with mesh tending to zero. By Proposition 5.37, both sides of (7.52) equal the limit of the sums

$$\sum_i X(\gamma(t_i^n))\big(Y(t_{i+1}^n) - Y(t_i^n)\big). \qquad \square$$

Lemma 7.25. *Suppose A is a nondecreasing cadlag function such that $A(0) = 0$, and Z is a nondecreasing real-valued cadlag function. Then*

$$\gamma(u) = \inf\{t \ge 0 : A(t) > u\}$$

defines a nondecreasing cadlag function with $\gamma(0) = 0$, and

$$\begin{aligned}\int_{(0,u]} Z \circ \gamma(s-)\, d(A \circ \gamma)(s) &\le \big(A(\gamma(u)) - u\big) Z \circ \gamma(u-) \\ &\quad + \int_{(0,u]} Z \circ \gamma(s-)\, ds.\end{aligned} \tag{7.53}$$

Proof. That γ is nondecreasing is immediate. To see right-continuity, let $s > \gamma(u)$. Then $A(s) > u$. Pick $\varepsilon > 0$ so that $A(s) > u + \varepsilon$. Then for $v \in [u, u+\varepsilon]$, $\gamma(v) \le s$.

Also, since $A(t) > u$ for $t > \gamma(u)$, by the cadlag property of A

$$A(\gamma(u)) \ge u. \tag{7.54}$$

Since $Z \circ \gamma$ is cadlag, the integrals (7.53) are limits of Riemann sums with integrand evaluated at the left endpoint of partition intervals (Lemma 1.12). Let $\{0 = s_0 < s_1 < \cdots < s_m = u\}$ be a partition of $[0, u]$. Next algebraic manipulations:

$$\begin{aligned}
&\sum_{i=0}^{m-1} Z(\gamma(s_i))\big(A(\gamma(s_{i+1})) - A(\gamma(s_i))\big) \\
&= \sum_{i=0}^{m-1} Z(\gamma(s_i))(s_{i+1} - s_i) \\
&\quad + \sum_{i=0}^{m-1} Z(\gamma(s_i))\big(A(\gamma(s_{i+1})) - s_{i+1} - A(\gamma(s_i)) + s_i\big) \\
&= \sum_{i=0}^{m-1} Z(\gamma(s_i))(s_{i+1} - s_i) + Z(\gamma(s_{m-1}))\big(A(\gamma(u)) - u\big) \\
&\quad - \sum_{i=1}^{m-1} \big(Z(\gamma(s_i)) - Z(\gamma(s_{i-1}))\big)\big(A(\gamma(s_i)) - s_i\big).
\end{aligned}$$

The sum on the last line above is nonnegative by the nondecreasing monotonicity of $Z \circ \gamma$ and by (7.54). Thus we have

$$\begin{aligned}
&\sum_{i=0}^{m-1} Z(\gamma(s_i))\big(A(\gamma(s_{i+1})) - A(\gamma(s_i))\big) \\
&\qquad \le \sum_{i=0}^{m-1} Z(\gamma(s_i))(s_{i+1} - s_i) + Z(\gamma(s_{m-1}))\big(A(\gamma(u)) - u\big).
\end{aligned}$$

Letting the mesh of the partition to zero turns the inequality above into (7.53). □

Let F be a cadlag BV function on $[0, T]$ and Λ_F its Lebesgue-Stieltjes measure. The total variation function of F was denoted by $V_F(t)$. The total variation measure $|\Lambda_F|$ satisfies $|\Lambda_F| = \Lambda_{V_F}$. For Lebesgue-Stieltjes integrals the general inequality (1.12) for signed measures can be written in this form:

$$\left| \int_{(0,T]} g(s)\, dF(s) \right| \le \int_{(0,T]} |g(s)|\, dV_F(s). \tag{7.55}$$

Lemma 7.26. *Let X be an adapted cadlag process and $\alpha > 0$. Let $\tau_1 < \tau_2 < \tau_3 < \cdots$ be the times of successive jumps in X of magnitude above α: with $\tau_0 = 0$,*

$$\tau_k = \inf\{t > \tau_{k-1} : |X(t) - X(t-)| > \alpha\}.$$

Then the $\{\tau_k\}$ are stopping times.

Proof. Fix k. We show $\{\tau_k \le t\} \in \mathcal{F}_t$. Consider the event

$$A = \bigcup_{\ell\ge1}\bigcup_{m\ge1}\bigcup_{N\ge1}\bigcap_{n\ge N}\Big\{\text{there exist integers } 0 < u_1 < u_2 < \cdots < u_k \le n \text{ such that } u_i - u_{i-1} \ge n/\ell, \text{ and } \Big|X\Big(\frac{u_i t}{n}\Big) - X\Big(\frac{u_i t}{n} - \frac{t}{n}\Big)\Big| > \alpha + \frac{1}{m}\Big\}.$$

We claim that $A = \{\tau_k \le t\}$.

Suppose $\omega \in A$. Let $s_{n,i} = u_i t/n$, $1 \le i \le k$, be the points whose existence follows for all $n \ge N$. Pass to a subsequence $\{n_j\}$ such that for each $1 \le i \le k$ we have convergence $s_{n_j,i} \to t_i \in [0,t]$ as $j \to \infty$. From the description of A there is an $1 \le \ell < \infty$ such that $t_i - t_{i-1} \ge t/\ell$. The convergence also forces $s_{n_j,i} - t/n \to t_i$. Since the increment in the path X cannot shrink below $\alpha + 1/m$ across the two points $s_{n_j,i} - t/n$ and $s_{n_j,i}$, it must be the case that $|X(t_i) - X(t_i-)| \ge \alpha + 1/m$.

Consequently the points $t_1, t_2, \ldots, t_k$ are times of jumps of magnitude above α, and thereby $\tau_k \le t$.

Conversely, suppose $\omega \in \{\tau_k \le t\}$ and let $t_1 < t_2 < \cdots < t_k$ be times of jumps of magnitude above α in $[0,t]$. Pick $\varepsilon > 0$ so that $|X(t_i) - X(t_i-)| \ge \alpha + 2\varepsilon$ and $t_i - t_{i-1} > 4\varepsilon$ for $2 \le i \le k$. Pick $\delta \in (0,\varepsilon)$ such that $r \in [t_i - \delta, t_i)$ and $s \in [t_i, t_i + \delta]$ imply

$$|X(s) - X(r)| > \alpha + \varepsilon.$$

Now let $\ell > t/\varepsilon$, $m > 1/\varepsilon$ and $N > t/\delta$. Then for $n \ge N$ find integers u_i such that

$$(u_i - 1)\frac{t}{n} < t_i \le u_i\frac{t}{n}.$$

Then also

$$t_i - \delta < (u_i - 1)\frac{t}{n} < t_i \le u_i\frac{t}{n} < t_i + \delta$$

from which

$$\Big|X\Big(\frac{u_i t}{n}\Big) - X\Big(\frac{u_i t}{n} - \frac{t}{n}\Big)\Big| > \alpha + \frac{1}{m}.$$

This shows $\omega \in A$. ∎

We need to consider equations for stopped processes, and also equations that are “restarted” at a stopping time.

Lemma 7.27. *Suppose F satisfies Assumption 7.16 and let*

$$\xi(t) = \int_{(0,t]} F(s, X)\, dY(s).$$

Let τ be a finite stopping time. Then

$$\xi^{\tau-}(t) = \int_{(0,t]} F(s, X^{\tau-})\, dY^{\tau-}(s). \tag{7.56}$$

In particular, suppose X satisfies equation (7.45). Then the equation continues to hold when all processes are stopped at $\tau-$. In other words,

$$X^{\tau-}(t) = H^{\tau-}(t) + \int_{(0,t]} F(s, X^{\tau-})\, dY^{\tau-}(s). \tag{7.57}$$

Proof. It suffices to check (7.56), the second conclusion is then immediate. Apply part (b) of Proposition 5.50 with $G(s) = F(s, X^{\tau-})$ and $J(s) = F(s, X)$. By the precise form of the Lipschitz property (7.46) of F, which is true for each fixed ω, for $0 \le s \le \tau$ we have

$$\begin{aligned} |F(s, X^{\tau-}) - F(s, X)| &\le L \cdot \sup_{u \in [0,s)} |X_s^{\tau-} - X_s| \\ &\le L \cdot \sup_{u \in [0,\tau)} |X_s - X_s| = 0. \end{aligned}$$

Then first by part (b), then by part (a) of Proposition 5.50,

$$\xi^{\tau-} = (J \cdot Y)^{\tau-} = (G \cdot Y)^{\tau-} = G \cdot Y^{\tau-}.$$ ∎

Lemma 7.28. *Assume the filtration $\{\mathcal{F}_t\}$ is right-continuous. Let σ be a finite stopping time for $\{\mathcal{F}_t\}$ and $\bar{\mathcal{F}}_t = \mathcal{F}_{\sigma+t}$.*

(a) *Let ν be a stopping time for $\{\bar{\mathcal{F}}_t\}$. Then $\sigma + \nu$ is a stopping time for $\{\mathcal{F}_t\}$ and $\bar{\mathcal{F}}_\nu \subseteq \mathcal{F}_{\sigma+\nu}$.*

(b) *Suppose τ is a stopping time for $\{\mathcal{F}_t\}$. Then $\nu = (\tau - \sigma)^+$ is an $\mathcal{F}_\tau$-measurable random variable, and an $\{\bar{\mathcal{F}}_t\}$-stopping time.*

(c) *Let Z be a cadlag process adapted to $\{\mathcal{F}_t\}$ and $\bar{Z}$ a cadlag process adapted to $\{\bar{\mathcal{F}}_t\}$. Define*

$$X(t) = \begin{cases} Z(t), & t < \sigma \\ Z(\sigma) + \bar{Z}(t - \sigma), & t \ge \sigma. \end{cases}$$

Then X is a cadlag process adapted to $\{\mathcal{F}_t\}$.

Proof. Part (a). Let $A \in \bar{\mathcal{F}}_\nu$. Observe that

$$A \cap \{\sigma + \nu < t\} = \bigcup_{r \in (0,t) \cap \mathbf{Q}} A \cap \{\nu < r\} \cap \{\sigma \le t - r\}.$$

(If $\varepsilon > 0$ satisfies $\sigma + \nu < t - \varepsilon$, then any rational $r \in (\nu, \nu + \varepsilon)$ will do.) By the definition of $\bar{\mathcal{F}}_\nu$,

$$A \cap \{\nu < r\} = \bigcup_{m \in \mathbf{N}} A \cap \{\nu \le r - \tfrac{1}{m}\} \in \bar{\mathcal{F}}_r = \mathcal{F}_{\sigma + r},$$

and consequently by the definition of $\mathcal{F}_{\sigma+r}$,

$$A \cap \{\nu < r\} \cap \{\sigma + r \le t\} \in \mathcal{F}_t.$$

If $A = \Omega$, we have showed that $\{\sigma + \nu < t\} \in \mathcal{F}_t$. By Lemma 2.6 and the right-continuity of $\{\mathcal{F}_t\}$, this implies that $\sigma + \nu$ is an $\{\mathcal{F}_t\}$-stopping time.

For the general $A \in \bar{\mathcal{F}}_\nu$ we have showed that

$$A \cap \{\sigma + \nu \le t\} = \bigcap_{m \ge n} A \cap \{\sigma + \nu < t + \tfrac{1}{m}\} \in \mathcal{F}_{t+(1/n)}.$$

Since this is true for any n,

$$A \cap \{\sigma + \nu \le t\} \in \bigcap_n \mathcal{F}_{t+(1/n)} = \mathcal{F}_t$$

where we used the right-continuity of $\{\mathcal{F}_t\}$ again.

Part (b). For $0 \le t < \infty$, $\{(\tau - \sigma)^+ \le t\} = \{\sigma + t \ge \tau\}$. By part (ii) of Lemma 2.3 this event lies in both $\mathcal{F}_\tau$ and $\mathcal{F}_{\sigma+t} = \bar{\mathcal{F}}_t$. This second part implies that $\nu = (\tau - \sigma)^+$ is an $\{\bar{\mathcal{F}}_t\}$-stopping time.

Part (c). Fix $0 \le t < \infty$ and a Borel set B on the state space of these processes.

$$\{X(t) \in B\} = \{\sigma > t,\, Z(t) \in B\} \cup \{\sigma \le t,\, Z(\sigma) + \bar{Z}(t - \sigma) \in B\}.$$

The first part $\{\sigma > t,\, Z(t) \in B\}$ lies in $\mathcal{F}_t$ because it is the intersection of two sets in $\mathcal{F}_t$. Let $\nu = (t - \sigma)^+$. The second part can be written as

$$\{\sigma \le t,\, Z(\sigma) + \bar{Z}(\nu) \in B\}.$$

Cadlag processes are progressively measurable, hence $Z(\sigma)$ is $\mathcal{F}_\sigma$-measurable and $\bar{Z}(\nu)$ is $\bar{\mathcal{F}}_\nu$-measurable (part (iii) of Lemma 2.3). Since $\sigma \le \sigma + \nu$, $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma+\nu}$. By part (a) $\bar{\mathcal{F}}_\nu \subseteq \mathcal{F}_{\sigma+\nu}$. Consequently $Z(\sigma) + \bar{Z}(\nu)$ is $\mathcal{F}_{\sigma+\nu}$-measurable. Since $\sigma \le t$ is equivalent to $\sigma + \nu \le t$, we can rewrite the second part once more as

$$\{\sigma + \nu \le t\} \cap \{Z(\sigma) + \bar{Z}(\nu) \in B\}.$$

By part (i) of Lemma 2.3 applied to the stopping times $\sigma + \nu$ and t, the set above lies in $\mathcal{F}_t$.

This concludes the proof that $\{X(t) \in B\} \in \mathcal{F}_t$. ∎

7.4.2. A Gronwall estimate for semimartingale equations. In this section we prove a Gronwall-type estimate for SDE's under fairly stringent assumptions on the driving semimartingale. When the result is applied, the assumptions can be relaxed through localization arguments. All the processes are defined for $0 \le t < \infty$. F is assumed to satisfy the conditions of Assumption 7.16, and in particular L is the Lipschitz constant of F. In this section we work with a class of semimartingales that satisfies the following definition.

Definition 7.29. Let $0 < \delta, K < \infty$ be constants. Let us say an $\mathbf{R}^m$-valued cadlag semimartingale $Y = (Y_1, \dots, Y_m)^T$ is of *type* (δ, K) if Y has a decomposition $Y = Y(0) + M + S$ where $M = (M_1, \dots, M_m)^T$ is an m-vector of L^2-martingales, $S = (S_1, \dots, S_m)^T$ is an m-vector of FV processes, and

$$|\Delta M_j(t)| \le \delta, \ |\Delta S_j(t)| \le \delta, \text{ and } V_t(S_j) \le K \tag{7.58}$$

for all $0 \le t < \infty$ and $1 \le j \le m$, almost surely.

This notion of (δ, K) type is of no general significance. We use it as a convenient abbreviation in the existence and uniqueness proofs that follow. As functions of the various constants that have been introduced, define the increasing process

$$A(t) = 16L^2 dm \sum_{j=1}^{m} [M_j]_t + 4KL^2 dm \sum_{j=1}^{m} V_{S_j}(t) + t, \tag{7.59}$$

the stopping times

$$\gamma(u) = \inf\{t \ge 0 : A(t) > u\}, \tag{7.60}$$

and the constant

$$c = c(\delta, K, L) = 16\delta^2 L^2 dm^2 + 4\delta K L^2 dm^2. \tag{7.61}$$

Let us make some observations about $A(t)$ and $\gamma(u)$. The term t is added to $A(t)$ to give $A(t) \ge t$ and make $A(t)$ strictly increasing. Then $A(u + \varepsilon) \ge u + \varepsilon > u$ for all $\varepsilon > 0$ gives $\gamma(u) \le u$. To see that $\gamma(u)$ is a stopping time, observe that $\{\gamma(u) \le t\} = \{A(t) \ge u\}$. First, assuming $\gamma(u) \le t$, by (7.54) and monotonicity $A(t) \ge A(\gamma(u)) \ge u$. Second, if $A(t) \ge u$ then by strict monotonicity $A(s) > u$ for all $s > t$, which implies $\gamma(u) \le t$.

Strict monotonicity of A gives continuity of γ. Right continuity of γ was already argued in Lemma 7.25. For left continuity, let $s < \gamma(u)$. Then $A(s) < u$ because $A(s) \ge u$ together with strict increasingness would imply the existence of a point $t \in (s, \gamma(u))$ such that $A(t) > u$, contradicting $t < \gamma(u)$. Then $\gamma(v) \ge s$ for $v \in (A(s), u)$ which shows left continuity.

In summary: $u \mapsto \gamma(u)$ is a continuous nondecreasing function of bounded stopping times such that $\gamma(u) \le u$. For any given ω and T, once $u > A(T)$ we have $\gamma(u) \ge T$, and so $\gamma(u) \to \infty$ as $u \to \infty$.

If Y is of type (δ_0, K_0) for any $\delta_0 \le \delta$ and $K_0 \le K$ then all jumps of A satisfy $|\Delta A(t)| \le c$. This follows because the jumps of quadratic variation and total variation obey $\Delta[M](t) = (\Delta M(t))^2$ and $\Delta V_{S_j}(t) = |\Delta S_j(t)|$.

For $\ell = 1, 2$, let H_ℓ, X_ℓ, and Z_ℓ be adapted $\mathbf{R}^d$-valued cadlag processes. Assume they satisfy the equations

$$Z_\ell(t) = H_\ell(t) + \int_{(0,t]} F(s, X_\ell)\, dY(s), \quad \ell = 1, 2, \tag{7.62}$$

for all $0 \le t < \infty$. Let

$$D_X(t) = \sup_{0 \le s \le t} |X_1(s) - X_2(s)|^2 \tag{7.63}$$

and

$$\phi_X(u) = E\big[D_X \circ \gamma(u)\big] = E\Big[\sup_{0 \le s \le \gamma(u)} |X_1(s) - X_2(s)|^2\Big]. \tag{7.64}$$

Make the same definitions with X replaced by Z and H. D_X, D_Z, and D_H are nonnegative, nondecreasing cadlag processes. ϕ_X, ϕ_Z, and ϕ_H are nonnegative nondecreasing functions, and cadlag at least on any interval on which they are finite. We assume that

$$\phi_H(u) = E\Big[\sup_{0 \le s \le \gamma(u)} |H_1(s) - H_2(s)|^2\Big] < \infty \tag{7.65}$$

for all $0 \le u < \infty$.

The proposition below is the key tool for both the uniqueness and existence proofs. Part (b) is a Gronwall type estimate for SDE's.

Proposition 7.30. *Suppose F satisfies Assumption 7.16 and Y is a semimartingale of type $(\delta, K - \delta)$ in Definition 7.29. Furthermore, assume (7.65), and let the pairs (X_ℓ, Z_ℓ) satisfy (7.62).*

(a) *For $0 \le u < \infty$,*

$$\phi_Z(u) \le 2\phi_H(u) + c\phi_X(u) + \int_0^u \phi_X(s)\, ds. \tag{7.66}$$

(b) *Suppose $Z_\ell = X_\ell$, in other words X_1 and X_2 satisfy the equations*

$$X_\ell(t) = H_\ell(t) + \int_{(0,t]} F(s, X_\ell)\, dY(s), \quad \ell = 1, 2. \tag{7.67}$$

Then $\phi_X(u) < \infty$. If δ is small enough relative to K and L so that $c < 1$, then for all $u > 0$,

$$\phi_X(u) \le \frac{2\phi_H(u)}{1 - c} \exp\Big\{\frac{u}{1 - c}\Big\}. \tag{7.68}$$

Before starting the proof of the proposition, we establish an auxiliary lemma.

Lemma 7.31. *Let $0 < \delta, K < \infty$ and suppose Y is a semimartingale of type (δ, K) as specified in Definition 7.29. Let $G = \{G_{i,j}\}$ be a bounded predictable $d \times m$-matrix valued process. Then for $0 \le u < \infty$ we have the bounds*

$$E\biggl[\sup_{0\le t\le\gamma(u)}\biggl|\int_{(0,t]} G(s)\,dY(s)\biggr|^2\biggr] \le \tfrac{1}{2}L^{-2}E\int_{(0,\gamma(u)]}|G(s)|^2\,dA(s) \le \tfrac{1}{2}L^{-2}(u+c)\|G\|_\infty^2. \tag{7.69}$$

Proof of Lemma 7.31. By the definition of Euclidean norm and by the inequality $(x+y)^2 \le 2x^2 + 2y^2$,

$$\biggl|\int_{(0,t]} G(s)\,dY(s)\biggr|^2 \le 2\sum_{i=1}^d\biggl(\sum_{j=1}^m\int_{(0,t]} G_{i,j}(s)\,dM_j(s)\biggr)^2 + 2\sum_{i=1}^d\biggl(\sum_{j=1}^m\int_{(0,t]} G_{i,j}(s)\,dS_j(s)\biggr)^2. \tag{7.70}$$

The first sum inside parentheses after the inequality in (7.70) is an L^2-martingale by the boundedness of G and because each M_j is an L^2-martingale. Take supremum over $0 \le t \le \gamma(u)$, on the right pass the supremum inside the i-sums, and take expectations. By Doob's inequality (3.13), Schwarz inequality in the form $(\sum_{i=1}^m a_i)^2 \le m\sum_{i=1}^m a_i^2$, and by the isometric property of stochastic integrals,

$$\begin{aligned}
&E\biggl[\sup_{0\le t\le\gamma(u)}\biggl(\sum_{j=1}^m\int_{(0,t]} G_{i,j}(s)\,dM_j(s)\biggr)^2\biggr]\\
&\le 4E\biggl[\biggl(\sum_{j=1}^m\int_{(0,\gamma(u)]} G_{i,j}(s)\,dM_j(s)\biggr)^2\biggr]\\
&\le 4m\sum_{j=1}^m E\biggl[\biggl(\int_{(0,\gamma(u)]} G_{i,j}(s)\,dM_j(s)\biggr)^2\biggr]\\
&\le 4m\sum_{j=1}^m E\int_{(0,\gamma(u)]} G_{i,j}(s)^2\,d[M_j](s).
\end{aligned} \tag{7.71}$$

Similarly handle the expectation of the last sum in (7.70). Instead of quadratic variation, use (7.55) and Schwarz another time.

$$
\begin{aligned}
&E\bigg[\sup_{0\le t\le\gamma(u)}\bigg(\sum_{j=1}^{m}\int_{(0,t]}G_{i,j}(s)\,dS_j(s)\bigg)^2\bigg]\\
&\le m\sum_{j=1}^{m}E\bigg[\sup_{0\le t\le\gamma(u)}\bigg(\int_{(0,t]}G_{i,j}(s)\,dS_j(s)\bigg)^2\bigg]\\
&\le m\sum_{j=1}^{m}E\bigg[\bigg(\int_{(0,\gamma(u)]}\big|G_{i,j}(s)\big|\,dV_{S_j}(s)\bigg)^2\bigg]\\
&\le m\sum_{j=1}^{m}E\bigg[V_{S_j}(\gamma(u))\int_{(0,\gamma(u)]}G_{i,j}(s)^2\,dV_{S_j}(s)\bigg].
\end{aligned}
\tag{7.72}
$$

Now we prove (7.69). Equations (7.70), (7.71) and (7.72), together with the hypothesis $V_{S_j}(t)\le K$, imply that

$$
\begin{aligned}
E\bigg[&\sup_{0\le t\le\gamma(u)}\bigg|\int_{(0,t]}G(s)\,dY(s)\bigg|^2\bigg]\\
&\le 8dm\sum_{j=1}^{m}E\int_{(0,\gamma(u)]}|G(s)|^2\,d[M_j](s)\\
&\qquad\qquad + 2Kdm\sum_{j=1}^{m}E\int_{(0,\gamma(u)]}|G(s)|^2\,dV_{S_j}(s)\\
&\le \tfrac12 L^{-2}E\int_{(0,\gamma(u)]}|G(s)|^2\,dA(s)\\
&\le \tfrac12 L^{-2}E\Big[\sup_{0\le t\le\gamma(u)}|G(s)|^2\cdot A(\gamma(u))\Big]\\
&\le \tfrac12 L^{-2}(u+c)\|G\|_\infty^2.
\end{aligned}
$$

From $A(\gamma(u)-)\le u$ and the bound c on the jumps in A came the bound $A(\gamma(u))\le u+c$ used above. This completes the proof of Lemma 7.31. □

Proof of Proposition 7.30. Step 1. We prove the proposition first under the additional assumption that there exists a constant C_0 such that $|F|\le C_0$, and under the relaxed assumption that Y is of type (δ,K). This small relaxation accommodates the localization argument that in the end removes the boundedness assumptions on F.

Use the inequality $(x+y)^2 \le 2x^2 + 2y^2$ to write

$$|Z_1(t) - Z_2(t)|^2 \le 2|H_1(t) - H_2(t)|^2 + 2\left| \int_{(0,t]} \big(F(s,X_1) - F(s,X_2)\big)\, dY(s) \right|^2. \tag{7.73}$$

We first check that

$$\phi_Z(u) = E\Big[\sup_{0\le t\le \gamma(u)} |Z_1(t) - Z_2(t)|^2 \Big] < \infty \quad \text{for all } 0 \le u < \infty. \tag{7.74}$$

Combine (7.69) with the hypothesis $|F| \le C_0$ to get the bound

$$E\bigg[\sup_{0\le t\le \gamma(u)} \left| \int_{(0,t]} \big(F(s,X_1) - F(s,X_2)\big)\, dY(s) \right|^2 \bigg] \le 2L^{-2}(u+c)C_0^2. \tag{7.75}$$

Now (7.74) follows from a combination of inequality (7.73), assumption (7.65), and bound (7.75). Note that (7.74) does not require a bound on X, due to the boundedness assumption on F and (7.65).

The Lipschitz assumption on F gives

$$\big|F(s,X_1) - F(s,X_2)\big|^2 \le L^2 D_X(s-).$$

Apply (7.69) together with the Lipschitz bound to get

$$\begin{aligned} &E\bigg[\sup_{0\le t\le \gamma(u)} \left| \int_{(0,t]} \big(F(s,X_1) - F(s,X_2)\big)\, dY(s) \right|^2 \bigg] \\ &\qquad \le \tfrac{1}{2} L^{-2} E \int_{(0,\gamma(u)]} \big|F(s,X_1) - F(s,X_2)\big|^2 \, dA(s) \\ &\qquad \le \tfrac{1}{2} E \int_{(0,\gamma(u)]} D_X(s-)\, dA(s). \end{aligned} \tag{7.76}$$

Now we prove part (a) under the assumption of bounded F. Take supremum over $0 \le t \le \gamma(u)$ in (7.73), take expectations, and apply (7.76) to write

$$\begin{aligned} \phi_Z(u) &= E\big[D_Z \circ \gamma(u)\big] = E\Big[\sup_{0\le t\le \gamma(u)} |Z_1(t) - Z_2(t)|^2 \Big] \\ &\le 2\phi_H(u) + E \int_{(0,\gamma(u)]} D_X(s-)\, dA(s). \end{aligned} \tag{7.77}$$

To the dA-integral above apply first the change of variable from Lemma 7.23 and then inequality (7.53). This gives

$$\begin{aligned} \phi_Z(u) &\le 2\phi_H(u) \\ &\quad + E\big[\big(A(\gamma(u)) - u\big) D_X \circ \gamma(u-)\big] + E \int_{(0,u]} D_X \circ \gamma(s-)\, ds. \end{aligned} \tag{7.78}$$

For a fixed ω, cadlag paths are bounded on bounded time intervals, so applying Lemmas 7.23 and 7.25 to the path-by-path integral $\int_{(0,\gamma(u)]} D_X(s-)\,dA(s)$ is not problematic. And then, since the resulting terms are nonnegative, their expectations exist.

By the definition of $\gamma(u)$, $A(s) \le u$ for $s < \gamma(u)$, and so $A(\gamma(u)-) \le u$. Thus by the bound c on the jumps of A,

$$A(\gamma(u)) - u \le A(\gamma(u)) - A(\gamma(u)-) \le c.$$

Applying this to (7.78) gives

$$\begin{aligned} \phi_Z(u) &\le 2\phi_H(u) + cE\big[D_X \circ \gamma(u-)\big] + \int_{(0,u]} E\big[D_X \circ \gamma(s-)\big]\,ds \\ &\le 2\phi_H(u) + c\phi_X(u) + \int_{(0,u]} \phi_X(u)\,ds. \end{aligned} \tag{7.79}$$

This is the desired conclusion (7.66), and the proof of part (a) for the case of bounded F is complete.

We prove part (b) for bounded F. By assumption, $Z_\ell = X_\ell$ and $c < 1$. Now $\phi_X(u) = \phi_Z(u)$, and by (7.74) this function is finite. Since it is nondecreasing, it is bounded on bounded intervals. Inequality (7.79) becomes

$$(1-c)\phi_X(u) \le 2\phi_H(u) + \int_0^u \phi_X(s)\,ds.$$

An application of Gronwall's inequality (Lemma A.20) gives the desired conclusion (7.68). This completes the proof of the proposition for the case of bounded F.

Step 2. Return to the original hypotheses: Assumption 7.16 for F without additional boundedness and Definition 7.29 for Y with type $(\delta, K-\delta)$. By part (iii) of Assumption 7.16, we can pick stopping times $\sigma_k \nearrow \infty$ and constants B_k such that

$$\big|\,\mathbf{1}_{(0,\sigma_k]}(s)F(s,X_\ell)\,\big| \le B_k \quad \text{for } \ell = 1, 2.$$

Define truncated functions by

$$F_{B_k}(s,\omega,\eta) = \big\{F(s,\omega,\eta) \wedge B_k\big\} \vee (-B_k). \tag{7.80}$$

By Lemma 7.27, assumption (7.62) gives the equations

$$Z_\ell^{\sigma_k-}(t) = H_\ell^{\sigma_k-}(t) + \int_{(0,t]} F(s, X_\ell^{\sigma_k-})\,dY^{\sigma_k-}(s), \quad \ell \in \{1,2\}.$$

Since $X_\ell = X_\ell^{\sigma_k-}$ on $[0,\sigma_k)$, $F(s,X_\ell^{\sigma_k-}) = F(s,X_\ell)$ on $[0,\sigma_k]$. The truncation has no effect on $F(s,X_\ell)$ if $0 \le s \le \sigma_k$, and so also

$$F(s, X_\ell^{\sigma_k-}) = F_{B_k}(s, X_\ell^{\sigma_k-}) \quad \text{on } [0,\sigma_k].$$

By part (b) of Proposition 5.50 we can perform this substitution in the integral, and get the equations

$$Z_\ell^{\sigma_k-}(t) = H_\ell^{\sigma_k-}(t) + \int_{(0,t]} F_{B_k}(s, X_\ell^{\sigma_k-})\,dY^{\sigma_k-}(s), \quad \ell \in \{1,2\}.$$

Now we have equations with bounded coefficients in the integral as required for Step 1.

We need to check what happens to the type in Definition 7.29 as we replace Y with Y^{σ_k-}. Originally Y was of type $(\delta, K-\delta)$. Decompose Y^{σ_k-} as

$$Y^{\sigma_k-}(t) = Y_0 + M^{\sigma_k}(t) + S^{\sigma_k-}(t) - \Delta M(\sigma_k)\mathbf{1}_{\{t\ge\sigma_k\}}.$$

The new martingale part M^{σ_k} is still an L^2-martingale. Its jumps are a subset of the jumps of M, hence bounded by δ. The jth component of the new FV part is $G_j(t) = S_j^{\sigma_k-}(t) - \Delta M_j(\sigma_k)\mathbf{1}_{\{t\ge\sigma_k\}}$. It has jumps of S_j on $[0,\sigma_k)$ and the jump of M_j at σ_k, hence all bounded by δ. The total variation V_{G_j} is at most $V_{S_j} + |\Delta M_j(\sigma_k)| \le K - \delta + \delta = K$. We conclude that Y^{σ_k-} is of type (δ, K).

We have verified all the hypotheses of Step 1 for the stopped processes and the function F_{B_k}. Consequently Step 1 applies and gives parts (a) and (b) for the stopped processes $Z_\ell^{\sigma_k-}$, $H_\ell^{\sigma_k-}$, and $X_\ell^{\sigma_k-}$. Note that

$$\begin{aligned} E\Big[\sup_{0\le s\le\gamma(u)} |X_1^{\sigma_k-}(s) - X_2^{\sigma_k-}(s)|^2\Big] &= E\Big[\sup_{0\le s\le\gamma(u),\, s<\sigma_k} |X_1(s) - X_2(s)|^2\Big] \\ &\le E\Big[\sup_{0\le s\le\gamma(u)} |X_1(s) - X_2(s)|^2\Big] = \phi_X(u) \end{aligned}$$

while

$$\lim_{k\to\infty} E\Big[\sup_{0\le s\le\gamma(u)} |X_1^{\sigma_k-}(s) - X_2^{\sigma_k-}(s)|^2\Big] = \phi_X(u)$$

by monotone convergence because $\sigma_k \nearrow \infty$. Same facts hold of course for Z and H too.

Using the previous inequality, the outcome from Step 1 can be written for part (a) as

$$\begin{aligned} &E\Big[\sup_{0\le s\le\gamma(u)} |Z_1^{\sigma_k-}(s) - Z_2^{\sigma_k-}(s)|^2\Big] \\ &\qquad\qquad \le 2\phi_H(u) + c\phi_X(u) + \int_0^u \phi_X(s)\,ds. \end{aligned}$$

and for part (b) as

$$E\Big[\sup_{0\le s\le\gamma(u)} |X_1^{\sigma_k-}(s) - X_2^{\sigma_k-}(s)|^2\Big] \le \frac{2\phi_H(u)}{1-c}\exp\Big\{\frac{u}{1-c}\Big\}.$$

As $k \nearrow \infty$ the left-hand sides of the inequalities above converge to the desired expectations. Parts (a) and (b) both follow, and the proof is complete. ☐

7.4.3. The uniqueness theorem.

Theorem 7.32. *Let H be an $\mathbf{R}^d$-valued adapted cadlag process and Y an $\mathbf{R}^m$-valued cadlag semimartingale. Assume F satisfies Assumption 7.16. Suppose X_1 and X_2 are two adapted $\mathbf{R}^d$-valued cadlag processes, and both are solutions to the equation*

$$X(t) = H(t) + \int_{(0,t]} F(s, X)\, dY(s), \quad 0 \le t < \infty. \tag{7.81}$$

Then for almost every ω, $X_1(t, \omega) = X_2(t, \omega)$ for all $0 \le t < \infty$.

Proof. At time zero $X_1(0) = H(0) = X_2(0)$.

Step 1. We first show that there exists a stopping time σ, defined in terms of Y, such that $P\{\sigma > 0\} = 1$, and for all choices of H and F, and for all solutions X_1 and X_2, $X_1(t) = X_2(t)$ for $0 \le t \le \sigma$.

To prove this, we stop Y so that the hypotheses of Proposition 7.30 are satisfied. Choose $0 < \delta < K/3 < \infty$ so that $c < 1$ for c defined by (7.61). By Theorem 3.21 (fundamental theorem of local martingales) we can choose the semimartingale decomposition $Y = Y(0) + M + G$ so that the local L^2-martingale M has jumps bounded by $\delta/2$. Fix a constant $0 < C < \infty$. Define the following stopping times:

$$\tau_1 = \inf\{t > 0 : |Y(t) - Y(0)| \ge C \text{ or } |Y(t-) - Y(0)| \ge C\},$$
$$\tau_2 = \inf\{t > 0 : |\Delta Y(t)| > \delta/2\},$$

and

$$\tau_3 = \inf\{t > 0 : V_{G_j}(t) \ge K - 2\delta \text{ for some } 1 \le j \le m\}.$$

Lemmas 2.9 and 7.26 guarantee that τ_1 and τ_2 are stopping times. For τ_3, observe that since $V_{G_j}(t)$ is nondecreasing and cadlag,

$$\{\tau_3 \le t\} = \bigcup_{j=1}^{m} \{V_{G_j}(t) \ge K - 2\delta\}.$$

Each of τ_1, τ_2 and τ_3 is strictly positive. A cadlag path satisfies $|Y(s) - Y(0)| < C$ for $s \le \delta$ for a positive δ that depends on the path, but then $\tau_1 \ge \delta$. The interval $[0, 1]$ contains only finitely many jumps of Y of magnitude above $\delta/2$, so there must be a first one, and this occurs at some positive time because $Y(0+) = Y(0)$. Finally, total variation V_{G_j} is cadlag and so $V_{G_j}(0+) = V_{G_j}(0) = 0$, hence $\tau_3 > 0$.

Let T be an arbitrary finite positive number and

$$\sigma = \tau_1 \wedge \tau_2 \wedge \tau_3 \wedge T. \tag{7.82}$$

$P(\sigma > 0) = 1$ by what was said above.

We claim that semimartingale $Y^{\sigma-}$ satisfies the hypotheses of Proposition 7.30. To see this, decompose $Y^{\sigma-}$ as

$$Y^{\sigma-}(t) = Y(0) + M^{\sigma}(t) + G^{\sigma-}(t) - \Delta M(\sigma) \cdot \mathbf{1}_{\{t \geq \sigma\}}. \tag{7.83}$$

Stopping does not introduce new jumps, so the jumps of M^σ are still bounded by $\delta/2$. The jumps of $Y^{\sigma-}$ are bounded by $\delta/2$ since $\sigma \leq \tau_2$. On $[0, \sigma)$ the FV part $S(t) = G^{\sigma-}(t) - \Delta M(\sigma) \cdot \mathbf{1}_{\{t \geq \sigma\}}$ has the jumps of $G^{\sigma-}$. These are bounded by δ because $\Delta G^{\sigma-}(t) = \Delta Y^{\sigma-}(t) - \Delta M^{\sigma-}(t)$. At time σ S has the jump $\Delta M(\sigma)$, bounded by $\delta/2$. The total variation of a component S_j of S is

$$\begin{aligned} V_{S_j}(t) &\leq V_{G_j^{\sigma-}}(t) + |\Delta M_j(\sigma)| \leq V_{G_j}(\tau_3-) + \delta/2 \leq K - 2\delta + \delta/2 \\ &\leq K - \delta. \end{aligned}$$

Since $|Y^{\sigma-} - Y(0)| \leq C$ and $|S_j| \leq |V_{S_j}| \leq K$, it follows that M^σ is bounded, and consequently an L^2-martingale.

We have verified that $Y^{\sigma-}$ is of type $(\delta, K - \delta)$ according to Definition 7.29.

By assumption equation (7.81) is satisfied by both X_1 and X_2. By Lemma 7.27, both $X_1^{\sigma-}$ and $X_2^{\sigma-}$ satisfy the equation

$$X(t) = H^{\sigma-}(t) + \int_{(0,t]} F(s, X)\, dY^{\sigma-}(s), \quad 0 \leq t < \infty. \tag{7.84}$$

To this equation we apply Proposition 7.30. In the hypotheses of Proposition 7.30 take $H_1 = H_2 = H^{\sigma-}$, so $\phi_H(u) = 0$ and assumption (7.65) is satisfied. All the hypotheses of part (b) of Proposition 7.30 are satisfied, and we get

$$E\Big[\sup_{0 \leq t \leq \gamma(u)} |X_1^{\sigma-}(t) - X_2^{\sigma-}(t)|^2\Big] = 0$$

for any $u > 0$, where $\gamma(u)$ is the stopping time defined by (7.60). As we let $u \nearrow \infty$ we get $X_1^{\sigma-}(t) = X_2^{\sigma-}(t)$ for all $0 \leq t < \infty$. This implies that $X_1(t) = X_2(t)$ for $0 \leq t < \sigma$.

At time σ,

$$\begin{aligned} X_1(\sigma) &= H(\sigma) + \int_{(0,\sigma]} F(s, X_1)\, dY(s) \\ &= H(\sigma) + \int_{(0,\sigma]} F(s, X_2)\, dY(s) \\ &= X_2(\sigma) \end{aligned} \tag{7.85}$$

because the integrand $F(s, X_1)$ depends only on $\{X_1(s) : 0 \le s < \sigma\}$ which agrees with the corresponding segment of X_2, as established in the previous paragraph. Now we know that $X_1(t) = X_2(t)$ for $0 \le t \le \sigma$. This concludes the proof of Step 1.

Step 2. Now we show that X_1 and X_2 agree up to an arbitrary finite time T. At this stage we begin to make use of the assumption of right-continuity of $\{\mathcal{F}_t\}$. Define

$$\tau = \inf\{t \ge 0 : X_1(t) \ne X_2(t)\}. \tag{7.86}$$

The time when the cadlag process $X_1 - X_2$ first enters the open set $\{0\}^c$ is a stopping time under a right-continuous filtration by Lemma 2.7. Hence τ is a stopping time. Since $X_1 = X_2$ on $[0, \tau)$, if $\tau < \infty$, a calculation like the one in (7.85) shows that $X_1 = X_2$ on $[0, \tau]$. From Step 1 we also know $\tau \ge \sigma > 0$.

The idea is to apply Step 1 again, starting from time τ. For this we need a lemma that enables us to restart the equation.

Lemma 7.33. *Assume F satisfies Assumption 7.16 and X satisfies equation (7.45) on $[0, \infty)$. Let σ be a bounded stopping time. Define a new filtration, new processes, and a new coefficient by $\bar{\mathcal{F}}_t = \mathcal{F}_{\sigma+t}$, $\bar{X}(t) = X(\sigma+t) - X(\sigma)$, $\bar{Y}(t) = Y(\sigma + t) - Y(\sigma)$, $\bar{H}(t) = H(\sigma + t) - H(\sigma)$, and*

$$\bar{F}(t, \omega, \eta) = F(\sigma + t, \omega, \zeta^{\omega,\eta})$$

where the cadlag path $\zeta^{\omega,\eta} \in D_{\mathbf{R}^d}[0, \infty)$ is defined by

$$\zeta^{\omega,\eta}(s) = \begin{cases} X(s), & 0 \le s < \sigma \\ X(\sigma) + \eta(s - \sigma), & s \ge \sigma. \end{cases}$$

Then under $\{\bar{\mathcal{F}}_t\}$, $\bar{X}$ and $\bar{H}$ are adapted cadlag processes, $\bar{Y}$ is a semimartingale, and $\bar{F}$ satisfies Assumption 7.16. $\bar{X}$ is a solution of the equation

$$\bar{X}(t) = \bar{H}(t) + \int_{(0,t]} \bar{F}(s, \bar{X})\, d\bar{Y}(s).$$

Proof of Lemma 7.33. We check that the new $\bar{F}$ satisfies all the hypotheses. The Lipschitz property is immediate. Let $\bar{Z}$ be a cadlag process adapted to $\{\bar{\mathcal{F}}_t\}$. Define the process Z by

$$Z(t) = \begin{cases} X(t), & t < \sigma \\ X(\sigma) + \bar{Z}(t - \sigma), & t \ge \sigma. \end{cases}$$

Then Z is a cadlag process adapted to $\{\mathcal{F}_t\}$ by Lemma 7.28. $\bar{F}(t, \bar{Z}) = F(\sigma + t, Z)$ is predictable under $\{\bar{\mathcal{F}}_t\}$ by Lemma 5.46. Find stopping times $\nu_k \nearrow \infty$ such that $\mathbf{1}_{(0,\nu_k]}(s) F(s, Z)$ is bounded for each k. Define $\rho_k =$

$(\nu_k - \sigma)^+$. Then $\rho_k \nearrow \infty$, by Lemma 7.28 ρ_k is a stopping time for $\{\bar{\mathcal{F}}_t\}$, and $\mathbf{1}_{(0,\rho_k]}(s)\bar{F}(s,\bar{Z}) = \mathbf{1}_{(0,\nu_k]}(\sigma+s)F(\sigma+s,Z)$ which is bounded.

$\bar{Y}$ is a semimartingale by Theorem 5.45. $\bar{X}$ and $\bar{H}$ are adapted to $\{\bar{\mathcal{F}}_t\}$ by part (iii) of Lemma 2.3 (recall that cadlag paths imply progressive measurability).

The equation for $\bar{X}$ checks as follows:

$$\begin{aligned}
\bar{X}(t) &= X(\sigma+t) - X(\sigma) \\
&= H(\sigma+t) - H(\sigma) + \int_{(\sigma,\sigma+t]} F(s,X)\,dY(s) \\
&= \bar{H}(t) + \int_{(0,t]} F(\sigma+s,X)\,d\bar{Y}(s) \\
&= \bar{H}(t) + \int_{(0,t]} \bar{F}(s,\bar{X})\,d\bar{Y}(s).
\end{aligned}$$

The next to last equality is from (5.48), and the last equality from the definition of $\bar{F}$ and $\zeta^{\omega,\bar{X}} = X$. ☐

We return to the proof of the uniqueness theorem. Let $0 < T < \infty$. By Lemma 7.33, we can restart the equations for X_1 and X_2 at time $\sigma = \tau \wedge T$. Since $X_1 = X_2$ on $[0,\sigma]$, both X_1 and X_2 lead to the same new coefficient $\bar{F}(t,\omega,\eta)$ for the restarted equation. Consequently we have

$$\bar{X}_\ell(t) = \bar{H}(t) + \int_{(0,t]} \bar{F}(s,\bar{X}_\ell)\,d\bar{Y}(s), \quad \ell = 1,2.$$

Applying Step 1 to this restarted equation, we find a stopping time $\overline{\sigma} > 0$ in the filtration $\{\mathcal{F}_{(\tau\wedge T)+t}\}$ such that $\bar{X}_1 = \bar{X}_2$ on $[0,\overline{\sigma}]$. This implies that $X_1 = X_2$ on $[0, \tau\wedge T + \overline{\sigma}]$. Hence by definition (7.86), $\tau \ge \tau \wedge T + \overline{\sigma}$, which implies that $\tau \ge T$. Since T was arbitrary, $\tau = \infty$. This says that X_1 and X_2 agree for all time. ☐

Remark 7.34. Let us note the crucial use of the fundamental theorem of local martingales in Step 1 of the proof above. Suppose we could not choose the decomposition $Y = Y(0) + M + G$ so that M has jumps bounded by $\delta/2$. Then to satisfy the hypotheses of Proposition 7.30, we could try to stop before either M or S has jumps larger than $\delta/2$. However, this attempt runs into trouble. The stopped local martingale part M^σ in (7.83) might have a large jump exactly at time σ. Replacing M^σ with $M^{\sigma-}$ would eliminate this problem, but there is no guarantee that the process $M^{\sigma-}$ is a local martingale.

7.4.4. Existence theorem. We begin with an existence theorem under stringent assumptions on Y. These hypotheses are subsequently relaxed with a localization argument.

Proposition 7.35. *A solution to* (7.45) *on* $[0,\infty)$ *exists under the following assumptions:*

(i) F *satisfies Assumption 7.16.*

(ii) *There are constants* $0 < \delta < K < \infty$ *such that* c *defined by* (7.61) *satisfies* $c < 1$, *and* Y *is of type* $(\delta, K-\delta)$ *as specified by Definition 7.29.*

We prove this proposition in several stages. The hypotheses are chosen so that the estimate in Proposition 7.30 can be applied.

We define a Picard iteration as follows. Let $X_0(t) = H(t)$, and for $n \ge 0$,

$$X_{n+1}(t) = H(t) + \int_{(0,t]} F(s, X_n)\, dY(s). \tag{7.87}$$

We need to stop the processes in order to get bounds that force the iteration to converge. By part (ii) of Assumption 7.16 fix stopping times $\nu_k \nearrow \infty$ and constants B_k such that $\mathbf{1}_{(0,\nu_k]}(s)|F(s,H)| \le B_k$, for all k. By Lemma 7.27 equation (7.87) continues to hold for stopped processes:

$$X^{\nu_k-}_{n+1}(t) = H^{\nu_k-}(t) + \int_{(0,t]} F(s, X^{\nu_k-}_n)\, dY^{\nu_k-}(s). \tag{7.88}$$

Fix k for a while. For $n \ge 0$ let

$$D_n(t) = \sup_{0\le s\le t} |X^{\nu_k-}_{n+1}(s) - X^{\nu_k-}_n(s)|^2.$$

Recall definition (7.60), and for $0 \le u < \infty$ let

$$\phi_n(u) = E\big[D_n \circ \gamma(u)\big].$$

Lemma 7.36. *The function* $\phi_0(u)$ *is bounded on bounded intervals.*

Proof. Because ϕ_0 is nondecreasing it suffices to show that $\phi_0(u)$ is finite for any u. First,

$$\begin{aligned} |X^{\nu_k-}_1(t) - X^{\nu_k-}_0(t)|^2 &= \left| \int_{(0,t]} F(s, H^{\nu_k-})\, dY^{\nu_k-}(s) \right|^2 \\ &= \left| \int_{(0,t]} F_{B_k}(s, H^{\nu_k-})\, dY^{\nu_k-}(s) \right|^2 \end{aligned}$$

where F_{B_k} denotes the truncated function

$$F_{B_k}(s,\omega,\eta) = \big\{F(s,\omega,\eta) \wedge B_k\big\} \vee (-B_k).$$

The truncation can be introduced in the stochastic integral because on $[0,\nu_k]$

$$F(s, H^{\nu_k-}) = F(s,H) = F_{B_k}(s,H) = F_{B_k}(s, H^{\nu_k-}).$$

We get the bound

$$\begin{aligned}\phi_0(u) &= E\Big[\sup_{0\le t\le\gamma(u)}|X_1^{\nu_k-}(t)-X_0^{\nu_k-}(t)|^2\Big]\\&\le E\bigg[\sup_{0\le t\le\gamma(u)}\bigg|\int_{(0,t]}F_{B_k}(s,H)\,dY(s)\bigg|^2\bigg]\\&\le \tfrac12 L^{-2}(u+c)B_k^2.\end{aligned}$$

The last inequality above is from Lemma 7.31. □

Part (a) of Proposition 7.30 applied to $(Z_1,Z_2)=(X_n^{\nu_k-},X_{n+1}^{\nu_k-})$ and $(H_1,H_2)=(H^{\nu_k-},H^{\nu_k-})$ gives

$$\phi_{n+1}(u)\le c\phi_n(u)+\int_0^u\phi_n(s)\,ds. \tag{7.89}$$

Note that we need no assumption on H because only the difference $H^{\nu_k-}-H^{\nu_k-}=0$ appears in hypothesis (7.65) for Proposition 7.30. By the previous lemma ϕ_0 is bounded on bounded intervals. Then (7.89) shows inductively that all ϕ_n are bounded functions on bounded intervals. The next goal is to prove that $\sum_n\phi_n(u)<\infty$.

Lemma 7.37. *Fix $0<T<\infty$. Let $\{\phi_n\}$ be nonnegative measurable functions on $[0,T]$ such that $\phi_0\le B$ for some constant B, and inequality (7.89) is satisfied for all $n\ge0$ and $0\le u\le T$. Then for all n and $0\le u\le t$,*

$$\phi_n(u)\le B\sum_{k=0}^n\frac{1}{k!}\binom{n}{k}u^kc^{n-k}. \tag{7.90}$$

Proof. Use induction. (7.90) is true for $n=0$ by hypothesis. Assume it is true for n. By (7.89),

$$\begin{aligned}\phi_{n+1}(u)&\le c\phi_n(u)+\int_0^u\phi_n(s)\,ds\\&\le B\sum_{k=0}^n\frac{c^{n+1-k}}{k!}\binom{n}{k}u^k+B\sum_{k=0}^n\frac{c^{n-k}}{k!}\binom{n}{k}\int_0^u s^k\,ds\\&=B\sum_{k=0}^n\frac{c^{n+1-k}}{k!}\binom{n}{k}u^k+B\sum_{k=1}^{n+1}\frac{c^{n+1-k}}{k!}\binom{n}{k-1}u^k\\&=B\sum_{k=0}^{n+1}\frac{c^{n+1-k}}{k!}\binom{n+1}{k}u^k.\end{aligned}$$

For the last step above, combine terms and use $\binom{n}{k}+\binom{n}{k-1}=\binom{n+1}{k}$. □

To (7.90) we apply the next limit.

Lemma 7.38. *For any* $0 < \delta < 1$ *and* $0 < u < \infty$,

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{k!}\binom{n}{k}u^k\delta^{n-k} = \frac{1}{1-\delta}\cdot\exp\Bigl\{\frac{u}{1-\delta}\Bigr\}.$$

Proof. First check this auxiliary equality for $0 < x < 1$ and $k \ge 0$:

$$\sum_{m=0}^{\infty}(m+1)(m+2)\cdots(m+k)x^m = \frac{k!}{(1-x)^{k+1}}. \tag{7.91}$$

One way to see this is to write the left-hand sum as

$$\begin{aligned}
\sum_{m=0}^{\infty}\frac{d^k}{dx^k}x^{m+k} &= \frac{d^k}{dx^k}\Bigl(\sum_{m=0}^{\infty}x^{m+k}\Bigr) = \frac{d^k}{dx^k}\Bigl(\frac{x^k}{1-x}\Bigr)\\
&= \sum_{j=0}^{k}\binom{k}{j}\frac{d^j}{dx^j}\Bigl(\frac{1}{1-x}\Bigr)\cdot\frac{d^{k-j}}{dx^{k-j}}(x^k)\\
&= \sum_{j=0}^{k}\binom{k}{j}\frac{j!}{(1-x)^{j+1}}\cdot k(k-1)\cdots(j+1)x^j\\
&= \frac{k!}{1-x}\sum_{j=0}^{k}\binom{k}{j}\cdot\Bigl(\frac{x}{1-x}\Bigr)^j = \frac{k!}{1-x}\Bigl(1+\frac{x}{1-x}\Bigr)^k\\
&= \frac{k!}{(1-x)^{k+1}}.
\end{aligned}$$

For an alternative proof of (7.91) see Exercise 7.8.

After changing the order of summation, the sum in the statement of the lemma becomes

$$\begin{aligned}
&\sum_{k=0}^{\infty}\frac{u^k}{(k!)^2}\sum_{n=k}^{\infty}n(n-1)\cdots(n-k+1)\delta^{n-k}\\
&\quad= \sum_{k=0}^{\infty}\frac{u^k}{(k!)^2}\sum_{m=0}^{\infty}(m+1)(m+2)\cdots(m+k)\delta^m\\
&\quad= \sum_{k=0}^{\infty}\frac{u^k}{(k!)^2}\cdot\frac{k!}{(1-\delta)^{k+1}} = \frac{1}{1-\delta}\sum_{k=0}^{\infty}\frac{1}{k!}\Bigl(\frac{u}{1-\delta}\Bigr)^k\\
&\quad= \frac{1}{1-\delta}\cdot\exp\Bigl\{\frac{u}{1-\delta}\Bigr\}.
\end{aligned}$$

□

As a consequence of the last two lemmas we get

$$\sum_{n=0}^{\infty}\phi_n(u) < \infty. \tag{7.92}$$

It follows by the next Borel-Cantelli argument that for almost every ω, the cadlag functions $\{X_n^{\nu_k-}(t) : n \in \mathbf{Z}_+\}$ on the interval $t \in [0, \gamma(u)]$ form a Cauchy sequence in the uniform norm. (Recall that we are still holding k fixed.) Pick $\alpha \in (c, 1)$. By Chebychev's inequality and (7.90),

$$\begin{aligned} \sum_{n=0}^{\infty} P\Big\{ \sup_{0 \le t \le \gamma(u)} |X_{n+1}^{\nu_k-}(t) - X_n^{\nu_k-}(t)| \ge \alpha^{n/2} \Big\} &\le B \sum_{n=0}^{\infty} \alpha^{-n} \phi_n(u) \\ &\le B \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{k!} \binom{n}{k} \Big(\frac{u}{\alpha}\Big)^k \Big(\frac{c}{\alpha}\Big)^{n-k}. \end{aligned}$$

This sum converges by Lemma 7.38. By the Borel-Cantelli lemma there exists an almost surely finite $N(\omega)$ such that for $n \ge N(\omega)$,

$$\sup_{0 \le t \le \gamma(u)} |X_{n+1}^{\nu_k-}(t) - X_n^{\nu_k-}(t)| < \alpha^{n/2}.$$

Consequently for $p > m \ge N(\omega)$,

$$\begin{aligned} \sup_{0 \le t \le \gamma(u)} |X_m^{\nu_k-}(t) - X_p^{\nu_k-}(t)| &\le \sum_{n=m}^{p-1} \sup_{0 \le t \le \gamma(u)} |X_{n+1}^{\nu_k-}(t) - X_n^{\nu_k-}(t)| \\ &\le \sum_{n=m}^{\infty} \alpha^{n/2} = \frac{\alpha^{m/2}}{1 - \alpha^{1/2}}. \end{aligned}$$

This last bound can be made arbitrarily small by taking m large enough. This gives the Cauchy property. By the completeness of cadlag functions under uniform distance (Lemma A.5), for almost every ω there exists a cadlag function $t \mapsto \widetilde{X}_k(t)$ on the interval $[0, \gamma(u)]$ such that

$$\lim_{n\to\infty} \sup_{0 \le t \le \gamma(u)} |\widetilde{X}_k(t) - X_n^{\nu_k-}(t)| = 0. \tag{7.93}$$

By considering a sequence of u-values increasing to infinity, we get a single event of probability one on which (7.93) holds for all $0 \le u < \infty$. For any fixed ω and $T < \infty$, $\gamma(u) \ge T$ for all large enough u. We conclude that, with probability one, there exists a cadlag function $\widetilde{X}_k$ on the interval $[0, \infty)$ such that

$$\lim_{n\to\infty} \sup_{0 \le t \le T} |\widetilde{X}_k(t) - X_n^{\nu_k-}(t)| = 0 \quad \text{for all } T < \infty. \tag{7.94}$$

Next we show, still with k fixed, that $\widetilde{X}_k$ is a solution of

$$\widetilde{X}_k(t) = H^{\nu_k-}(t) + \int_{(0,t]} F(s, \widetilde{X}_k)\, dY^{\nu_k-}(s). \tag{7.95}$$

For this we let $n \to \infty$ in (7.88) to obtain (7.95) in the limit. The left side of (7.88) converges to the left side of (7.95) almost surely, uniformly on compact time intervals by (7.94). For the the right side of (7.88) we apply Theorem 5.44. From the Lipschitz property of F

$$\big|F(t,\widetilde{X}_k) - F(t,X_n^{\nu_k-})\big| \le L\cdot \sup_{0\le s<t} |\widetilde{X}_k(s) - X_n^{\nu_k-}(s)|.$$

In Theorem 5.44 take $H_n(t) = F(t,\widetilde{X}_k) - F(t,X_n^{\nu_k-})$ and the cadlag bound

$$G_n(t) = L\cdot \sup_{0\le s\le t} |\widetilde{X}_k(s) - X_n^{\nu_k-}(s)|.$$

(Note the range of the supremum over s.) The convergence in (7.94) gives the hypothesis in Theorem 5.44, and we conclude that the right side of (7.88) converges to the right side of (7.95), in probability, uniformly over t in compact time intervals. We can get almost sure convergence along a subsequence. Equation (7.95) has been verified.

This can be repeated for all values of k. The limit (7.94) implies that if $k<m$, then $\widetilde{X}_m = \widetilde{X}_k$ on $[0,\nu_k)$. Since $\nu_k \nearrow \infty$, we conclude that there is a single well-defined cadlag process X on $[0,\infty)$ such that $X=\widetilde{X}_k$ on $[0,\nu_k)$.

On $[0,\nu_k)$ equation (7.95) agrees term by term with the equation

$$X(t) = H(t) + \int_{(0,t]} F(s,X)\,dY(s). \tag{7.96}$$

For the integral term this follows from part (b) of Proposition 5.50 and the manner of dependence of F on the path: since $X=\widetilde{X}_k$ on $[0,\nu_k)$, $F(s,X) = F(s,\widetilde{X}_k)$ on $[0,\nu_k]$.

We have found a solution to equation (7.96). The proof of Proposition 7.35 is thereby complete.

The main result of this section is the existence of a solution without extra assumptions on Y. This theorem will also complete the proof of Theorem 7.17.

Theorem 7.39. *A solution to (7.45) on $[0,\infty)$ exists under Assumption 7.16 on F, for an arbitrary semimartingale Y and cadlag process H.*

Given an arbitrary semimartingale Y and a cadlag process H, fix constants $0<\delta<K/3<\infty$ so that c defined by (7.61) satisfies $c<1$. Pick a decomposition $Y=Y_0+M+G$ such that the local L^2-martingale M has jumps bounded by $\delta/2$. Fix a constant $0<C<\infty$. Define the following stopping times ρ_k, σ_k and τ_k^i for $1\le i\le 3$ and $k\in \mathbf{Z}_+$. First

$\rho_0 = \sigma_0 = \tau_0^i = 0$ for $1 \le i \le 3$. For $k \ge 1$,

$$\begin{aligned}
\tau_k^1 &= \inf\{t > 0 : |Y(\rho_{k-1} + t) - Y(\rho_{k-1})| \ge C \\
&\qquad\qquad\qquad\qquad \text{or } |Y((\rho_{k-1} + t)-) - Y(\rho_{k-1})| \ge C\}, \\
\tau_k^2 &= \inf\{t > 0 : |\Delta Y(\rho_{k-1} + t)| > \delta/2\}, \\
\tau_k^3 &= \inf\{t > 0 : V_{G_j}(\rho_{k-1} + t) - V_{G_j}(\rho_{k-1}) \ge K - 2\delta \\
&\qquad\qquad\qquad\qquad \text{for some } 1 \le j \le m\}, \\
\sigma_k &= \tau_k^1 \wedge \tau_k^2 \wedge \tau_k^3 \wedge 1, \quad \text{and} \quad \rho_k = \sigma_1 + \cdots + \sigma_k.
\end{aligned}$$

Each $\sigma_k > 0$ for the same reasons that σ defined by (7.82) was positive. Consequently $0 = \rho_0 < \rho_1 < \rho_2 < \cdots$

We claim that $\rho_k \nearrow \infty$. To see why, suppose to the contrary that $\rho_k(\omega) \nearrow u < \infty$ for some sample point ω. By the existence of the limit $Y(u-)$ there exists $\beta > 0$ such that $|Y(s) - Y(t)| \le \delta/4$ for $s, t \in [u-\beta, u)$. But then, once $\rho_k \in [u - \beta/2, u)$, it must be that $\rho_k + \tau_{k+1}^i \ge u$ for both $i = 1, 2$. (Tacitly assuming here that $C > \delta/4$.) since V_{G_j} is nondecreasing, $\tau_{k_m}^3 \nearrow u$ would force $V_{G_j}(u-) = \infty$.

By iterating part (a) of Lemma 7.28 one can conclude that for each k, ρ_k is a stopping time for $\{\mathcal{F}_t\}$, and then σ_{k+1} is a stopping time for $\{\mathcal{F}_{\rho_k + t} : t \ge 0\}$. (Recall that we are assuming $\{\mathcal{F}_t\}$ right-continuous now.)

The heart of the existence proof is an iteration which we formulate with the next lemma.

Lemma 7.40. *For each k, there exists an adapted cadlag process $X_k(t)$ such that the equation*

$$X_k(t) = H^{\rho_k}(t) + \int_{(0,t]} F(s, X_k)\, dY^{\rho_k}(s) \tag{7.97}$$

is satisfied.

Proof. The proof is by induction. After each ρ_k, we restart the equation but stop before the hypotheses of Proposition 7.35 are violated. This way we can apply Proposition 7.35 to construct a segment of the solution, one interval (ρ_k, ρ_{k+1}) at a time. An explicit definition takes the solution up to time ρ_{k+1}, and then we are ready for the next iteration.

For $k = 1$, $\rho_1 = \sigma_1$. The semimartingale $Y^{\sigma_1 -}$ satisfies the hypotheses of Proposition 7.35. This argument is the same as in the uniqueness proof of the previous section, at (7.83). Consequently by Proposition 7.35 there exists a solution $\widetilde{X}$ of the equation

$$\widetilde{X}(t) = H^{\rho_1 -}(t) + \int_{(0,t]} F(s, \widetilde{X})\, dY^{\rho_1 -}(s). \tag{7.98}$$

Define

$$X_1(t) = \begin{cases} \widetilde{X}(t), & 0 \le t < \rho_1 \\ X_1(\rho_1-) + \Delta H(\rho_1) + F(\rho_1, \widetilde{X})\Delta Y(\rho_1), & t \ge \rho_1. \end{cases}$$

$F(t, \widetilde{X}) = F(t, X_1)$ for $0 \le t \le \rho_1$ because $X_1 = \widetilde{X}$ on $[0, \rho_1)$. Then X_1 solves (7.97) for $k = 1$ and $0 \le t < \rho_1$. For $t \ge \rho_1$, recalling (5.46)–(5.47) for left limits and jumps of stochastic integrals:

$$\begin{aligned} X_1(t) &= X_1(\rho_1-) + \Delta H(\rho_1) + F(\rho_1, \widetilde{X})\Delta Y(\rho_1) \\ &= H(\rho_1) + \int_{(0,\rho_1]} F(s, X_1)\, dY(s) \\ &= H^{\rho_1}(t) + \int_{(0,t]} F(s, X_1)\, dY^{\rho_1}(s). \end{aligned}$$

The equality of stochastic integrals of the last two lines above is an instance of the general identity $G \cdot Y^\tau = (G \cdot Y)^\tau$ (Proposition 5.39). The case $k = 1$ of the lemma has been proved.

Assume a process $X_k(t)$ solves (7.97). Define $\bar{\mathcal{F}}_t = \mathcal{F}_{\rho_k + t}$,

$$\begin{aligned} \bar{H}(t) &= H(\rho_k + t) - H(\rho_k), \\ \bar{Y}(t) &= Y(\rho_k + t) - Y(\rho_k), \\ \text{and} \quad \bar{F}(t, \omega, \eta) &= F(\rho_k + t, \omega, \zeta^{\omega,\eta}) \end{aligned}$$

where the cadlag path $\zeta^{\omega,\eta}$ is defined by

$$\zeta^{\omega,\eta}(s) = \begin{cases} X_k(s), & 0 \le s < \rho_k \\ X_k(\rho_k) + \eta(s - \rho_k), & s \ge \rho_k. \end{cases}$$

Now we find a solution $\bar{X}$ to the equation

$$\bar{X}(t) = \bar{H}^{\sigma_{k+1}-}(t) + \int_{(0,t]} \bar{F}(s, \bar{X})\, d\bar{Y}^{\sigma_{k+1}-}(s) \tag{7.99}$$

under the filtration $\{\bar{\mathcal{F}}_t\}$. We need to check that this equation is the type to which Proposition 7.35 applies. Semimartingale $\bar{Y}^{\sigma_{k+1}-}$ satisfies the assumption of Proposition 7.35, again by the argument already used in the uniqueness proof. $\bar{F}$ satisfies Assumption 7.16 exactly as was proved earlier for Lemma 7.33.

The hypotheses of Proposition 7.35 have been checked, and so there exists a process $\bar{X}$ that solves (7.99). Note that $\bar{X}(0) = \bar{H}(0) = 0$. Define

$$X_{k+1}(t) = \begin{cases} X_k(t), & t < \rho_k \\ X_k(\rho_k) + \bar{X}(t - \rho_k), & \rho_k \le t < \rho_{k+1} \\ X_{k+1}(\rho_{k+1}-) + \Delta H(\rho_{k+1}) & \\ \qquad + F(\rho_{k+1}, X_{k+1})\Delta Y(\rho_{k+1}), & t \ge \rho_{k+1}. \end{cases}$$

The last case of the definition above makes sense because it depends on the segment $\{X_{k+1}(s) : 0 \le s < \rho_{k+1}\}$ defined by the two preceding cases.

By induction X_{k+1} satisfies the equation (7.97) for $k+1$ on $[0, \rho_k]$. From the definition of $\bar{F}$, $\bar{F}(s, \bar{X}) = F(\rho_k + s, X_{k+1})$ for $0 \le s \le \sigma_{k+1}$. Then for $\rho_k < t < \rho_{k+1}$,

$$\begin{aligned} X_{k+1}(t) &= X_k(\rho_k) + \bar{X}(t - \rho_k) \\ &= X_k(\rho_k) + \bar{H}(t - \rho_k) + \int_{(0, t-\rho_k]} \bar{F}(s, \bar{X})\, d\bar{Y}(s) \\ &= X_k(\rho_k) + H(t) - H(\rho_k) + \int_{(\rho_k, t]} F(s, X_{k+1})\, dY(s) \\ &= H(t) + \int_{(0,t]} F(s, X_{k+1})\, dY(s) \\ &= H^{\rho_{k+1}}(t) + \int_{(0,t]} F(s, X_{k+1})\, dY^{\rho_{k+1}}(s). \end{aligned}$$

The last line of the definition of X_{k+1} extends the validity of the equation to $t \ge \rho_{k+1}$:

$$\begin{aligned} X_{k+1}(t) &= X_{k+1}(\rho_{k+1}-) + \Delta H(\rho_{k+1}) + F(\rho_{k+1}, X_{k+1})\Delta Y(\rho_{k+1}) \\ &= H(\rho_{k+1}-) + \Delta H(\rho_{k+1}) \\ &\qquad + \int_{(0,\rho_{k+1})} F(s, X_{k+1})\, dY(s) + F(\rho_{k+1}, X_{k+1})\Delta Y(\rho_{k+1}) \\ &= H(\rho_{k+1}) + \int_{(0,\rho_{k+1}]} F(s, X_{k+1})\, dY(s) \\ &= H^{\rho_{k+1}}(t) + \int_{(0,t]} F(s, X_{k+1})\, dY^{\rho_{k+1}}(s). \end{aligned}$$

By induction, the lemma has been proved. □

We are ready to finish off the proof of Theorem 7.39. If $k < m$, stopping the processes of the equation

$$X_m(t) = H^{\rho_m}(t) + \int_{(0,t]} F(s, X_m)\, dY^{\rho_m}(s)$$

at ρ_k gives the equation

$$X_m^{\rho_k}(t) = H^{\rho_k}(t) + \int_{(0,t]} F(s, X_m^{\rho_k})\, dY^{\rho_k}(s).$$

By the uniqueness theorem, $X_m^{\rho_k} = X_k$ for $k < m$. Consequently there exists a process X that satisfies $X = X_k$ on $[0, \rho_k]$ for each k. Then for $0 \le t \le \rho_k$, equation (7.97) agrees term by term with the desired equation

$$X(t) = H(t) + \int_{(0,t]} F(s, X)\, dY(s).$$

Hence this equation is valid on every $[0, \rho_k]$, and thereby on $[0, \infty)$. The existence and uniqueness theorem has been proved.

Exercises

Exercise 7.1. (a) Show that for any $g \in C[0,1]$,

$$\lim_{t \nearrow 1} (1-t) \int_0^t \frac{g(s)}{(1-s)^2}\, ds = g(1).$$

(b) Let the process X_t be defined by (7.13) for $0 \le t < 1$. Show that $X_t \to 0$ as $t \to 1$.

Hint. Apply Exercise 6.8 and then part (a).

Exercise 7.2. Assume $a(t)$, $b(t)$, $g(t)$ and $h(t)$ are given deterministic Borel functions on $\mathbf{R}_+$ that are bounded on each compact time interval. Solve the following linear SDE with a suitable integrating factor, and use Itô's formula to check that your tentative solution solves the equation:

$$dX_t = \big(a(t)X_t + b(t)\big)\, dt + \big(g(t)X_t + h(t)\big)\, dB_t. \tag{7.100}$$

Hints. To guess at the right integrating factor, rewrite the equation as

$$dX - X(a\, dt + g\, dB) = b\, dt + h\, dB$$

and look at the examples in Section 7.1.1 for inspiration. You can also solve an easier problem first by setting some of the coefficient functions equal to constants or even zero.

Exercise 7.3. Generalize the integrating factor approach further to solve

$$dX_t = a(t)X_t\, d[Y]_t + b(t)X_t\, dY_t \tag{7.101}$$

where Y is a given continuous semimartingale and functions $a(t)$ and $b(t)$ are locally bounded Borel functions.

Exercise 7.4. Show that the proof of Theorem 7.12 continues to work if the Lipschitz assumption on σ and b is weakened to this local Lipschitz condition: for each $n \in \mathbf{N}$ there exists a constant $L_n < \infty$ such that

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \le L_n |x - y| \tag{7.102}$$

for all $t \in \mathbf{R}_+$ and $x, y \in \mathbf{R}^d$ such that $|x|, |y| \le n$.

Exercise 7.5. Let a, b, ρ, σ be locally Lipschitz functions on $\mathbf{R}^d$, G an open subset of $\mathbf{R}^d$, and assume that on G we have the equalities $a = b$ and $\rho = \sigma$. Let $\xi \in G$ and let X and Y be continuous $\mathbf{R}^d$-valued processes that satisfy

$$X_t = \xi + \int_0^t a(X_s)\, ds + \int_0^t \rho(X_s)\, dB_s$$

and

$$Y_t = \xi + \int_0^t b(Y_s)\, ds + \int_0^t \sigma(Y_s)\, dB_s.$$

Let

$$\alpha = \inf\{t \ge 0 : X_t \in G^c\} \quad \text{and} \quad \beta = \inf\{t \ge 0 : Y_t \in G^c\}.$$

Show that the processes X^α and Y^β are indistinguishable. *Hint.* Use Gronwall's inequality on a second moment, along the lines of the proof of Theorem 7.12. Note that if G is bounded, then the functions a, b, ρ, σ are bounded on G.

Exercise 7.6. (a) Let Y be a cadlag semimartingale. Find $[[Y]]$ (the quadratic variation of the quadratic variation) and the covariation $[Y, [Y]]$. For use in part (b) you should find the simplest possible formulas in terms of the jumps of Y.

(b) Let Y be a continuous semimartingale. Show that

$$X_t = X_0 \exp(\alpha t + \beta Y_t - \tfrac{1}{2}\beta^2 [Y]_t)$$

solves the SDE

$$dX = \alpha X\, dt + \beta X\, dY.$$

Exercise 7.7 (Fisk-Stratonovich SDE)**.** Recall the definition of the Fisk-Stratonovich integral from Exercise 6.14. Let $B_{\boldsymbol{\cdot}}$ be standard Brownian motion, b a continuous function on $\mathbf{R}$ and $\sigma \in C^2(\mathbf{R})$. Assume that X is a continuous semimartingale. Show that these two integral equations are equivalent:

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s) \circ dB_s \tag{7.103}$$

and

$$X_t = X_0 + \int_0^t \bigl(b(X_s) + \tfrac{1}{2}\sigma(X_s)\sigma'(X_s)\bigr)\, ds + \int_0^t \sigma(X_s)\, dB_s. \tag{7.104}$$

In other words, a Fisk-Stratonovich SDE corresponds to adding another drift to an Itô equation.

Exercise 7.8. Here is an alternative inductive proof of the identity (7.91) used in the existence proof for solutions of SDE's. Fix $-1 < x < 1$ and let

$$a_k = \sum_{m=0}^{\infty} (m+1)(m+2)\cdots(m+k)x^m$$

and

$$b_k = \sum_{m=0}^{\infty} m(m+1)(m+2)\cdots(m+k)x^m.$$

Compute a_1 explicitly, then derive the identities $b_k = xa_{k+1}$ and $a_{k+1} = (k+1)a_k + b_k$.

Chapter 8

Applications of Stochastic Calculus

8.1. Local time

In Section 8.1.1 we construct local time for Brownian motion and then in Section 8.1.2 we derive Tanaka's formula and the distribution of local time.

8.1.1. Existence of local time for Brownian motion. $(\Omega, \mathcal{F}, P)$ is a fixed complete probability space with a complete filtration $\{\mathcal{F}_t\}$, and $B = \{B_t\}$ a one-dimensional Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$ (Definition 2.26). By redefining B on an event of probability zero if necessary we can assume that $B_t(\omega)$ is continuous in t for each $\omega \in \Omega$. The initial point B_0 is an arbitrary, possibly random point in $\mathbf{R}$.

We construct the *local time process* $L(t, x)$ of Brownian motion. This process measures the time spent at point x up to time t. Note that it does not make sense to look "literally" at the amount of time spent at x up to time t: this would be the random variable $\int_0^t \mathbf{1}\{B_s = x\}\, ds$ which vanishes almost surely (Exercise 8.1). The proper thing to look for is a Radon-Nikodym derivative with respect to Lebesgue measure. This idea is captured by the requirement that

$$\int_0^t g(B_s(\omega))\, ds = \int_{\mathbf{R}} g(x) L(t, x, \omega)\, dx \tag{8.1}$$

for all bounded Borel functions g. Here is the existence theorem that summarizes the main properties.

Theorem 8.1. *There exists a process* $\{L(t,x,\omega) : t \in \mathbf{R}_+,\ x \in \mathbf{R}\}$ *on* $(\Omega, \mathcal{F}, P)$ *and an event* Ω_0 *such that* $P(\Omega_0) = 1$ *with the following properties.* $L(t,x)$ *is* $\mathcal{F}_t$*-measurable for each* (t,x). *For each* $\omega \in \Omega_0$ *the following statements hold.*

(a) $L(t,x,\omega)$ *is jointly continuous in* (t,x) *and nondecreasing in* t.

(b) *Identity* (8.1) *holds for all* $t \in \mathbf{R}_+$ *and bounded Borel functions* g *on* $\mathbf{R}$.

(c) *For each fixed* $x \in \mathbf{R}$,

$$\int_0^\infty \mathbf{1}\{B_t(\omega) \neq x\}\, L(dt, x, \omega) = 0. \tag{8.2}$$

The integral in (8.2) is the Lebesgue-Stieltjes integral over time whose integrator is the nondecreasing function $t \mapsto L(t,x,\omega)$. The continuity of $L(t,x)$ and (8.1) imply a pointwise characterization of local time as

$$L(t,x,\omega) = \lim_{\varepsilon \searrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbf{1}_{(x-\varepsilon, x+\varepsilon)}(B_s(\omega))\, ds \quad \text{for } \omega \in \Omega_0. \tag{8.3}$$

We do not base the proof of Theorem 8.1 on a direct construction such as (8.3), intuitively attractive as it is. But note that (8.3) does imply the claim about the monotonicity of $L(t,x,\omega)$ as a function of t.

We shall first give a precise definition of $L(t,x)$, give some intuitive justification for it, and then prove Theorem 8.1. We use stochastic integrals of the type $\int_0^t \mathbf{1}_{[x,\infty)}(B_s)\, dB_s$. To justify their well-definedness note that the integrand $\mathbf{1}_{[x,\infty)}(B_s(\omega))$ is predictable, by Exercise 5.2(b). Alternately, it is enough to observe that the integrand is a measurable process and appeal to the fact that measurability is enough for integration with Brownian motion (Chapter 4 or Section 5.5).

To define $L(t,x)$ we need a nice version of these integrals. We postpone the proof of this lemma to the end of the section.

Lemma 8.2. *There exists a process* $\{I(t,x) : t \in \mathbf{R}_+,\ x \in \mathbf{R}\}$ *such that* $(t,x) \mapsto I(t,x,\omega)$ *is continuous for all* $\omega \in \Omega$ *and for each* (t,x),

$$I(t,x) = \int_0^t \mathbf{1}_{[x,\infty)}(B_s)\, dB_s \quad \textit{with probability 1.} \tag{8.4}$$

Define the process $L(t,x)$ by the equation

$$\tfrac{1}{2} L(t,x,\omega) = (B_t(\omega) - x)^+ - (B_0(\omega) - x)^+ - I(t,x,\omega). \tag{8.5}$$

$L(t,x,\omega)$ is continuous in (t,x) for each $\omega \in \Omega$ because the terms on the right are continuous.

Heuristic justification. Let us explain informally the idea behind this definition. The limit in (8.3) captures the idea of local time as a derivative with respect to Lebesgue measure. But Brownian paths are rough so we prefer to take the $\varepsilon \searrow 0$ limit in a situation with more regularity. Itô's formula gives us a way out. Among the terms that come out of applying Itô's formula to a function φ is the integral $\frac{1}{2}\int_0^t \varphi''(B_s)\,ds$. So fix a small $\varepsilon > 0$ and take $\varphi''(z) = (2\varepsilon)^{-1}\mathbf{1}_{(x-\varepsilon,x+\varepsilon)}(z)$ so that Itô's formula captures the integral on the right of (8.3). Two integrations show that φ should satisfy

$$\varphi'(z) = \begin{cases} 0, & z < x-\varepsilon \\ (2\varepsilon)^{-1}(z-x+\varepsilon), & x-\varepsilon \le z \le x+\varepsilon \\ 1, & z > x+\varepsilon \end{cases}$$

and

$$\varphi(z) = \begin{cases} 0, & z < x-\varepsilon \\ (4\varepsilon)^{-1}(z-x+\varepsilon)^2, & x-\varepsilon \le z \le x+\varepsilon \\ z-x, & z > x+\varepsilon. \end{cases}$$

However, φ'' is not continuous so we cannot apply Itô's formula to φ. But *if* we could, the outcome would be

$$\text{(8.6)} \qquad \frac{1}{4\varepsilon}\int_0^t \mathbf{1}_{(x-\varepsilon,x+\varepsilon)}(B_s(\omega))\,ds = \varphi(B_t) - \varphi(B_0) - \int_0^t \varphi'(B)\,dB.$$

As $\varepsilon \searrow 0$, $\varphi(z) \to (z-x)^+$ and $\varphi'(z) \to \mathbf{1}_{[x,\infty)}(z)$, except at $z = x$, but this single point would not make a difference to the stochastic integrals. The upshot is that definition (8.5) represents the unjustified $\varepsilon \searrow 0$ limit of the unjustified equation (8.6).

Returning to the rigorous proof, next we show that $L(t,x)$ functions as an occupation density.

Proposition 8.3. *There exists an event Ω_0 such that $P(\Omega_0) = 1$ and for all $\omega \in \Omega_0$ the process $L(t,x,\omega)$ defined by (8.5) satisfies (8.1) for all $t \in \mathbf{R}_+$ and all bounded Borel functions g on $\mathbf{R}$.*

Proof. We show that (8.1) holds for a special class of functions. The class we choose comes from the proof given in [11]. The rest of the proof is the usual clean-up.

Step 1. We verify that (8.1) holds almost surely for a fixed t and a fixed piecewise linear continuous function $g = g_{q,r,\eta}$ of the following type:

$$g(x) = \begin{cases} 0, & x \le q-\eta \text{ or } x \ge r+\eta \\ (x-q+\eta)/\eta, & q-\eta \le x \le q \\ 1, & q \le x \le r \\ (r+\eta-x)/\eta, & r \le x \le r+\eta \end{cases} \tag{8.7}$$

where $q < r$ and $\eta > 0$ are rationals. As η becomes small, g approximates the indicator function of the closed interval $[q, r]$. Abbreviate $a = q - \eta$ and $b = r + \eta$ so that g vanishes outside $[a, b]$.

We need a Fubini theorem of sorts for stochastic integrals. We sketch the proof of this lemma after the main proof.

Lemma 8.4. *The equality*

$$\int_0^t \biggl(\int_a^b g(x) \mathbf{1}_{[x,\infty)}(B_s)\, dx \biggr) dB_s = \int_a^b g(x)\, I(t,x)\, dx \tag{8.8}$$

holds almost surely.

The trick is now to write down a version of Itô's formula where the integral $\int_0^t g(B_s)\, ds$ appears as the quadratic variation part. To this end, define

$$f(z) = \int_{-\infty}^z dy \int_{-\infty}^y dx\, g(x) = \int_a^b g(x)(z-x)^+\, dx$$

and note that

$$f'(z) = \int_{-\infty}^z g(x)\, dx = \int_a^b g(x) \mathbf{1}_{[x,\infty)}(z)\, dx \quad \text{and} \quad f''(z) = g(z).$$

By Itô's formula

$$\tfrac{1}{2} \int_0^t g(B_s)\, ds = f(B_t) - f(B_0) - \int_0^t f'(B_s)\, dB_s$$

$$= \int_a^b g(x) \bigl[(B_t - x)^+ - (B_0 - x)^+ \bigr] dx - \int_0^t \biggl(\int_a^b g(x) \mathbf{1}_{[x,\infty)}(B_s)\, dx \biggr) dB_s$$

[by (8.8)]

$$= \int_a^b g(x) \bigl[(B_t - x)^+ - (B_0 - x)^+ - I(t,x) \bigr] dx$$

[by (8.5)]

$$= \tfrac{1}{2} \int_a^b g(x) L(t,x)\, dx.$$

Comparing the first and last expressions of the calculation verifies (8.1) almost surely for a particular g of type (8.7) and a fixed t.

Step 2. Let Ω_0 be the event of full probability on which (8.1) holds for all rational $t \geq 0$ and the countably many $g_{q,r,\eta}$ for rational (q, r, η). For a fixed $g_{q,r,\eta}$ both sides of (8.1) are continuous in t. (The right side because $L(t, x, \omega)$ is uniformly continuous for (t, x) in a fixed rectangle which can be taken large enough to contain the support of g.) Consequently on Ω_0 (8.1) extends from rational t to all $t \in \mathbf{R}_+$.

At this point we can see that $L(t, x, \omega) \geq 0$. If $L(t, x, \omega) < 0$ then by continuity there exists $\delta > 0$ and an interval (a, b) around x such that $L(t, y, \omega) \leq -\delta$ for $y \in (a, b)$. But for g of type (8.7) supported inside (a, b) we have

$$\int_{\mathbf{R}} g(x) L(t, x, \omega)\, dx = \int_0^t g(B_s(\omega))\, ds \geq 0,$$

a contradiction.

To extend (8.1) to all functions with $\omega \in \Omega_0$ fixed, think of both sides of (8.1) as defining a measure on $\mathbf{R}$. By letting $q \searrow -\infty$ and $r \nearrow \infty$ in $g_{q,r,\eta}$ we see that both sides define finite measures with total mass t. By letting $\eta \searrow 0$ we see that the two measures agree on closed intervals $[q, r]$ with rational endpoints. These intervals form a π-system that generates $\mathcal{B}_{\mathbf{R}}$. By Lemma B.5 the two measures agree on all Borel sets, and consequently integrals of bounded Borel functions agree also. □

It remains to prove claim (8.2). By Exercise 1.3 it suffices to show that if t is a point of strict increase of $L(\,\cdot\,, x, \omega)$ with $x \in \mathbf{R}$ and $\omega \in \Omega_0$ fixed, then $B_t(\omega) = x$. So suppose t is such a point, so that in particular $L(t, x, \omega) < L(t_1, x, \omega)$ for all $t_1 > t$. It follows from (8.3) that for all small enough $\varepsilon > 0$,

$$\int_0^t \mathbf{1}_{(x-\varepsilon, x+\varepsilon)}(B_s(\omega))\, ds < \int_0^{t_1} \mathbf{1}_{(x-\varepsilon, x+\varepsilon)}(B_s(\omega))\, ds$$

which would not be possible unless there exists $s \in (t, t_1)$ such that $B_s(\omega) \in (x - \varepsilon, x + \varepsilon)$. Take a sequence $\varepsilon_j \searrow 0$ and for each ε_j pick $s_j \in (t, t_1)$ such that $B_{s_j}(\omega) \in (x - \varepsilon_j, x + \varepsilon_j)$. By compactness there is a convergent subsequence $s_{j_k} \to s \in [t, t_1]$. By path-continuity $B_{s_{j_k}}(\omega) \to B_s(\omega)$ and by choice of s_j's, $B_{s_{j_k}}(\omega) \to x$. We conclude that for each $t_1 > t$ there exists $s \in [t, t_1]$ such that $B_s(\omega) = x$. Taking $t_1 \searrow t$ gives $B_t(\omega) = x$.

We have now proved all the claims of Theorem 8.1. It remains to prove the two lemmas used along the way.

Proof of Lemma 8.4. Recall that we are proving the almost sure identity

$$\int_0^t \biggl(\int_a^b g(x) \mathbf{1}_{[x,\infty)}(B_s)\, dx \biggr)\, dB_s = \int_a^b g(x)\, I(t,x)\, dx \tag{8.9}$$

where g is given by (8.7), $a = q - \eta$ and $b = r + \eta$, and g vanishes outside $[a,b]$.

The left-hand stochastic integral is well-defined because the integrand is a bounded continuous process. The continuity can be seen from

$$\int_a^b g(x) \mathbf{1}_{[x,\infty)}(B_s)\, dx = \int_{-\infty}^{B_s} g(x)\, dx. \tag{8.10}$$

Since $g(x)I(t,x)$ is continuous the integral on the right of (8.9) can be written as a limit of Riemann sums. Introduce partitions $x_i = a + (i/n)(b - a)$, $0 \le i \le n$, with mesh $\delta = (b-a)/n$. Stochastic integrals are linear, so on the right of (8.9) we have almost surely

$$\begin{aligned} \lim_{n\to\infty} \delta \sum_{i=1}^n g(x_i) I(t,x_i) &= \lim_{n\to\infty} \delta \sum_{i=1}^n g(x_i) \int_0^t \mathbf{1}_{[x_i,\infty)}(B_s)\, dB_s \\ &= \lim_{n\to\infty} \int_0^t \delta \sum_{i=1}^n g(x_i) \mathbf{1}_{[x_i,\infty)}(B_s)\, dB_s. \end{aligned}$$

By the L^2 isometry of stochastic integration we can assert that this limit equals the stochastic integral on the left of (8.9) if

$$\lim_{n\to\infty} \int_0^t E\biggl[\biggl(\int_a^b g(x) \mathbf{1}_{[x,\infty)}(B_s)\, dx \;-\; \delta \sum_{i=1}^n g(x_i)\, \mathbf{1}_{[x_i,\infty)}(B_s) \biggr)^2 \biggr]\, ds = 0.$$

We leave this as an exercise. □

Proof of Lemma 8.2. Abbreviate $M(t,x) = \int_0^t \mathbf{1}_{[x,\infty)}(B_s)\, dB_s$. Recall that the task is to establish the existence of a continuous version for $M(t,x)$, as a function of $(t,x) \in \mathbf{R}_+ \times \mathbf{R}$. We will show that for any $T < \infty$ and $m \in \mathbf{N}$ there exists a finite constant $C = C(T,m)$ so that

$$E\bigl[\, |M(s,x) - M(t,y)|^{2m} \bigr] \le C |(s,x) - (t,y)|^m \tag{8.11}$$

for $(s,x), (t,y) \in [0,T] \times \mathbf{R}$. After (8.11) has been shown the lemma follows from the Kolmogorov-Centsov criterion (Theorem B.20). Precisely speaking, here is what needs to be said to derive the lemma from Theorem B.20.

The index set has dimension two. So if we take $m > 2$, Theorem B.20 implies that for each positive integer k the process $M(t,x)$ has a continuous version $\{Y^{(k)}(t,x) : (t,x) \in [0,k] \times [-k,k]\}$ on the compact index set $[0,k] \times [-k,k]$. The processes $Y^{(k)}$ are then combined to yield a single continuous process indexed by $\mathbf{R}_+ \times \mathbf{R}$ via the following reasoning. If $k < \ell$, then for rational $(t,x) \in [0,k] \times [-k,k]$ the equality $Y^{(k)}(t,x) = M(t,x) = Y^{(\ell)}(t,x)$

holds with probability 1. Since rationals are countable, there is a single event $\Omega_{k,\ell}$ such that $P(\Omega_{k,\ell}) = 1$ and for $\omega \in \Omega_{k,\ell}$, $Y^{(k)}(t,x,\omega) = Y^{(\ell)}(t,x,\omega)$ for all rational $(t,x) \in [0,k] \times [-k,k]$. All values of a continuous function are completely determined by the values on a dense set, hence for $\omega \in \Omega_{k,\ell}$, $Y^{(k)}(t,x,\omega) = Y^{(\ell)}(t,x,\omega)$ for all $(t,x) \in [0,k] \times [-k,k]$. Let $\Omega_0 = \bigcap_{k<\ell} \Omega_{k,\ell}$. Then $P(\Omega_0) = 1$ and for $\omega \in \Omega_0$ we can consistently define $I(t,x,\omega) = Y^{(k)}(t,x,\omega)$ for any k such that $(t,x) \in [0,k] \times [-k,k]$. Outside Ω_0 we can define for example $I(t,x,\omega) \equiv 0$ to get a process that is continuous in (t,x) at all ω.

We turn to prove (8.11). We can assume $s < t$. First using additivity of stochastic integrals and abbreviating $a = x \wedge y$, $b = x \vee y$, almost surely

$$M(s,x) - M(t,y) = \operatorname{sign}(y-x) \int_0^s \mathbf{1}_{[a,b)}(B)\,dB - \int_s^t \mathbf{1}_{[y,\infty)}(B)\,dB.$$

Using $|u+v|^k \le 2^k(|u|^k + |v|^k)$ and then the inequality in Proposition 6.16

$$\begin{aligned} &E|M(s,x) - M(t,y)|^{2m} \\ &\le CE\biggl[\,\Bigl|\int_0^s \mathbf{1}_{[a,b)}(B)\,dB\Bigr|^{2m}\biggr] + CE\biggl[\,\Bigl|\int_s^t \mathbf{1}_{[y,\infty)}(B)\,dB\Bigr|^{2m}\biggr] \\ &\le CE\biggl[\Bigl(\int_0^s \mathbf{1}_{[a,b)}(B_u)\,du\Bigr)^m\biggr] + CE\biggl[\Bigl(\int_s^t \mathbf{1}_{[y,\infty)}(B_u)\,du\Bigr)^m\biggr]. \end{aligned} \tag{8.12}$$

The integrals on the last line above appear because the quadratic variation of $\int_0^s X\,dB$ is $\int_0^s X^2\,du$ (Proposition 5.55). For the second integral we need to combine this with the step given in Theorem 5.45:

$$\int_s^t \mathbf{1}_{[y,\infty)}(B_u)\,dB_u = \int_0^{t-s} \mathbf{1}_{[y,\infty)}(B_{s+u})\,d\bar{B}_u$$

where $\bar{B}_u = B_{s+u} - B_s$ is the restarted Brownian motion (which of course is again a standard Brownian motion).

The second integral on line (8.12) is immediately bounded by $|t-s|^m$. We estimate the first integral with the help of independent Brownian increments. Let us abbreviate $d\mathbf{u} = du_m \cdots du_1$ for a multivariate integral and $\Delta u_j = u_j - u_{j-1}$ for increments of the time variables, and introduce $u_0 = 0$ so that $\Delta u_1 = u_1$.

Suppose f is a symmetric function of m variables. This means that permuting the variables does not change the value of f:

$$f(u_1, u_2, \ldots, u_m) = f(u_{i_1}, u_{i_2}, \ldots, u_{i_m})$$

for any rearrangement $(u_{i_1}, u_{i_2}, \dots, u_{i_m})$ of $(u_1, u_2, \dots, u_m)$. Then

$$\int_0^s \cdots \int_0^s f(u_1, u_2, \dots, u_m)\, d\mathbf{u} = m! \int\!\cdots\!\int_{0<u_1<\cdots<u_m<s} f(u_1, u_2, \dots, u_m)\, d\mathbf{u}.$$

Check this as an exercise.

Now for the first integral in (8.12) switch around the expectation and the integration and use the above trick, to make it equal

$$E\int_0^s \cdots \int_0^s \prod_{i=1}^m \mathbf{1}_{[a,b)}(B_{u_i})\, d\mathbf{u}$$

$$= m! \int\!\cdots\!\int_{0<u_1<\cdots<u_m<s} E\big[\mathbf{1}_{[a,b)}(B_{u_1})\mathbf{1}_{[a,b)}(B_{u_2})\cdots\mathbf{1}_{[a,b)}(B_{u_m})\big]\, d\mathbf{u}. \tag{8.13}$$

We estimate the expectation one indicator at a time, beginning with a conditioning step:

$$E\big[\mathbf{1}_{[a,b)}(B_{u_1})\cdots\mathbf{1}_{[a,b)}(B_{u_{m-1}})\,\mathbf{1}_{[a,b)}(B_{u_m})\big]$$
$$= E\big[\mathbf{1}_{[a,b)}(B_{u_1})\cdots\mathbf{1}_{[a,b)}(B_{u_{m-1}})\,E\{\mathbf{1}_{[a,b)}(B_{u_m})\,|\,\mathcal{F}_{u_{m-1}}\}\,\big]$$

To handle the conditional expectation recall that the increment $B_{u_m} - B_{u_{m-1}}$ is independent of $\mathcal{F}_{u_{m-1}}$ and has the $\mathcal{N}(0, \Delta u_m)$ distribution. Use property (x) of conditional expectations given in Theorem 1.26.

$$E\{\mathbf{1}_{[a,b)}(B_{u_m})\,|\,\mathcal{F}_{u_{m-1}}\} = E\{\mathbf{1}_{[a,b)}(B_{u_{m-1}} + B_{u_m} - B_{u_{m-1}})\,|\,\mathcal{F}_{u_{m-1}}\}$$
$$= \int_{-\infty}^{\infty} \mathbf{1}_{[a,b)}(B_{u_{m-1}} + x)\frac{e^{-x^2/(2\Delta u_m)}}{\sqrt{2\pi\Delta u_m}}\, dx$$
$$\le \frac{b-a}{\sqrt{2\pi\Delta u_m}}.$$

In the last upper bound the exponential was simply dropped (it is ≤ 1) and then the integral taken.

This step can be applied repeatedly to the expectation in (8.13) to replace each indicator with the upper bound $(b-a)/\sqrt{2\pi\Delta u_m}$. At the end of this step

$$\text{line (8.13)} \;\le\; (2\pi)^{-m/2} m!(b-a)^m \int\!\cdots\!\int_{0<u_1<\cdots<u_m<s} \frac{1}{\sqrt{\Delta u_1\cdots\Delta u_m}}\, d\mathbf{u}.$$

The integral can be increased by letting s increase to T, then integrated to a constant that depends on T and m, and we get

$$\text{line (8.13)} \;\le\; C(m,T)(b-a)^m.$$

Tracing the development back up we see that this is an upper bound for the first integral on line (8.12).

We now have for line (8.12) the upper bound $|t-s|^m + C(m,T)(b-a)^m$ which is bounded by $C(m,T)|(s,x)-(t,y)|^m$. We have proved inequality (8.11) and thereby the lemma. $\square$

8.1.2. Tanaka's formula and reflection. We begin the next development by extending the defining equation (8.5) of local time to an Itô formula for the convex function defined by the absolute value. The sign function is defined by $\operatorname{sign}(u) = \mathbf{1}\{u>0\} - \mathbf{1}\{u<0\}$.

Theorem 8.5 (Tanaka's formula)**.** *For each $x \in \mathbf{R}$ we have the identity*

$$|B_t - x| - |B_0 - x| = \int_0^t \operatorname{sign}(B_s - x)\, dB_s + L(t,x) \tag{8.14}$$

in the sense that the processes indexed by $t \in \mathbf{R}_+$ on either side of the equality sign are indistinguishable.

Remark 8.6. Equation (8.14) does agree with the usual Itô formula

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\, dB_s + \tfrac{1}{2}\int_0^t f''(B_s)\, ds$$

if we allow ourselves to use weak derivatives f' and f''. Let f be a locally integrable function on $\mathbf{R}$. This means that $\int_K |f|\, dx < \infty$ for any compact set $K \subseteq \mathbf{R}$. Then the measure μ is the *weak derivative* or *distributional derivative* of f (and we write $\mu = f'$) if

$$\int_{\mathbf{R}} \varphi\, d\mu = -\int_{\mathbf{R}} \varphi'(x) f(x)\, dx \tag{8.15}$$

for every $\varphi \in C_c^\infty(\mathbf{R})$ (compactly supported, infinitely differentiable φ). This notion does generalize the classical pointwise derivative because if f' exists and is continuous, then integration parts gives the above identity for $\mu(dx) = f'(x)dx$. Similarly, measure $\nu = f''$ if

$$\int_{\mathbf{R}} \varphi\, d\nu = \int_{\mathbf{R}} \varphi''(x) f(x)\, dx \qquad \forall\, \varphi \in C_c^\infty(\mathbf{R}). \tag{8.16}$$

Note that integration by parts leaves no boundary terms because $\varphi = 0$ outside some bounded interval. Simple integration by parts exercises now verify that if $f(z) = |z-x|$ for a fixed x, then its weak derivatives are $f'(z) = \operatorname{sign}(z-x)$ and $f'' = 2\delta_x$. Here δ_x denotes the pointmass at x, that is, the measure defined for Borel sets A by $\delta_x(A) = \mathbf{1}_A(x)$. The identity that needs to be checked for $f'' = 2\delta_x$ is

$$2\varphi(x) = \int_{\mathbf{R}} |z-x|\, \varphi''(z)\, dz$$

for $\varphi \in C_c^\infty(\mathbf{R})$. Now we see how (8.14) makes sense: the last term can be thought of as $\frac{1}{2}\int_0^t f''(B_s)\,ds$ with $f''(z) = 2\delta_x(z)$ and the rigorous meaning of $\int_0^t \delta_x(B_s)\,ds$ is $L(t,x)$.

This remark is a minute glimpse into the *theory of distributions* where all functions have derivatives in this weak sense. The term *distribution* is used in a sense different from its probabilistic meaning. A synonym four this distribution is the term *generalized function*. See textbooks [8] or [**15**] for more.

Proof of Theorem 8.5. All the processes in (8.14) are continuous in t (the stochastic integral by its construction as a limit in $\mathcal{M}_2^c$), hence it suffices to show almost sure equality for a fixed t.

Let L^- and I^- denote the local time and the stochastic integral in (8.4) associated to the Brownian motion $-B$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ as B. Let Ω_0^- be the full-probability event where the properties of Theorem 8.1 hold for $-B$ and L^-. On $\Omega_0 \cap \Omega_0^-$ we can apply (8.3) to both B and $-B$ to get $L^-(t,x) = L(t,-x)$ as common sense would dictate.

$$\begin{aligned} L(t,x) &= \tfrac{1}{2}L(t,x) + \tfrac{1}{2}L^-(t,-x) \\ &= (B_t(\omega) - x)^+ - (B_0(\omega) - x)^+ - I(t,x,\omega) \\ &\qquad + (-B_t(\omega) + x)^+ - (-B_0(\omega) + x)^+ - I^-(t,-x,\omega) \\ &= |B_t(\omega) - x| - |B_0(\omega) - x| \\ &\qquad - \int_0^t \mathbf{1}_{[x,\infty)}(B_s)\,dB_s - \int_0^t \mathbf{1}_{[-x,\infty)}(-B_s)\,d(-B)_s. \end{aligned}$$

The construction of the stochastic integral shows that if the signs of all the B-increments are switched, $\int_0^t X\,d(-B) = -\int_0^t X\,dB$ a.s. Consequently, almost surely,

$$\begin{aligned} &\int_0^t \mathbf{1}_{[x,\infty)}(B_s)\,dB_s + \int_0^t \mathbf{1}_{[-x,\infty)}(-B_s)\,d(-B)_s \\ &= \int_0^t \big[\mathbf{1}_{[x,\infty)}(B_s) - \mathbf{1}_{(-\infty,x]}(B_s)\big]\,dB_s = \int_0^t \operatorname{sign}(B_s - x)\,dB_s. \end{aligned}$$

In fact the convention of the sign function at 0 is immaterial because $\int_0^\infty \mathbf{1}\{B_s = x\}\,ds = 0$ almost surely for any fixed x. Substituting the last integral back up finishes the proof. □

While local time is a subtle object, its distribution has a surprisingly simple description. Considering a particular fixed Brownian motion B, let

$$M_t = \sup_{0 \le s \le t} B_s \tag{8.17}$$

be its *running maximum*, and write $L = \{L_t = L(t,0) : t \in \mathbf{R}_+\}$ for the local time process at the origin.

Theorem 8.7. *Let B be a standard Brownian motion. Then we have this equality in distribution of processes:*

$$(M - B, M) \overset{d}{=} (|B|, L).$$

The absolute value $|B_t|$ is thought of as Brownian motion reflected at the origin. The first equality $M - B \overset{d}{=} |B|$ indicates that Brownian motion reflects off its running maximum statistically in exactly the same way as off the fixed point 0. Recall from (2.42) that the *reflection principle* of Brownian motion gives the fixed time distributional equality $M_t \overset{d}{=} |B_t|$ but this does not extend to a process-level equality.

To prove Theorem 8.7 we introduce Skorohod's solution to the *reflection problem*. $C(\mathbf{R}_+)$ is the space of continuous functions on $\mathbf{R}_+$.

Lemma 8.8. *Let $b \in C(\mathbf{R}_+)$ such that $b(0) \geq 0$. Then there is a unique pair (a, ℓ) of functions in $C(\mathbf{R}_+)$ determined by these properties:*

(i) *$a = b + \ell$ and $a(t) \geq 0$ for all $t \in \mathbf{R}_+$.*

(ii) *$\ell(0) = 0$, ℓ is nondecreasing and*

$$\int_0^\infty a(t)\, d\ell(t) = 0. \tag{8.18}$$

Remark 8.9. Condition (8.18) is equivalent to $\Lambda\{t \in \mathbf{R}_+ : a(t) > 0\} = 0$ where Λ is the Lebesgue-Stieltjes measure of ℓ. This follows from the basic integration fact that for a nonnegative function a, $\int a\, d\Lambda = 0$ if and only if $a = 0$ Λ-a.e. In particular, if $a > 0$ on (s,t), by the continuity of ℓ,

$$0 = \Lambda(s,t) = \ell(t-) - \ell(s) = \ell(t) - \ell(s).$$

The point of (i)–(ii) is that ℓ provides the minimal amount of upward push to keep a nonnegative. (8.18) says that no push happens when $a > 0$ which would be wasted effort. (See Exercise 8.2 on this point of the optimality of the solution.) The pair (a, ℓ) is said to solve the *reflection problem* for b.

Proof of Lemma 8.8. *Existence.* Define

$$\ell(t) = \sup_{s \in [0,t]} b^-(s) \quad \text{and} \quad a(t) = b(t) + \ell(t). \tag{8.19}$$

We check that the functions ℓ and a defined above satisfy the required properties. We leave it as an exercise to check that ℓ is continuous.

For part (i) it remains only to observe that

$$a(t) = b(t) + \ell(t) \ge b(t) + b^-(t) = b^+(t) \ge 0.$$

For part (ii), $b(0) \ge 0$ implies $\ell(0) = b^-(0) = 0$ and the definition of ℓ makes it nondecreasing. By Exercise 1.3, to show (8.18) it suffices to show that if t is a point of strict increase for ℓ then $a(t) = 0$. So fix t and suppose that $\ell(t_1) > \ell(t)$ for all $t_1 > t$. From the definition it then follows that for each $k \in \mathbf{N}$ there exists $s_k \in [t, t + k^{-1}]$ such that

$$0 \le \ell(t) < b^-(s_k) \le \ell(t + k^{-1}). \tag{8.20}$$

Let $k \nearrow \infty$ which will take $s_k \to t$. From $b^-(s_k) > 0$ it follows that $b(s_k) < 0$ and then by continuity $b(t) \le 0$. Again by continuity, in the limit (8.20) becomes $\ell(t) = b^-(t)$. Consequently $a(t) = b(t) + \ell(t) = -b^-(t) + b^-(t) = 0$.

Uniqueness. Suppose (a, ℓ) and $(\tilde{a}, \tilde{\ell})$ are two pairs of $C(\mathbf{R}_+)$-functions that satisfy (i)–(ii), and $\tilde{a}(t) > a(t)$ for some $t > 0$. Let $u = \sup\{s \in [0, t] : \tilde{a}(s) = a(s)\}$. By continuity $\tilde{a}(u) = a(u)$ and so $u < t$. Then $\tilde{a}(s) > a(s)$ for all $s \in (u, t]$ because (again by continuity) $\tilde{a} - a$ cannot change sign without passing through 0.

Since $a(s) \ge 0$ we see that $\tilde{a}(s) > 0$ for all $s \in (u, t]$, and then by property (8.18) (or by its equivalent form in Remark 8.9), $\tilde{\ell}(t) = \tilde{\ell}(u)$. Since

$$a - \ell = b = \tilde{a} - \tilde{\ell} \quad \text{implies} \quad \tilde{a} - a = \tilde{\ell} - \ell$$

and ℓ is nondecreasing, we get

$$\tilde{a}(t) - a(t) = \tilde{\ell}(t) - \ell(t) \le \tilde{\ell}(u) - \ell(u) = \tilde{a}(u) - a(u) = 0$$

contradicting $\tilde{a}(t) > a(t)$.

Thus $\tilde{a} \le a$. The roles of $\tilde{a}$ and a can be switched in the argument, and we conclude that $\tilde{a} = a$. The above implication then gives $\tilde{\ell} = \ell$. □

Tanaka's formula (8.14) and the properties of local time established in Theorem 8.1 show that the pair $(|B - x|, L(\cdot, x))$ solves the reflection problem for the process

$$\widetilde{B}_t = |B_0 - x| + \int_0^t \operatorname{sign}(B_s - x)\, dB_s. \tag{8.21}$$

So by (8.19) and the uniqueness part of Lemma 8.8 we get the new identity

$$L(t, x) = \sup_{0 \le s \le t} \widetilde{B}_s^-. \tag{8.22}$$

Lévy's criterion (Theorem 6.14) tells us that $\widetilde{B}$ is a Brownian motion. For this we only need to check the quadratic variation of the stochastic integral:

$$\Big[\int_0^{\cdot} \operatorname{sign}(B_s - x)\,dB_s\Big]_t = \int_0^t \big(\operatorname{sign}(B_s - x)\big)^2\,ds = t.$$

Proof of Theorem 8.7. Take $B_0 = x = 0$ in (8.14) and (8.21) and let $\bar{B} = -\widetilde{B}$ which is now another standard Brownian motion. (8.14) takes the form $|B| = -\bar{B} + L$. This together with property (8.2) of local time implies that $(|B|, L)$ solves the reflection problem for $-\bar{B}$.

On the other hand, we can use (8.19) to obtain a solution (a, ℓ) with

$$\ell(t) = \sup_{0\le s\le t} (-\bar{B})_s^- = \sup_{0\le s\le t} \bar{B}_s \equiv \bar{M}_t$$

where the last identity defines the running maximum process $\bar{M}_t$. Thus we have two solutions for the reflection problem for $-\bar{B}$, namely $(|B|, L)$ and $(\bar{M} - \bar{B}, \bar{M})$. By uniqueness $(|B|, L) = (\bar{M} - \bar{B}, \bar{M})$. Since $(\bar{M} - \bar{B}, \bar{M}) \stackrel{d}{=} (M - B, M)$, we have the claimed distributional equality $(|B|, L) \stackrel{d}{=} (M - B, M)$. $\square$

8.2. Change of measure

In this section we take up a new idea, changing a random variable to a different one by changing the probability measure on the underlying sample space. This elementary example illustrates. Let X be a $\mathcal{N}(0,1)$ random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\alpha \in \mathbf{R}$. On Ω define the random variable $Z = \exp(\alpha X - \alpha^2/2)$. Then $Z \geq 0$ and $E(Z) = 1$. Consequently we can use Z as a Radon-Nikodym derivative to define a new probability measure Q on $(\Omega, \mathcal{F})$ by $Q(A) = E(Z\mathbf{1}_A)$ for measurable sets $A \in \mathcal{F}$. Now that we have two measures P and Q on the same measurable space, expectations under these two measures need to be distinguished notationally. A common way to do this is to write $E^P(f) = \int f\,dP$ for expectation under P, and similarly E^Q for expectation under Q.

To see what happens to the distribution of X when we replace P with Q, let us derive the density of X under Q. Let f be a bounded Borel function on $\mathbf{R}$. Using the density of X under P we see that

$$\begin{aligned} E^Q[f(X)] &= E^P[Z \cdot f(X)] = E^P[e^{\alpha X - \alpha^2/2} f(X)] \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x) e^{\alpha x - \alpha^2/2 - x^2/2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} f(x) e^{-(x-\alpha)^2/2}\,dx. \end{aligned}$$

Consequently, under Q, X is a normal variable with mean α and variance 1. The switch from P to Q added a mean to X. The transformation works also in the opposite direction. If we start with Q and then define P by $dP = Z^{-1}dQ$ then the switch from P to Q removes the mean from X.

Our goal is to achieve the same with Brownian motion. By a change of the underlying measure with an explicitly given Radon-Nikodym derivative we can add a drift to a Brownian motion or change the initial drift.

Let now $(\Omega, \mathcal{F}, P)$ be a complete probability space with a complete filtration $\{\mathcal{F}_t\}$, and $B(t) = [B_1(t), \dots, B_d(t)]^T$ a d-dimensional Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$. $B(0)$ can be an arbitrary $\mathbf{R}^d$-valued $\mathcal{F}_0$-measurable random variable. Let $H(t) = (H_1(t), \dots, H_d(t))$ be an $\mathbf{R}^d$-valued adapted, measurable process such that

$$\int_0^T |H(t,\omega)|^2\,dt < \infty \quad \text{for all } T < \infty, \text{ for } P\text{-almost every } \omega. \tag{8.23}$$

Under these conditions the real valued stochastic integral

$$\int_0^t H(s)\,dB(s) = \sum_{i=1}^d \int_0^t H_i(s)\,dB_i(s)$$

is well-defined, as explained in either Chapter 4 or Section 5.5. Define the stochastic exponential

$$Z_t = \exp\Bigl\{\int_0^t H(s)\,dB(s) - \tfrac{1}{2}\int_0^t |H(s)|^2\,ds\Bigr\}. \tag{8.24}$$

Here it is important that $|x| = (x_1^2 + \cdots + x_d^2)^{1/2}$ is the Euclidean norm. By using a continuous version of the stochastic integral we see that Z_t is a continuous process. Itô's formula shows that

$$Z_t = 1 + \int_0^t Z_s H(s)\,dB(s) \tag{8.25}$$

and so Z_t is a continuous local martingale. Let $\{\tau_n\}$ be a localizing sequence for Z_t. Since $Z_t \ge 0$, Fatou's lemma can be applied to see that

$$EZ_t = E(\lim_{n\to\infty} Z_{t\wedge\tau_n}) \le \varliminf_{n\to\infty} E(Z_{t\wedge\tau_n}) = EZ_0 = 1. \tag{8.26}$$

If will be of fundamental importance for the sequel to make sure that Z_t *is a martingale.* This can be guaranteed by suitable assumptions on H. Let us return to this point later and continue now under the assumption that Z_t is a martingale. Then

$$EZ_t = EZ_0 = 1 \qquad \text{for all } t \ge 0.$$

Thus as in the opening paragraph, each Z_t qualifies as a Radon-Nikodym derivative that gives a new probability measure Q_t. If we restrict Q_t to $\mathcal{F}_t$, the entire family $\{Q_t\}_{t\in\mathbf{R}_+}$ acquires a convenient consistency property. So for each $t \in \mathbf{R}_+$ let us define a probability measure Q_t on $(\Omega, \mathcal{F}_t)$ by

$$Q_t(A) = E^P[\mathbf{1}_A Z_t] \quad \text{for } A \in \mathcal{F}_t. \tag{8.27}$$

Then by the martingale property, for $s < t$ and $A \in \mathcal{F}_s$,

$$\begin{aligned} Q_t(A) &= E^P[\mathbf{1}_A Z_t] = E^P[\mathbf{1}_A E^P(Z_t|\mathcal{F}_s)] = E^P[\mathbf{1}_A Z_s] \\ &= Q_s(A). \end{aligned} \tag{8.28}$$

This is the consistency property.

Additional conditions are needed if we wish to work with a single measure Q instead of the family $\{Q_t\}_{t\in\mathbf{R}_+}$. But as long as our computations are restricted to a fixed finite time horizon $[0, T]$ we have all that is needed. The consistency property (8.28) allows us to work with just one measure Q_T.

Continuing towards the main result, define the continuous, adapted $\mathbf{R}^d$-valued process W by

$$W(t) = B(t) - B(0) - \int_0^t H(s)\,ds.$$

The theorem, due to Cameron, Martin and Girsanov, states that under the transformed measure W is a Brownian motion.

Theorem 8.10. *Fix $0 < T < \infty$. Assume that $\{Z_t : t \in [0,T]\}$ is a martingale, H satisfies (8.23), and define the probability measure Q_T on $\mathcal{F}_T$ by (8.27). Then on the probability space $(\Omega, \mathcal{F}_T, Q_T)$ the process $\{W(t) : t \in [0,T]\}$ is a d-dimensional standard Brownian motion relative to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$.*

Remark 8.11. Here is a quick hand-waving proof of Theorem 8.10. Lévy's criterion (Theorem 6.14) is employed to check that W is a Brownian motion under Q_T. $[W] = [B]$ because $\int_0^t H(s)\,ds$ is a FV process. This works under both P and Q_T because these measures are equivalent in the sense that they have the same sets of measure zero. Under P

$$\begin{aligned} d(WZ) &= W\,dZ + Z\,dW + d[W,Z] \\ &= WZH\,dB + Z\,dB - ZH\,dt + ZH\,dt \\ &= Z(WH+1)\,dB. \end{aligned}$$

Thus WZ is a continuous local L^2 martingale under P. Let us calculate as if WZ were a martingale under P and W integrable under Q_T. Let $0 \le s < t \le T$ and $A \in \mathcal{F}_s$.

$$\begin{aligned} E^{Q_T}[\mathbf{1}_A W_t] &= E^{Q_t}[\mathbf{1}_A W_t] = E^P[\mathbf{1}_A W_t Z_t] = E^P[\mathbf{1}_A W_s Z_s] \\ &= E^{Q_s}[\mathbf{1}_A W_s] = E^{Q_T}[\mathbf{1}_A W_s]. \end{aligned}$$

This would show that $\{W_t : t \in [0,T]\}$ is a martingale under Q_T. A martingale W with quadratic variation $[W]_t = t$ is a Brownian motion. At the end of the section we make this argument precise with localization, and also generalize it to a statement about how local martingales under P turn into semimartingales under Q_T.

Let us illustrate how Theorem 8.10 can be used in practice. We also see that the restriction to finite time horizons is not necessarily a handicap.

Example 8.12. Let $a < 0$, $\mu \in \mathbf{R}$, and B_t a standard Brownian motion. Let σ be the first time when B_t hits the space-time line $x = a - \mu t$. We find the probability distribution of σ.

To apply Girsanov's theorem we formulate the question in terms of $X_t = B_t + \mu t$, Brownian motion with drift μ. We can express σ as

$$\sigma = \inf\{t \ge 0 : X_t = a\}.$$

Define $Z_t = e^{-\mu B_t - \mu^2 t/2}$ and $Q_t(A) = E^P(Z_t \mathbf{1}_A)$. Check that Z_t is a martingale. Thus under Q_t, $\{X_s : 0 \le s \le t\}$ is a standard Brownian motion. Compute as follows.

$$\begin{aligned} P(\sigma > t) &= E^Q(Z_t^{-1}\mathbf{1}\{\sigma > t\}) = e^{-\mu^2 t/2} E^Q\big(e^{\mu X_t}\mathbf{1}\{\inf_{0\le s\le t} X_s > a\}\big) \\ &= e^{-\mu^2 t/2} E^P\big(e^{-\mu B_t}\mathbf{1}\{ M_t < -a\}\big). \end{aligned}$$

On the last line we introduced a new standard Brownian motion $B = -X$ with running maximum $M_t = \sup_{0\le s\le t} B_s$, and then switched back to using the distribution P of standard Brownian motion. From (2.48) in Exercise 2.24 we can read off the density of (B_t, M_t). After some calculus

$$P(\sigma > t) = \int_{-\infty}^{-t^{-1/2}a+\mu\sqrt{t}} \frac{e^{x^2/2}}{\sqrt{2\pi}}\,dx \;-\; e^{2\mu a}\int_{-\infty}^{t^{-1/2}a+\mu\sqrt{t}} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx$$

Let us ask for $P(\sigma = \infty)$, that is, the probability that X stays forever in (a, ∞). By taking $t \to \infty$ above we get

$$P(\sigma = \infty) = \begin{cases} 0 & \mu \le 0 \\ 1 - e^{2\mu a} & \mu > 0. \end{cases}$$

We can deduce the same answer by letting $b \nearrow \infty$ in (6.21).

Before proving the theorem, let us address the question of when Z_t is a martingale. One sufficient hypothesis one can find in the literature is *Novikov's condition*

$$E\Big[\exp\Big\{\tfrac{1}{2}\int_0^T |H(s)|^2\,ds\Big\}\Big] < \infty. \tag{8.29}$$

(See Section 3.5 in [11].) We give here proofs for two cases. First a simple Gronwall argument for a bounded H. Then Itô arguments for a case where H can be an unbounded function of Brownian motion.

Theorem 8.13. *Suppose $\{H(s) : s \in [0, T]\}$ is a bounded, adapted, measurable $\mathbf{R}^d$-valued process. Then the stochastic exponential $\{Z_s : s \in [0, T]\}$ defined by (8.24) is an L^2 bounded martingale.*

Proof. Begin with (8.25), but stop it at $\sigma_m = \inf\{t \ge 0 : Z_t \ge m\}$.

$$Z_{t\wedge\sigma_m} = 1 + \int_0^t Z_s \mathbf{1}_{[0,\sigma_m]}(s) H(s)\,dB(s).$$

Square both sides, use $(a + b)^2 \le 2a^2 + 2b^2$.

$$Z_{t\wedge\sigma_m}^2 \le 2 + 2\Big(\int_0^t Z_s \mathbf{1}_{[0,\sigma_m]}(s) H(s)\,dB(s)\Big)^2.$$

With a bounded integrand, the stochastic integral on the right is a martingale. Use Doob's inequality, the isometry of stochastic integration, and the boundedness of H.

$$\begin{aligned}
E\big[\sup_{s\in[0,t]} Z^2_{s\wedge\sigma_m}\big] &\le 2 + 2E\Big[\sup_{s\in[0,t]}\Big(\int_0^s Z_u \mathbf{1}_{[0,\sigma_m]}(u)H(u)\,dB(u)\Big)^2\Big] \\
&\le 2 + 2CE\Big[\Big(\int_0^t Z_u \mathbf{1}_{[0,\sigma_m]}(u)H(u)\,dB(u)\Big)^2\Big] \\
&= 2 + 2CE\int_0^t Z_u^2 \mathbf{1}_{[0,\sigma_m]}(u)|H(u)|^2\,du \\
&\le 2 + 2CE\int_0^t Z^2_{u\wedge\sigma_m}\,du \\
&\le 2 + 2C\int_0^t E\big[\sup_{r\in[0,u]} Z^2_{r\wedge\sigma_m}\big]\,du.
\end{aligned}$$

An application of Gronwall's inequality (Lemma A.20) gives

$$E\big[\sup_{s\in[0,t]} Z^2_{s\wedge\sigma_m}\big] \le 2e^{2Ct}, \quad t\in[0,T].$$

Let $m \nearrow \infty$ to obtain $E[(Z_T^*)^2] \le 4e^{4CT}$. Exercise 3.8 gives the martingale property of Z_t. □

Next we extend the result above to a class of unbounded integrands.

Theorem 8.14. *Let $b : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ be a Borel function for which there exist constants $0 < R, C < \infty$ such that, for all $t \in \mathbf{R}_+$,*

(8.30) $$|b(t,x)| \le C \text{ for } |x| \le R \quad\text{and}\quad |b(t,x)| \le C|x| \text{ for } |x| > R.$$

Let B be a d-dimensional Brownian motion with $B_0 \in L^2(P)$. Then

(8.31) $$Z_t = \exp\Big\{\int_0^t b(s,B(s))\,dB(s) - \tfrac12\int_0^t |b(s,B(s))|^2\,ds\Big\}$$

is a martingale.

Proof. The integrals in the exponential in Z_t are well-defined because $B(s)$ is continuous. For a fixed $T < \infty$ and ω the path $\{B(s,\omega) : s \in [0,T]\}$ lies in some compact ball. Consequently $b(s, B(s))$ is bounded as s varies over $[0,T]$ and condition (8.23) is satisfied.

Define

$$\tau_n = \inf\{t \ge 0 : |B_t| \ge n\}.$$

The localized process can be written as

$$\begin{aligned} Z_{t\wedge\tau_n} &= \exp\Big\{\int_0^{t\wedge\tau_n} b(s,B(s))\,dB(s) - \tfrac12\int_0^{t\wedge\tau_n} |b(s,B(s))|^2\,ds\Big\} \\ &= \exp\Big\{\int_0^t \mathbf{1}_{(0,\tau_n]}(s) b(s,B(s))\,dB(s) - \tfrac12\int_0^t \mathbf{1}_{(0,\tau_n]}(s)|b(s,B(s))|^2\,ds\Big\}. \end{aligned}$$

Theorem 8.13 applies to the bounded integrand $H(s) = \mathbf{1}_{(0,\tau_n]}(s)b(s,B(s))$. Hence $Z_{t\wedge\tau_n}$ is an L^2 martingale and thereby $\{\tau_n\}$ a localizing sequence of stopping times for Z_t.

By Exercise 3.9, to show that Z_t is a martingale, it suffices to show that for each fixed $t > 0$, the sequence $\{Z_{t\wedge\tau_n}\}_{n\in\mathbf{N}}$ is uniformly integrable. The case $t = 0$ is clear since $Z_0 = 1$.

By assumption (8.30) we can fix a constant c such that

$$|x\cdot b(t,x)| \le c(1+|x|^2).$$

Itô's formula (6.9) (without the jump terms) applied to the process

$$f(t,Z_t,B(t)) = Z_t(1+|B(t)|^2)e^{-2(d+c)t}$$

gives

$$\begin{aligned} &Z_t(1+|B(t)|^2)e^{-2(d+c)t} = 1+|B(0)|^2 \\ &+\int_0^t (1+|B(s)|^2)e^{-2(d+c)s}\,dZ_s + 2\sum_{i=1}^d \int_0^t Z_sB_i(s)e^{-2(d+c)s}\,dB_i(s) \\ &\qquad - 2(d+c)\int_0^t Z_s(1+|B(s)|^2)e^{-2(d+c)s}\,ds \\ &\quad + 2d\int_0^t Z_se^{-2(d+c)s}\,ds + 2\sum_{i=1}^d\int_0^t B_i(s)e^{-2(d+c)s}\,d[Z,B_i]_s. \end{aligned} \tag{8.32}$$

Rewrite and bound the last term as follows:

$$\begin{aligned} 2\sum_{i=1}^d \int_0^t B_i(s)e^{-2(d+c)s}\,d[Z,B_i]_s &= 2\int_0^t e^{-2(d+c)s}Z_s\,B(s)\cdot b(s,B(s))\,ds \\ &\le 2c\int_0^t Z_s(1+|B(s)|^2)e^{-2(d+c)s}\,ds. \end{aligned}$$

This bound shows that the sum of three last terms in (8.32) is nonpositive. Use (8.25) to substitute for dZ and we have the following inequality:

$$\begin{aligned} Z_t(1+|B(t)|^2)e^{-2(d+c)t} &\le 1+|B(0)|^2 \\ &+\int_0^t (1+|B(s)|^2)e^{-2(d+c)s} Z_s\, b(s,B(s))\,dB(s) \\ &+2\int_0^t Z_s e^{-2(d+c)s} B(s)\,dB(s). \end{aligned}$$

Replace t with $t\wedge\tau_n$, and use (5.32) to take the restriction to $(0,\tau_n]$ inside the stochastic integrals:

$$\begin{aligned} Z_{t\wedge\tau_n}(1+|B(t\wedge\tau_n)|^2)e^{-2(d+c)t\wedge\tau_n} &\le 1+|B(0)|^2 \\ &+\int_0^t \mathbf{1}_{(0,\tau_n]}(s)(1+|B(s)|^2)e^{-2(d+c)s} Z_s\, b(s,B(s))\,dB(s) \\ &+2\int_0^t \mathbf{1}_{(0,\tau_n]}(s) Z_s e^{-2(d+c)s} B(s)\,dB(s). \end{aligned}$$

The integrands in the two stochastic integrals are members of $\mathcal{L}_2(B)$ as defined below (4.1). Consequently the integrals are mean zero L^2 martingales, and we can take expectations over the inequality above. We further restrict the integral of the leftmost member to the event $\tau_n < t$:

$$E\big[Z_{t\wedge\tau_n}(1+|B(t\wedge\tau_n)|^2)e^{-2(d+c)t\wedge\tau_n}\,\mathbf{1}\{\tau_n<t\}\big] \le 1+E[\,|B(0)|^2\,].$$

On the right we have a finite constant due to the assumption $B(0)\in L^2$. On the event $\tau_n<t$,

$$|B(t\wedge\tau_n)| = |B(\tau_n)| \ge n$$

where the possibility $|B(\tau_n)|>n$ can happen if $\tau_n=0$. Consequently

$$E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{\tau_n<t\}\big] \le \frac{(1+E|B(0)|^2)e^{2(d+c)t}}{1+n^2} \le C(t)n^{-2}.$$

Now we can show the required uniform integrability.

$$\begin{aligned} &E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{Z_{t\wedge\tau_n}\ge r\}\big] \\ &\le E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{\tau_n<t\}\big] + E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{Z_{t\wedge\tau_n}\ge r\}\,\mathbf{1}\{\tau_n\ge t\}\big] \\ &\le C(t)n^{-2} + E\big[Z_t\,\mathbf{1}\{Z_t\ge r\}\big]. \end{aligned}$$

Given $\varepsilon>0$ pick n_0 so that $C(t)n_0^{-2}<\varepsilon/2$. By the integrability of variables $Z_{t\wedge\tau_n}$ and Z_t ((8.26) for Z_t) we can pick $0<r_0<\infty$ so that

$$\max_{1\le n\le n_0} E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{Z_{t\wedge\tau_n}\ge r_0\}\big]\le\varepsilon \quad\text{and}\quad E\big[Z_t\,\mathbf{1}\{Z_t\ge r\}\big]\le\varepsilon/2.$$

The combination of all the bounds gives

$$\sup_n E\big[Z_{t\wedge\tau_n}\,\mathbf{1}\{Z_{t\wedge\tau_n}\ge r\}\big]\le\varepsilon \quad \text{for } r\ge r_0. \qquad \square$$

Having illustrated the result and the hypotheses, we turn to proving the main result itself.

Proof of Theorem 8.10. Theorem 8.10 is proved with Lévy's criterion, so we need to check that $\{W(t) : t \in [0,T]\}$ is a local martingale under Q_T with quadratic covariation $[W_i, W_j]_t = \delta_{i,j} t$.

The covariation follows by collecting facts we know already. Since $Q_T \ll P$ on $\mathcal{F}_T$, any P-almost sure property on $\mathcal{F}_T$ is also Q_T-almost sure. Condition (8.23) implies that $\int_0^t H(s)\,ds$ is a continuous BV process on $[0,T]$ (Exercise 1.5). Thus by Lemma A.10 $[H_i, H_j] = [B_i, H_j] = 0$. We also get $[B_i, B_j]_t = \delta_{i,j} t$ under Q in a few lines. Given a partition $\pi = \{t_k\}$ of $[0,t]$ with $t \le T$ and $\varepsilon > 0$, define the event

$$A_\pi = \left\{ \left| \sum_{k=0}^{m(\pi)-1} \bigl(B_i(t_{k+1}) - B_i(t_k)\bigr)\bigl(B_j(t_{k+1}) - B_j(t_k)\bigr) - \delta_{i,j} t \right| \ge \varepsilon \right\}.$$

Property $[B_i, B_j]_t = \delta_{i,j} t$ under P means precisely that, for any $\varepsilon > 0$, $P(A_\pi) \to 0$ as $\operatorname{mesh}(\pi) \to 0$. Dominated convergence and the assumed integrability of Z_T give the same under Q_T: $Q_T(A_\pi) = E^P[Z_T \mathbf{1}_{A_\pi}] \to 0$. The application of the dominated convergence theorem is not quite the usual one. We can assert that the integrand $Z_T \mathbf{1}_{A_\pi} \to 0$ in probability, because

$$P\{Z_T \mathbf{1}_{A_\pi} \ge \delta\} \le P(A_\pi) \to 0 \quad \text{for any } \delta > 0.$$

Convergence in probability is a sufficient hypothesis for the dominated convergence theorem (Theorem B.12).

It remains to check that $\{W(t) : t \in [0,T]\}$ is a local martingale under Q_T. This follows by taking $M_t = B_i(t) - B_i(0)$ in Lemma 8.15 below, for each i in turn. With this we can consider Theorem 8.10 proved. □

The above proof is completed by addressing the question of how to transform a local martingale under P so that it becomes a local martingale under Q_T. Continuing with all the assumptions we have made, let $\{M_t : t \in [0,T]\}$ be a continuous local martingale under P such that $M_0 = 0$. Define

$$N_t = M_t - \sum_{j=1}^{d} \int_0^t H_j(s)\, d[M, B_j]_s. \tag{8.33}$$

By the Kunita-Watanable inequality (Proposition 2.19)

$$\left| \int_0^t H_j(s)\, d[M, B_j]_s \right| \le [M]_t^{1/2} \left\{ \int_0^t H_j(s)^2\, ds \right\}^{1/2} \tag{8.34}$$

so the integrals in the definition of N_t are finite. Furthermore, $[M, B_j]_t$ is continuous (Proposition 2.16 or Theorem 3.27) and consequently the integral $\int_0^t H_j(s)\, d[M, B_j]_s$ is a continuous BV process (Exercise 1.5). When we switch measures from P to Q_T it turns out that M continues to be a semimartingale. Equation (8.33) gives us the semimartingale decomposition of M under Q_T because N is a local martingale as stated in the next lemma.

Lemma 8.15. *$\{N_t : t \in [0,T]\}$ is a continuous local martingale under Q_T.*

Proof. By its definition N is a semimartingale under P. Integration by parts (Proposition 5.61), equation (8.25) for Z, and the substitution rules (Corollary 5.60 and Theorem 5.63) give

$$\begin{aligned}
N_t Z_t &= \int_0^t N\,dZ + \int_0^t Z\,dN + [N,Z]_t \\
&= \int_0^t NZH\,dB + \int_0^t Z\,dM - \sum_j \int_0^t ZH_j\,d[M,B_j] \\
&\qquad\qquad + \sum_j \int_0^t ZH_j\,d[M,B_j] \\
&= \int_0^t NZH\,dB + \int_0^t Z\,dM.
\end{aligned}$$

The last line shows that NZ is a continuous local martingale under P.

Now we split the remaining proof into two cases. Assume first that N is uniformly bounded: $|N_t(\omega)| \le C$ for all $t \in [0,T]$ and $\omega \in \Omega$. We shall show that in this case NZ is actually a martingale.

Let $\tau_n \nearrow \infty$ be stopping times that localize NZ under P. Let $0 \le s < t \le T$ and $A \in \mathcal{F}_s$. From the martingale property of $(NZ)^{\tau_n}$ we have

$$E^P\big[\mathbf{1}_A N_{t\wedge\tau_n} Z_{t\wedge\tau_n}\big] = E^P\big[\mathbf{1}_A N_{s\wedge\tau_n} Z_{s\wedge\tau_n}\big]. \tag{8.35}$$

We claim that

$$\lim_{n\to\infty} E^P\big[\mathbf{1}_A N_{t\wedge\tau_n} Z_{t\wedge\tau_n}\big] \;=\; E^P\big[\mathbf{1}_A N_t Z_t\big]. \tag{8.36}$$

This follows from the generalized dominated convergence theorem (Theorem A.15). We have the almost sure convergences $N_{t\wedge\tau_n} \to N_t$ and $Z_{t\wedge\tau_n} \to Z_t$ and the domination $|N_{t\wedge\tau_n} Z_{t\wedge\tau_n}| \le C Z_{t\wedge\tau_n}$. We also have the limit $E^P[\mathbf{1}_A Z_{t\wedge\tau_n}] \to E^P[\mathbf{1}_A Z_t]$ by the following reasoning. We have assumed that Z_t is a martingale, hence by optional stopping (Theorem 3.6) $Z_{t\wedge\tau_n} = E^P(Z_t \,|\, \mathcal{F}_{t\wedge\tau_n})$. Consequently by Lemma B.16 the sequence $Z_{t\wedge\tau_n}$ is uniformly integrable. The limit $E[\mathbf{1}_A Z_{t\wedge\tau_n}] \to E[\mathbf{1}_A Z_t]$ finally comes from the almost sure convergence, uniform integrability and Theorem 1.20(iv).

Taking limit (8.36) on both sides of (8.35) verifies the martingale property of NZ. Then we can check that N is a martingale under Q_T:

$$E^{Q_T}[\mathbf{1}_A N_t] = E^P\big[\mathbf{1}_A N_t Z_t\big] = E^P\big[\mathbf{1}_A N_s Z_s\big] = E^{Q_T}[\mathbf{1}_A N_s].$$

Now the general case. Define

$$\tau_n = \inf\Big\{t \ge 0 : |M_t| + [M]_t + \int_0^t |H(s)|^2 \, ds \ge n\Big\}.$$

By the continuity of the processes in question, these are stopping times (Corollary 2.10) such that, Q_T-almost surely, $\tau_n \ge T$ for large enough n. We apply the part already proved to M^{τ_n} in place of M, and let $N^{(n)}$ denote the process defined by (8.33) with M^{τ_n} in place of M.

Observe that

$$\int_0^t H_j(s) \, d[M^{\tau_n}, B_j]_s = \int_0^t H_j(s) \, d[M, B_j]_s^{\tau_n} = \int_0^{t\wedge\tau_n} H_j(s) \, d[M, B_j]_s.$$

This tells us two things. First, combined with inequality (8.34) and $|M^{\tau_n}| \le n$, we see from the definition (8.33) that $N^{(n)}$ is bounded. Thus the part of the proof already done implies that $N^{(n)}$ is a martingale under Q_T. Second, looking at the definition again we see that $N^{(n)} = N^{\tau_n}$.

We have found a sequence of stopping times τ_n such that $\tau_n \ge T$ for large enough n Q_T-almost surely, and $\{N_t^{\tau_n} : t \in [0, T]\}$ is a martingale. In other words, we have shown that N is a local martingale under Q_T. □

8.3. Weak solutions for Itô equations

Consider again the Itô equation

$$dX_t = b(t, X_t) \, dt + \sigma(t, X_t) \, dB_t \tag{8.37}$$

with given Borel functions $b : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ and $\sigma : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^{d\times m}$.

A *weak solution* of (8.37) is a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$ and two adapted processes X and B such that

(i) X is $\mathbf{R}^d$-valued,

(ii) B is a standard $\mathbf{R}^m$-valued Brownian motion under $\{\mathcal{F}_t\}$, and

(iii) these two conditions are satisfied:

$$P\Big\{\forall\, T < \infty : \int_0^T |b(s, X_s)| \, ds + \int_0^T |\sigma(s, X_s)|^2 \, ds < \infty\Big\} = 1. \tag{8.38}$$

and

$$X_t = X_0 + \int_0^t b(s, X_s) \, ds + \int_0^t \sigma(s, X_s) \, dB_s, \quad 0 \le t < \infty, \tag{8.39}$$

in the sense that the processes on the right and left are indistinguishable. In other words, to produce a weak solution, one must construct a probability space on which a strong solution exists. (One such construction is enough for existence of a weak solution, while strong existence requires that a solution can be constructed for any Brownian motion and filtration.) We say that *weak existence* holds for equation (8.37) if for any Borel probability distribution ν on $\mathbf{R}^d$ there exists a weak solution X with initial distribution $X_0 \sim \nu$.

Weak solutions can possess two types of uniqueness.

(i) *Pathwise uniqueness* means the following. Suppose in the description above there is a single space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}$ and Brownian motion B, but two solution processes X and $\widetilde{X}$. Then if $X_0 = \widetilde{X}_0$ a.s., the processes X and $\widetilde{X}$ are indistinguishable.

(ii) *Weak uniqueness*, also called *uniqueness in law*, means that any two weak solutions X and $\widetilde{X}$ with the same initial distribution ν have the same distribution as processes. That is, for any measurable set A on the path space C, $P\{X \in A\} = \widetilde{P}\{\widetilde{X} \in A\}$. Note that these statements assume implicitly that in the background we have the structures necessary for two weak solutions: $(\Omega, \mathcal{F}, P)$, $\{\mathcal{F}_t\}$, X and B, and $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$, $\{\widetilde{\mathcal{F}}_t\}$, $\widetilde{X}$ and $\widetilde{B}$.

Under Lipschitz assumption (7.20) Theorem 7.12 showed pathwise uniqueness, and under both assumptions (7.20) and (7.21) Theorem 7.14 showed weak uniqueness. The literature contains another formulation of pathwise uniqueness that allows two distinct filtrations $\{\mathcal{F}_t\}$ and $\{\widetilde{\mathcal{F}}_t\}$, with X adapted to $\{\mathcal{F}_t\}$ and $\widetilde{X}$ adapted to $\{\widetilde{\mathcal{F}}_t\}$. (See for example [11, Section 5.3].) Our formulation agrees with [3, Section 10.4] and [14, Section IX.1].

Let us look at a well-known example of Tanaka where weak existence and uniqueness hold but pathwise uniqueness and strong existence fail.

Example 8.16. Define the sign function now by

$$\operatorname{sign}(x) = \mathbf{1}\{x \ge 0\} - \mathbf{1}\{x < 0\} \tag{8.40}$$

so that $\operatorname{sign}(x)^2 = 1$. Consider the SDE

$$X_t = \int_0^t \operatorname{sign}(X_s)\, dB_s, \qquad 0 \le t < \infty. \tag{8.41}$$

First we argue weak uniqueness. Let (X, B) be a given pair that solves the equation. (Implicitly understood that there is a probability space and a filtration in the background.) Since the integrand $\operatorname{sign}(X_s)$ is bounded, X is a continuous martingale, and

$$[X]_t = \int_0^t \operatorname{sign}(X_s)^2\, ds = t.$$

Thus by Lévy's criterion (Theorem 6.14) X is a standard Brownian motion. We have conclusively characterized every weak solution as a standard Brownian motion, and thereby weak uniqueness holds.

Continue with a solution pair (X, B). Since X is a Brownian motion,

$$\int_0^\infty P(X_s = 0)\, ds = 0$$

and consequently the stochastic integral in (8.41) is indistinguishable from the one with integrand $\mathbf{1}\{X_s > 0\} - \mathbf{1}\{X_s < 0\}$ (Proposition 4.10(c)). From this we see that $(-X, B)$ is also a solution. Thus pathwise uniqueness fails.

Next we show that a strong solution of (8.41) cannot exist. Suppose B is a Brownian motion and let $\{\mathcal{F}_t\}$ be the augmentation of the filtration generated by B. To get a contradiction, assume there exists a process X adapted to $\{\mathcal{F}_t\}$ that satisfies (8.41). Then, as observed, X is a standard Brownian motion and we can apply Tanaka's formula (8.14) to it. Let $L(t, x)$ denote local time for X, as defined in Theorem 8.1. Since $\operatorname{sign}(X)dB = dX$,

$$B_t = \int_0^t \operatorname{sign}(X_s)^2\, dB_s = \int_0^t \operatorname{sign}(X_s)\, dX_s = |X_t| - L(t, 0). \tag{8.42}$$

Formula (8.3) shows that $L(t, 0)$ is measurable with respect to the σ-algebra $\mathcal{G}_t = \sigma\{|X_s| : s \in [0, t]\}$. Hence (8.42) above shows that B_t is $\mathcal{G}_t$-measurable. We conclude that $\mathcal{F}_t$ is contained in the augmentation of $\mathcal{G}_t$. But this last statement cannot hold. For example, the event $\{X_t \geq 1\}$ has positive probability and cannot lie in $\mathcal{G}_t$ because any function of $\{|X_s| : s \in [0, t]\}$ would give the same value for X and $-X$.

But note that weak solutions to (8.41) do exist. Simply turn the previous argument around. *Start* with a Brownian motion X and define another Brownian motion $B_t = \int_0^t \operatorname{sign}(X_s)\, dX_s$. Then $dX_t = \operatorname{sign}(X_t)^2\, dX_t = \operatorname{sign}(X_t)\, dB_t$ which shows that X is a solution. What is the distinction between this and what was just done above? The point is that now B is not any given Brownian motion and X is not adapted to the filtration of B.

We show how Theorem 8.10 enables us to go beyond Lipschitz drifts to prove the existence of a weak solution to an SDE with a more general drift. Let $0 < T < \infty$, $b : [0, T] \times \mathbf{R}^d \to \mathbf{R}^d$ Borel measurable, and B_t denote standard d-dimensional Brownian motion. Consider the following equation on $[0, T] \times \mathbf{R}^d$:

$$dX_t = b(t, X_t)\, dt + dB_t. \tag{8.43}$$

Theorem 8.17. *Assume that b is a bounded Borel function. Then equation (8.43) has a weak solution for any initial distribution ν on $\mathbf{R}^d$.*

Proof. Let X_t be d-dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$ with respect to filtration $\{\mathcal{F}_t\}$ and with initial distribution $X_0 \sim \nu$. By the boundedness of b and Theorem 8.13,

$$Z_t = \exp\Big\{ \int_0^t b(s, X_s)\, dX_s - \tfrac{1}{2} \int_0^t |b(s, X_s)|^2\, ds \Big\}, \quad 0 \le t \le T, \tag{8.44}$$

is a martingale. Define probability measure Q_T on $\mathcal{F}_T$ by $dQ_T = Z_T dP$. By Theorem 8.10, the process

$$B_t = X_t - X_0 - \int_0^t b(s, X_s)\, ds, \quad t \in [0, T], \tag{8.45}$$

is a standard Brownian motion on the probability space $(\Omega, \mathcal{F}_T, Q_T)$, under the original filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. Equation (8.45) is just the integral form of (8.43). On $\mathcal{F}_0$ measures Q_T and P coincide because $Z_0 = 1$. Consequently $X_0 \sim \nu$ also under Q_T. To summarize, process $(X_t)_{t\in[0,T]}$ and Brownian motion $(B_t)_{t\in[0,T]}$ with filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ on $(\Omega, \mathcal{F}_T, Q_T)$ form a weak solution of (8.43) with initial distribution ν. □

Next the above idea of using Girsanov's theorem to create weak solutions is applied to verify the invariant distribution of certain diffusion processes. Let $\varphi \in C^2(\mathbf{R}^d)$ have bounded second derivatives, and consider solutions X to the SDE

$$dX_t = \nabla\varphi(X_t)dt + B_t. \tag{8.46}$$

With bounded second derivatives $\nabla\varphi$ satisfies the Lipschitz and growth assumptions (7.20) and (7.21). Hence solutions to this equation exist, possess both weak and strong uniqueness, and are Markov processes.

We show that the evolution prescribed by (8.46) preserves the Borel measure with density $e^{2\varphi(x)}$ on $\mathbf{R}^d$. If $\int_{\mathbf{R}^d} e^{2\varphi(x)}\, dx < \infty$ we can subtract a constant from φ (without changing (8.46)) and have an invariant probability distribution for the diffusion defined by (8.46). As before, let P^x (with expectation E^x) denote the probability distribution on $C = C_{\mathbf{R}^d}[0,\infty)$ of the process X that solves (8.46) with deterministic initial point $X_0 = x$, for $x \in \mathbf{R}^d$.

Theorem 8.18. *For any Borel function $f \ge 0$ on $\mathbf{R}^d$ and $t > 0$,*

$$\int_{\mathbf{R}^d} E^x[f(X_t)]\, e^{2\varphi(x)}\, dx = \int_{\mathbf{R}^d} f(x)\, e^{2\varphi(x)}\, dx. \tag{8.47}$$

For the proof we need a symmetry property of standard Brownian motion, namely that a Brownian path between two fixed points looks the same in both directions. This is in fact a property of Brownian bridge.

Let Y be a real-valued random variable. Below we use a conditional expectation of the type $E[X \mid Y = y]$ which is a function of a point $y \in \mathbf{R}$. It is defined rigorously as follows. Since $E(X \mid Y)$ is a $\sigma(Y)$-measurable random variable, there exists a Borel function g such that $E(X \,|\, Y) = g(Y)$. Then let $E[X \,|\, Y = y] = g(y)$. If Y has density f_Y on $\mathbf{R}$, then this conditional expectation can be characterized by the requirement

$$E[\, X\, \mathbf{1}_B(Y)] = \int_B E[\, X \,|\, Y = y] f_Y(y)\, dy \tag{8.48}$$

that should hold for all Borel sets $B \subseteq \mathbf{R}$.

Lemma 8.19. *Let P^x denote the distribution of $\mathbf{R}^d$-valued Brownian motion started at $B_0 = x$. Fix $T \in (0, \infty)$ and let $F \geq 0$ be a Borel function on $C[0, T]$. Let B_t denote the coordinate process on C, and $\widetilde{B}_t = B_{T-t}$ for $t \in [0, T]$. Then for $a, b \in \mathbf{R}^d$*

$$E^a[\, F(B_\cdot) \,|\, B_T = b\,] = E^b[\, F(\widetilde{B}_\cdot) \,|\, B_T = a\,] \tag{8.49}$$

Proof of Lemma 8.19. By the standard π-λ and limit arguments, it suffices to consider a function $F(\omega) = F(\omega(0), \omega(s_1), \dots, \omega(s_n), \omega(T))$ for finitely many time points $0 = s_0 < s_1 < \cdots < s_n < s_{n+1} = T$. (A function like this is sometimes called a *local* or a *cylinder* function.) Then one can check that the conditional expectation is given by the following formula:

$$\begin{aligned} &E^a[\, F(B_0, B_{s_1}, \dots, B_{s_n}, B_T) \,|\, B_T = b\,] \\ &\quad = \int_{\mathbf{R}^{nd}} F(a, x_1, \dots, x_n, b)\, \frac{\prod_{i=1}^{n+1} p_{s_i - s_{i-1}}(x_i - x_{i-1})}{p_T(b-a)}\, dx_{1,n} \end{aligned} \tag{8.50}$$

Utilizing the symmetries of the Gaussian kernel a change of variable shows that the above equals

$$E^b[\, F(\widetilde{B}_0, \widetilde{B}_{s_1}, \dots, \widetilde{B}_{s_n}, \widetilde{B}_T) \,|\, B_T = a\,] \tag{8.51}$$

The details of this proof are left for Exercise 8.4 □

Proof of Theorem 8.18. We need several probability measures and processes here, so the notation needs to stretch a little. $E^x[F(X_\cdot)]$ will refer to the distributions of the solution X of the SDE (8.46), as defined above Theorem 8.18. $E^x[F(B_\cdot)]$ will refer to the distributions of $\mathbf{R}^d$-valued Brownian motion as in Lemma 8.19.

Note that Itô's formula gives

$$d\varphi(B_t) = \nabla\varphi(B_t) dB_t + \tfrac{1}{2}\Delta\varphi(B_t) dt. \tag{8.52}$$

We start by constructing a weak solution to (8.46) via Girsanov's Theorem. On a probability space $(\Omega, \mathcal{H}, \mathbf{P})$ let Y be Brownian motion started at x and adapted to the complete filtration $\{\mathcal{H}_t\}$. Define

$$Z_t = e^{\int_0^t \nabla\varphi(Y_s)\, dY_s - \frac{1}{2}\int_0^t |\nabla\varphi(Y_s)|^2\, ds}. \tag{8.53}$$

Function $b(x) = \nabla\varphi(x)$ satisfies assumption (8.30) due to the assumption of bounded second derivatives, and hence Z_t is a martingale by Theorem 8.14. Define the probability measure Q_T on $\mathcal{H}_T$ by $Q_T(A) = E^{\mathbf{P}}[Z_T \mathbf{1}_A]$. By Theorem 8.10, the process

$$W_t = Y_t - x - \int_0^t \nabla\varphi(Y_s)\, dY_s, \quad t \in [0, T] \tag{8.54}$$

is a standard Brownian motion. Equation (8.54) is the same as (8.46), and so Y is a weak solution of (8.46) on time interval $[0, T]$. By weak uniqueness, $\{Y_t\}_{t\in[0,T]}$ has distribution P^x from (8.47).

In the next calculation we represent the solution of (8.46) with Y under Q_T. Then, since under $\mathbf{P}$ Y is Brownian motion started at x, we switch notation to the Brownian path variable $B_{\boldsymbol{\cdot}}$ under its distribution P^x. After this condition on the endpoint $y = B_t$ and in the last step use the symmetry of the Gaussian kernel and (8.49).

$$\begin{aligned}
&E^x[f(X_t)] = E^{Q_T}[f(Y_t)] = E^{\mathbf{P}}[Z_t f(Y_t)] \\
&= E^x\big[f(B_t)e^{\int_0^t \nabla\varphi(B_s)\, dB_s - \frac{1}{2}\int_0^t |\nabla\varphi(B_s)|^2\, ds}\big] \\
&= E^x\big[f(B_t)e^{\varphi(B_t)-\varphi(x)-\frac{1}{2}\int_0^t (\Delta\varphi(B_s)+|\nabla\varphi(B_s)|^2)\, ds}\big] \\
&= \int dy\, p_t(y-x) f(y) e^{\varphi(y)-\varphi(x)} E^x\big[e^{-\frac{1}{2}\int_0^t (\Delta\varphi(B_s)+|\nabla\varphi(B_s)|^2)\, ds} \,\big|\, B_t = y\big] \\
&= \int dy\, p_t(x-y) f(y) e^{\varphi(y)-\varphi(x)} E^y\big[e^{-\frac{1}{2}\int_0^t (\Delta\varphi(B_s)+|\nabla\varphi(B_s)|^2)\, ds} \,\big|\, B_t = x\big].
\end{aligned}$$

Integrate against the density $e^{2\varphi(x)}$ and rearrange the integral.

$$\begin{aligned}
\int e^{2\varphi(x)} E^x[f(X_t)]\, dx &= \int dy\, f(y) e^{2\varphi(y)} \int dx\, p_t(x-y) e^{\varphi(x)-\varphi(y)} \\
&\quad \times E^y\big[e^{-\frac{1}{2}\int_0^t (\Delta\varphi(B_s)+|\nabla\varphi(B_s)|^2)\, ds} \,\big|\, B_t = x\big].
\end{aligned} \tag{8.55}$$

Remove again the conditioning to see that the dx-integral above equals

$$\begin{aligned}
&E^y\big[e^{\varphi(B_t)-\varphi(y)-\frac{1}{2}\int_0^t (\Delta\varphi(B_s)+|\nabla\varphi(B_s)|^2)\, ds}\big] \\
&= E^y\big[e^{\int_0^t \nabla\varphi(B_s)\, dB_s - \frac{1}{2}\int_0^t |\nabla\varphi(B_s)|^2\, ds}\big] \;=\; 1
\end{aligned}$$

where at the end we noticed that we have again an expectation of the martingale Z_t. Equality (8.55) simplifies to give the desired conclusion (8.47). □

Example 8.20. (Ornstein-Uhlenbeck process) Let $\alpha > 0$ and take $\varphi(x) = -\frac{1}{2}\alpha|x|^2$. Then the SDE (8.46) becomes

$$dX_t = -\alpha X_t\, dt + B_t \tag{8.56}$$

and the invariant distribution is the d-dimensional Gaussian $\mathcal{N}(0, \frac{1}{2\alpha} I)$.

Exercises

Exercise 8.1. Show that for a given x, $\int_0^t \mathbf{1}\{B_s = x\}\,ds = 0$ for all $t \in \mathbf{R}_+$ almost surely. *Hint.* Simply take the expectation. Note that this example again shows the interplay between countable and uncountable in probability: in some sense the result says that "for any x, Brownian motion spends zero time at x." Yet of course Brownian motion is somewhere!

Exercise 8.2. Let (a, ℓ) be the solution of the reflection problem in Lemma 8.8. If $(\tilde{a}, \tilde{\ell})$ is a pair of $C(\mathbf{R}_+)$-functions such that $\tilde{\ell}$ is nondecreasing and $\tilde{a} = b + \tilde{\ell} \geq 0$, show that $\tilde{a} \geq a$ and $\tilde{\ell} \geq \ell$.

Exercise 8.3. Let $B_t^{(\mu)} = B_t + \mu t$ be standard Brownian motion with drift μ and running maximum $M_t^{(\mu)} = \sup_{0 \leq s \leq t} B_s^{(\mu)}$. Find the joint density of $(B_t^{(\mu)}, M_t^{(\mu)})$. Naturally you will utilize (2.48).

Exercise 8.4. Fill in the details in the proof of Lemma 8.19. Namely, use the convolution properties of the Gaussian kernel to check that the right-hand side of (8.50) defines a probability measure on $\mathbf{R}^{nd}$. Then check (8.50) itself by verifying that the right-hand side satisfies

$$\begin{aligned} &E^a\big[f(B_T)F(B_\bullet)\big] \\ &= \int_{\mathbf{R}^d} db\, p_T(b-a)\, f(b)\, E^a[\,F(B_0, B_{s_1}, \dots, B_{s_n}, B_T)\,|\,B_T = b\,]. \end{aligned} \tag{8.57}$$

Finally check the equality of (8.50) and (8.51).

Chapter 9

White Noise and a Stochastic Partial Differential Equation

In the first section of this chapter we develop a stochastic integral over time and space, $\int_{(0,t]\times\mathbf{R}^d} Y(s,x)\,dW(s,x)$, where the integrand is a real-valued stochastic process $Y(s,x,\omega)$ indexed by time and space, and the integrator W is white noise. White noise is a random assignment of Gaussian-distributed signed mass on space-time so that masses on disjoint sets are independent. This is a natural generalization of the independent increments property of Brownian motion. In passing we define a multiparameter generalization of Brownian motion called the Brownian sheet. Armed with this new integral, we can make rigorous sense of partial differential equations with noise. In Section 9.2 we develop a key example of such a stochastic partial differential equation (SPDE), namely a heat equation with stochastic noise.

9.1. Stochastic integral with respect to white noise

9.1.1. White noise and the isonormal process. Let $(X,\mathcal{B},\lambda)$ be a σ-finite measure space. A *white noise* on the space X relative to the measure λ is a mean zero Gaussian process $\{W(A) : A \in \mathcal{B},\ \lambda(A) < \infty\}$ with covariance

$$E[W(A)W(B)] = \lambda(A \cap B). \tag{9.1}$$

We also say that W is a white noise based on the space $(X,\mathcal{B},\lambda)$. Exercise 9.1 asks you to use Exercise 1.19 to establish the existence of such a process.

White noise is a finitely additive random set function: if $A \cap B = \emptyset$ then $W(A)$ and $W(B)$ are independent and $W(A \cup B) = W(A) + W(B)$ a.s. (Exercise 9.2). In particular $W(\emptyset) = 0$ a.s., which can also be seen from

$$E[W(\emptyset)^2] = \lambda(\emptyset) = 0.$$

White noise is an L^2-valued measure, which means that countable additivity holds in the L^2 sense: if $\{A_k\}$ are pairwise disjoint, $A = \bigcup_{k\ge 1} A_k$ and $\lambda(A) < \infty$, then

$$\begin{aligned} E\Big[\Big(W(A) - \sum_{k=1}^{n} W(A_k)\Big)^2\Big] &= E\Big[\Big(W\Big(\bigcup_{k=n+1}^{\infty} A_k\Big)\Big)^2\Big] \\ &= \lambda\Big(\bigcup_{k=n+1}^{\infty} A_k\Big) \underset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{9.2}$$

If we complement this with a standard result on random series (Theorem B.11) we also get

$$\sum_{k=1}^{\infty} W(A_k) = W(A) \quad \text{almost surely.} \tag{9.3}$$

In general, white noise is not a signed measure for any fixed ω. (Exercise 9.4).

White noise $\{W(A) : A \in \mathcal{B},\ \lambda(A) < \infty\}$ extends readily to a mean zero Gaussian process $\{W(h) : h \in L^2(\lambda)\}$ with covariance

$$E[W(g)W(h)] = \int_X gh\,d\lambda, \quad g, h \in L^2(\lambda). \tag{9.4}$$

This is the *isonormal process* on $L^2(\lambda)$. The construction uses the L^2-isometry inherent already in formula (9.1).

Suppose first $h = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$ is a simple L^2 function where the α_i are real and nonzero and the $\{A_i\}$ are pairwise disjoint. The assumption $h \in L^2(\lambda)$ then guarantees that $\lambda(A_i) < \infty$ for each i. Define

$$W(h) = \sum_{i=1}^{n} \alpha_i W(A_i). \tag{9.5}$$

Observe that $W(h)$ is a mean zero Gaussian random variable with variance

$$E[W(h)^2] = \sum_{i=1}^{n} \alpha_i^2 \lambda(A_i) = \|h\|^2_{L^2(\lambda)}. \tag{9.6}$$

Here we have the isometry: for simple functions h, $\|W(h)\|_{L^2(P)} = \|h\|_{L^2(\lambda)}$.

Next we extend. Given $h \in L^2(\lambda)$, find simple functions $h_n \in L^2(\lambda)$ such that $\|h - h_n\|_{L^2(\lambda)} \to 0$. Then $\{h_n\}$ is a Cauchy sequence in $L^2(\lambda)$, in other words, given $\varepsilon > 0$ there exists $n_0 \in \mathbf{N}$ such that $\|h_m - h_n\|_{L^2(\lambda)} < \varepsilon$ for

$m, n \geq n_0$. Since $\|W(h_m) - W(h_n)\|_{L^2(P)} = \|h_m - h_n\|_{L^2(\lambda)}$, $\{W(h_n)\}$ is a Cauchy sequence in $L^2(P)$. By the completeness of $L^2(P)$, $W(h_n)$ converges in $L^2(P)$ to a random variable which we call $W(h)$. This constructs the isonormal process, except for a few bits left to tidy up, relegated to Exercise 9.7.

It is quite obvious that we were really constructing an integral with respect to the random L^2 measure W, and consequently we could also use the notation $W(h) = \int h\,dW$ for the isonormal process. The shortcoming here is that we did not treat random integrands. This will be achieved in Section 9.1.3 with the help of martingale theory, in the spirit of Itô integrals.

9.1.2. White noise, Brownian motion and Brownian sheet. Let W be a white noise on $\mathbf{R}_+$ with respect to Lebesgue measure m. Then $B_t = W(0, t]$ defines a Gaussian process with the covariance of Brownian motion: $EB_sB_t = m((0, s] \cap (0, t]) = s \wedge t$. Thus a continuous version of this process, whose existence is given by the Kolmogorov-Centsov criterion Theorem B.20, is standard Brownian motion.

The Brownian sheet is a generalization of Brownian motion to a process with a multidimensional index set. Let $\mathbf{R}_+^d = \{t = (t_1, \dots, t_d) \in \mathbf{R}^d : t_i \geq 0 \,\forall\, i\}$ denote the nonnegative quadrant of d-dimensional space and let W be a white noise on $(\mathbf{R}_+^d, \mathcal{B}_{\mathbf{R}_+^d}, m)$ where m is Lebesgue measure. The *Brownian sheet* $\{B_t : t \in \mathbf{R}_+^d\}$ is defined by $B_t = W((0, t])$ where for $t = (t_1, \dots, t_d) \in \mathbf{R}_+^d$ the d-dimensional interval is $(0, t] = \prod_{i=1}^d (0, t_i]$. An equivalent way to define the distribution of the Brownian sheet is to say that $\{B_t : t \in \mathbf{R}_+^d\}$ is a mean zero Gaussian process with covariance

$$E(B_sB_t) = (s_1 \wedge t_1) \cdot (s_2 \wedge t_2) \cdots (s_d \wedge t_d). \tag{9.7}$$

It can be shown that, just like Brownian motion, Brownian sheet $B_{\boldsymbol{\cdot}}$ has a continuous version (Exercise 9.8).

9.1.3. The stochastic integral with respect to white noise. To treat SDEs rigorously we developed a stochastic integral with respect to processes indexed by time. In order to treat SPDEs rigorously we need integration for processes in space-time. This need gives rise to a theory of integration with respect to martingale measures. We state some general definitions, but treat thoroughly only the case of white noise.

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a complete filtration $\{\mathcal{F}_t\}$. We take as our state space $\mathbf{R}^d$, although a more general metric space would do just as well. Let $\mathcal{A}$ be a collection of Borel subsets of $\mathbf{R}^d$ such that $\mathcal{A}$ is closed under unions and set differences. That is, $A, B \in \mathcal{A}$ imply that $A \cup B \in \mathcal{A}$ and $A \smallsetminus B \in \mathcal{A}$. Such a collection is sometimes called a *ring* of sets [4, Sect. 3.1]. A ring is not quite an algebra: a ring $\mathcal{A}$ is an algebra iff

$\mathbf{R}^d \in \mathcal{A}$. An example of a ring is the collection of bounded Borel sets. (A Borel set is *bounded* if it is contained in some cube $[-m, m]^d$.)

Let U be a real-valued function on $\mathcal{A} \times \Omega$ such that for each set $A \in \mathcal{A}$, $U(A, \omega)$ is a square-integrable random variable: $E[U(A)^2] < \infty$. U is an *L^2-valued measure* on $\mathcal{A}$ if the next two conditions are satisfied.

(i) U is finitely additive: $U(A \cup B) = U(A) + U(B)$ a.s. whenever $A \cap B = \emptyset$.

(ii) U is countably additive on $\mathcal{A}$ in the L^2 sense: whenever $\{A_k\}$ are pairwise disjoint elements of $\mathcal{A}$ and $A = \bigcup_{k \geq 1} A_k \in \mathcal{A}$, then

$$\lim_{n\to\infty} E\Big[\Big(U(A) - \sum_{k=1}^{n} U(A_k)\Big)^2\Big] = 0. \tag{9.8}$$

An L^2-valued measure U is *σ-finite* if there exists a sequence $\{B_n\}_{n\in\mathbf{N}} \subseteq \mathcal{A}$ such that $\mathbf{R}^d = \bigcup B_n$ and

$$\forall n \in \mathbf{N}: \; \sup\{E[U(A)^2] : A \in \mathcal{A},\, A \subseteq B_n\} < \infty.$$

White noise on a Borel subset of a Euclidean space with respect to Lebesgue measure m is an obvious example of a σ-finite L^2-valued measure. The ring $\mathcal{A}$ could be the collection of bounded Borel sets, or the collection of sets with finite Lebesgue measure.

Definition 9.1. A process $\{M_t(A) : t \in \mathbf{R}_+, A \in \mathcal{A}\}$ is a martingale measure if the following conditions are satisfied.

(i) For each $t > 0$, $M_t(\cdot)$ is a σ-finite L^2-valued measure on $\mathcal{A}$.

(ii) For each $A \in \mathcal{A}$, $M_t(A)$ is a martingale with respect to $\{\mathcal{F}_t\}$ with initial value $M_0(A) = 0$.

We get a martingale measure out of white noise as follows. Let W be white noise based on $\mathbf{R}_+ \times \mathbf{R}^d$ with Lebesgue measure, defined on a complete probability space $(\Omega, \mathcal{F}, P)$. Define $W_0(A) = 0$ and $W_t(A) = W((0,t] \times A)$ for $0 < t < \infty$ and bounded Borel sets $A \subseteq \mathbf{R}^d$. The filtration $\{\mathcal{F}_t\}$ is defined by augmenting $\mathcal{F}_t^W = \sigma\{W(D) : D \in \mathcal{B}_{(0,t]\times\mathbf{R}^d} \text{ is bounded}\}$. The martingale property comes from independent increments: for $0 \leq s < t$,

$$W_t(A) = W_s(A) + W((s,t] \times A)$$

where $W_s(A)$ is $\mathcal{F}_s$-measurable and $W((s,t] \times A)$ is independent of $\mathcal{F}_s$ (Exercise 9.3). Thus

$$\begin{aligned} E[W_t(A) \,|\, \mathcal{F}_s] &= E[W_s(A) \,|\, \mathcal{F}_s] + E[\, W((s,t] \times A) \,|\, \mathcal{F}_s] \\ &= W_s(A) + E[\, W((s,t] \times A)\,] = W_s(A). \end{aligned} \tag{9.9}$$

For $s < t$ the random variable $W((s,t] \times A)$ has $\mathcal{N}(0, (t-s)m(A))$ distribution, and so Exercise 1.12 implies that

$$E|W_t(A) - W_s(A)|^p = E|W((s,t] \times A)|^p = C_p m(A)^{p/2} |t-s|^{p/2}. \tag{9.10}$$

Together with Theorem B.20 this moment calculation shows that, for each bounded set A, there exists a continuous version of the martingale $W_{\cdot}(A)$.

A particularly useful class of martingale measures are the orthogonal martingale measures. By definition, a martingale measure is *orthogonal* if the product $M_t(A)M_t(B)$ is a martingale whenever $A \cap B = \emptyset$. This is equivalent to saying that the covariations $[M(A), M(B)]_t$ and $\langle M(A), M(B)\rangle_t$ vanish for disjoint A and B. That white noise is an orthogonal martingale measure follows from the independence of the martingales $W_{\cdot}(A)$ and $W_{\cdot}(B)$ for disjoint A and B (Exercise 9.9).

The goal is now to develop a stochastic integral of a time-space indexed process $Y(s, x, \omega)$ with respect to white noise. Alternative notations for this integral are

$$(Y \cdot W)_t(A) = \int_{(0,t] \times A} Y(s,x)\, W(ds, dx) = \int_{(0,t] \times A} Y(s,x)\, dW(s,x) \tag{9.11}$$

where $t \in \mathbf{R}_+$ is the time parameter and $A \subseteq \mathbf{R}^d$ a Borel subset. As the notation indicates, the integral itself is also a martingale measure. For a fixed set A it will be continuous t. The development parallels closely the development of the Itô integral $\int_0^t Y\, dB$ in Chapter 4, with the difference that there is now an additional spatial variable to keep track of.

The *covariance functional* of a martingale measure M is defined by

$$\bar{Q}_t(A, B) = \langle M(A), M(B)\rangle_t. \tag{9.12}$$

As Exercise 9.9 shows, for white noise this is $\bar{Q}_t(A,B) = tm(A \cap B)$. Define the *covariance measure* Q of white noise on $\mathbf{R}_+ \times \mathbf{R}^d \times \mathbf{R}^d$ (time $\times$ space $\times$ space) by

$$\int_{\mathbf{R}_+ \times \mathbf{R}^d \times \mathbf{R}^d} \varphi(s,x,y)\, Q(ds, dx, dy) = \int_{\mathbf{R}_+ \times \mathbf{R}^d} \varphi(s,x,x)\, ds\, dx. \tag{9.13}$$

The covariance measure extends the definition of the covariance functional $\bar{Q}$ in the sense that $\bar{Q}_t(A,B) = Q((0,t] \times A \times B)$.

The special pleasant property of white noise is that $\bar{Q}$ extends effortlessly to a deterministic nonnegative measure Q. This is analogous to the simple form of the Doléans measure of Brownian motion (Example 5.2). For general martingale measures the covariance functional $\bar{Q}$ is random and takes both signs. To extend it to a (random) signed measure Q one assumes the existence of a dominating measure. Martingale measures that satisfy this condition are called *worthy*. We refer to Walsh's lectures [17] for the general theory.

It is also convenient that Q is supported on $\mathbf{R}_+ \times \Delta(\mathbf{R}^d)$ where $\Delta(\mathbf{R}^d) = \{(x,x) : x \in \mathbf{R}^d\}$ is the diagonal of $\mathbf{R}^d \times \mathbf{R}^d$. This feature is common to all orthogonal martingale measures. However, we do not need the covariance measure for anything in the sequel, because we only discuss white noise and things stay simple.

Our integrands will be real-valued functions on $\mathbf{R}_+ \times \mathbf{R}^d \times \Omega$ (time $\times$ space $\times$ probability space). Measurability of a function $f : \mathbf{R}_+ \times \mathbf{R}^d \times \Omega \to \mathbf{R}$ will mean measurability with respect to the completion of $\mathcal{B}_{\mathbf{R}_+} \otimes \mathcal{B}_{\mathbf{R}^d} \otimes \mathcal{F}$ under $m \otimes m \otimes P$. The L^2-norm of such a function over the set $[0,T] \times \mathbf{R}^d \times \Omega$ is

$$\|f\|_{2,T} = \left(E \int_{[0,T]\times\mathbf{R}^d} |f(t,x,\omega)|^2 dt\, dx \right)^{1/2}. \tag{9.14}$$

We also need a notion of adaptedness. Begin with the filtration $\{\mathcal{B}_{\mathbf{R}^d} \otimes \mathcal{F}_t\}_{t\in\mathbf{R}_+}$ on the space $\mathbf{R}^d \times \Omega$, and let $\{\mathcal{G}_t\}_{t\in\mathbf{R}_+}$ be its augmentation under the measure $m \otimes P$, as defined in Section 2.1. Call a measurable function $f : \mathbf{R}_+ \times \mathbf{R}^d \times \Omega \to \mathbf{R}$ an *adapted process* if $(x,\omega) \mapsto f(t,x,\omega)$ is $\mathcal{G}_t$-measurable for each $t \in \mathbf{R}_+$. Let $\mathcal{L}_2(W)$ be the space of such adapted processes for which $\|f\|_{2,T} < \infty$ for all $T < \infty$. Analogously with our development in Chapters 4 and 5, we define a metric on $\mathcal{L}_2(W)$ by

$$d_{\mathcal{L}_2(W)}(X,Y) = \sum_{k=1}^{\infty} 2^{-k}\big(1 \wedge \|X-Y\|_{2,k}\big). \tag{9.15}$$

A special class of adapted processes are *simple predictable processes*

$$Y(t,x,\omega) = \sum_{i=1}^{n} \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t)\mathbf{1}_{A_i}(x) \tag{9.16}$$

where n is a finite integer, A_i is a bounded Borel subset of $\mathbf{R}^d$, $0 \le t_i < t_{i+1}$, ξ_i is a bounded $\mathcal{F}_{t_i}$-measurable random variable on $(\Omega,\mathcal{F},P)$, and the sets $\{(t_i,t_{i+1}] \times A_i\}$ are pairwise disjoint. Let $\mathcal{S}_2$ denote the space of simple predictable processes.

We record the approximation step first.

Proposition 9.2. *Suppose $Y \in \mathcal{L}_2(W)$. Then there exists a sequence of simple predictable processes $\{Y_n\}$ such that $d_{\mathcal{L}_2(W)}(Y,Y_n) \to 0$.*

Proof. As in earlier similar proofs (Lemmas 4.2 and 5.10), this goes in stages. It suffices to consider a fixed T and find simple predictable $\{Y_n\}$ such that $\|Y - Y_n\|_{2,T} \to 0$. Next, we may assume Y bounded and for the x-coordinate supported on a fixed compact set $K \subseteq \mathbf{R}^d$. That is, there exists a constant $C < \infty$ such that $|Y(t,x,\omega)| \le C$ for all (t,x,ω), and $Y(t,x,\omega) = 0$ unless $x \in K$. This follows from the dominated convergence theorem: if

$$Y^{(m)}(t,x,\omega) = [(m \wedge Y(t,x,\omega)) \vee (-m)] \cdot \mathbf{1}_{[-m,m]^d}(x)$$

then $\|Y - Y^{(m)}\|_{2,T} \to 0$ as $m \to \infty$.

For each $n \in \mathbf{N}$ and $s \in [0,1]$, define

$$Z^{n,s}(t,x,\omega) = \sum_{j \in \mathbf{Z}} Y(s + 2^{-n}j, x, \omega) \mathbf{1}_{(s+2^{-n}j,\, s+2^{-n}(j+1)]}(t) \cdot \mathbf{1}_{(0,T]}(t).$$

$Z^{n,s}$ is jointly measurable as a function of (s,t,x,ω) and it is an adapted process, but it is not a simple predictable process due to the nature of the dependence on x. Thus after approximating Y with a sequence of $Z^{n,s}$ processes, we have to approximate $Z^{n,s}$ with a simple predictable process. The formula above involves $Y(t,x,\omega)$ for $t < 0$, so we extend Y to negative times by setting it equal to zero. The proof of Lemma 4.2 shows that

$$\lim_{n\to\infty} E \int_{[0,T]\times K} dt\, dx \int_0^1 ds \left| Z^{n,s}(t,x,\omega) - Y(t,x,\omega) \right|^2 = 0. \tag{9.17}$$

This proof can be followed almost word for word: simply add the x-variables inside the functions and replace integrals $\int_0^T dt$ with $\int_{[0,T]\times K} dt\, dx$.

Then, as justified in that proof, there exists a fixed s and a subsequence $Z^{n_k,s}$ such that

$$\lim_{k\to\infty} E \int_{[0,T]\times K} dt\, dx \left| Z^{n_k,s}(t,x,\omega) - Y(t,x,\omega) \right|^2 = 0. \tag{9.18}$$

As the last step we show that, given fixed (n,s) and $\delta > 0$, we can find a simple predictable process $h^{n,s}$ such that

$$E \int_{[0,T]\times K} dt\, dx \left| Z^{n,s}(t,x,\omega) - h^{n,s}(t,x,\omega) \right|^2 < \delta. \tag{9.19}$$

For fixed (n,s,j) consider the $\mathcal{G}_{s+2^{-n}j}$-measurable, bounded L^2 function $(x,\omega) \mapsto Y(s+2^{-n}j, x, \omega)$ on $K \times \Omega$. Given $\delta_0 > 0$, we can find a simple function

$$g^{n,s,j}(x,\omega) = \sum_{i=1}^{\ell} c_i \mathbf{1}_{G_i}(x,\omega), \tag{9.20}$$

where the $\{G_i\}$ are $\mathcal{G}_{s+2^{-n}j}$-measurable subsets of $K \times \Omega$, such that

$$E \int_K |Y(s+2^{-n}j, x, \omega) - g^{n,s,j}(x,\omega)|^2 \, dx < \delta_0.$$

Product sets $\{B \times U : B \in \mathcal{B}_K, U \in \mathcal{F}_{s+2^{-n}j}\}$ form a semialgebra that generates the product σ-algebra $\mathcal{B}_K \otimes \mathcal{F}_{s+2^{-n}j}$. Each set G_i is a.e. equal to a set in $\mathcal{B}_K \otimes \mathcal{F}_{s+2^{-n}j}$. By a combination of Lemmas B.1 and B.2, each G_i in (9.20) can be approximated by a finite disjoint union $\bar{G}_i = \bigcup_j B_{i,j} \times U_{i,j}$ such that, for a given $\delta_1 > 0$,

$$E \int_K |\mathbf{1}_{G_i}(x,\omega) - \mathbf{1}_{\bar{G}_i}(x,\omega)|\, dx < \delta_1.$$

Put these together, choose δ_1 small enough, and rename the indices to create a function $\bar{g}^{n,s,j}(x,\omega) = \sum_{i=1}^m \mathbf{1}_{B_i}(x)\xi_i(\omega)$ such that

$$E \int_K |g^{n,s,j}(x,\omega) - \bar{g}^{n,s,j}(x,\omega)|^2\, dx < \delta_0.$$

Each B_i is a Borel subset of K, hence a bounded set. Each $\xi_i(\omega)$ is a bounded $\mathcal{F}_{s+2^{-n}j}$-measurable function and in fact of the form $c\mathbf{1}_U(\omega)$.

From these ingredients we construct the simple predictable process

$$h^{n,s}(t,x,\omega) = \sum_{j\in\mathbf{Z}} \bar{g}^{n,s,j}(x,\omega)\mathbf{1}_{(s+2^{-n}j,s+2^{-n}(j+1)]}(t) \cdot \mathbf{1}_{(0,T]}(t).$$

Inequality (9.19) can be satisfied by choosing δ_0 small enough □

Now we define the stochastic integral, beginning with simple predictable processes. For Y as in (5.6), the stochastic integral $Y \cdot W$ is the martingale measure defined, for $0 \le t < \infty$ and $B \in \mathcal{B}_{\mathbf{R}^d}$, by

$$(9.21)\qquad \begin{aligned} (Y \cdot W)_t(B) &= \sum_{i=1}^n \xi_i(\omega)\big(W_{t\wedge t_{i+1}}(A_i \cap B) - W_{t\wedge t_i}(A_i \cap B)\big) \\ &= \sum_{i=1}^n \xi_i(\omega) W\big((t\wedge t_i, t\wedge t_{i+1}] \times (A_i \cap B)\big). \end{aligned}$$

Note that since the A_i are by assumption bounded Borel sets, we do not need to impose that restriction on B. As in Lemma 4.4 or Lemma 5.8, at this point we need to check that the integral $Y \cdot W$ is linear in Y and does not depend on the particular representation chosen for Y. We leave these as exercises.

Lemma 9.3. *Let Y be a simple predictable process as in* (5.6) *and $B, G, H \subseteq \mathbf{R}^d$ be Borel sets.*

(i) *$Y \cdot W$ is a finite martingale measure. For each Borel set B there exists a t-continuous version of $(Y \cdot W)_t(B)$.*

(ii) *The process*

$$(9.22)\qquad Z_t = (Y \cdot W)_t(G) \cdot (Y \cdot W)_t(H) - \int_{(0,t]\times(G\cap H)} Y(s,x)^2\, ds\, dx$$

is a martingale. (The middle dot above is ordinary multiplication.)

(iii) *We have this isometry and bound:*

$$E[(Y\cdot W)_t(G)^2] = E\int_{(0,t]\times G} Y(s,x,\omega)^2\,ds\,dx = \|Y\|^2_{L^2((0,t]\times G\times\Omega)} \le \|Y\|^2_{2,t}. \tag{9.23}$$

Proof. (i) The martingale property of the process $(Y\cdot W)_t(B)$ on the right-hand side of (9.21) is checked, term by term, exactly as in the proof of Lemma 5.9 in Chapter 5. Continuity in t is evident from formula (9.21) if we choose continuous versions of the white noise martingales, as we can do by the observation around (9.10).

That $(Y\cdot W)_t$ is a finite L^2-valued measure on $\mathbf{R}^d$ follows from the corresponding properties of W. Let $B=\bigcup B_k$ be a disjoint union of Borel subsets of $\mathbf{R}^d$. It suffices to consider a single term $\xi\mathbf{1}_{(u,v]}\mathbf{1}_A$ of the type that appears in (5.6), with $0\le u<v<\infty$, a fixed bounded Borel set A, and bounded $\mathcal{F}_u$-measurable ξ. The L^2 convergence

$$\sum_{k=1}^m \xi W((t\wedge u,t\wedge v]\times(A\cap B_k)) \to \xi W((t\wedge u,t\wedge v]\times(A\cap B))$$

and the bound

$$E\big[\xi^2 W((t\wedge u,t\wedge v]\times(A\cap B))^2\big] \le E(\xi^2)(v-u)m(A)$$

uniformly in B and t follow from properties of W, including the independence of ξ and $W((t\wedge u,t\wedge v]\times(A\cap B))$ if $t\ge u$. (If $t<u$ we simply have $\xi W(\emptyset)=0$.)

For the subsequent parts we separate two calculations as lemmas.

Lemma 9.4. *Consider two simple predictable processes as in* (5.6) *that consist of a single term:*

$$f(s,x,\omega)=\eta(\omega)\mathbf{1}_{(a,b]}(s)\mathbf{1}_A(x) \quad\text{and}\quad g(s,x,\omega)=\xi(\omega)\mathbf{1}_{(u,v]}(s)\mathbf{1}_D(x).$$

Assume that one of these cases holds:

$$\begin{aligned}&\text{(i)}\ (a,b]=(u,v]\ \text{and}\ A\cap D=\emptyset,\\ &\text{(ii)}\ (a,b]\cap(u,v]=\emptyset.\end{aligned} \tag{9.24}$$

Then for any Borel sets $G,H\subseteq\mathbf{R}^d$ the product $(f\cdot W)_t(G)\cdot(g\cdot W)_t(H)$ is a martingale. (Note indeed that the middle $\cdot$ is an ordinary product of real numbers.)

Proof. The task is to check that, for $s<t$,

$$\begin{aligned}&E\big[\eta\xi\, W((t\wedge a,t\wedge b]\times(A\cap G))\,W((t\wedge u,t\wedge v]\times(D\cap H))\,\big|\,\mathcal{F}_s\big]\\ &=\eta\xi\, W((s\wedge a,s\wedge b]\times(A\cap G))\,W((s\wedge u,s\wedge v]\times(D\cap H)).\end{aligned} \tag{9.25}$$

This is fairly straightforward, keeping in mind the independence properties of white noise. The location of s splits the calculations into cases. For example, if $s > a$ then decompose the white noise values into sums according to

$$(t \wedge a, t \wedge b] = (a, s \wedge b] \cup (s \wedge b, t \wedge b]. \tag{9.26}$$

Furthermore, if $s \geq b$ then the second piece is empty. We leave the details as Exercise 9.11. □

Lemma 9.5. *With notation as in Lemma 9.4, for $s < t$*

$$\begin{aligned} &E\big[(f \cdot W)_t(G) \cdot (f \cdot W)_t(H) \,|\, \mathcal{F}_s\big] \\ &\quad = E(\eta^2 \,|\, \mathcal{F}_s)(t \wedge b - t \wedge a) m(A \cap G \cap H) \\ &\quad + (f \cdot W)_s(G) \cdot (f \cdot W)_s(H) \;-\; \eta^2 (s \wedge b - s \wedge a) m(A \cap G \cap H). \end{aligned} \tag{9.27}$$

Proof. Suppose first $s \leq a$. Then on the right-hand side only the first term is nonzero. If also $t \leq a$ then everything reduces to zero. If $s \leq a < t$, condition on $\mathcal{F}_a$ on the left-hand side and use independence to get

$$\begin{aligned} &E\Big[\eta^2\, E\big\{W((a, t \wedge b] \times (A \cap G))\, W((a, t \wedge b] \times (A \cap H)) \,\big|\, \mathcal{F}_a\big\} \,\Big|\, \mathcal{F}_s\Big] \\ &= E(\eta^2 \,|\, \mathcal{F}_s)(t \wedge b - t \wedge a) m(A \cap G \cap H). \end{aligned}$$

If $s > a$, split the white noise terms according to (9.26) and use independence and the covariation of white noise to deduce

$$\begin{aligned} &E\big[(f \cdot W)_t(G) \cdot (f \cdot W)_t(H) \,|\, \mathcal{F}_s\big] \\ &= \eta^2\, E\big[W((a, t \wedge b] \times (A \cap G))\, W((a, t \wedge b] \times (A \cap H)) \,\big|\, \mathcal{F}_s\big] \\ &= \eta^2\, W((a, s \wedge b] \times (A \cap G))\, W((a, s \wedge b] \times (A \cap H)) \\ &\qquad + \eta^2 (t \wedge b - s \wedge b) m(A \cap G \cap H) \\ &= (f \cdot W)_s(G) \cdot (f \cdot W)_s(H) \\ &\qquad + E(\eta^2 \,|\, \mathcal{F}_s)\, (t \wedge b - t \wedge a) m(A \cap G \cap H) \\ &\qquad - \eta^2 (s \wedge b - s \wedge a) m(A \cap G \cap H). \end{aligned}$$

In the last equality we simply identified the first term as $(f \cdot W)_s(G) \cdot (f \cdot W)_s(H)$, used $\eta^2 = E(\eta^2 \,|\, \mathcal{F}_s)$ for $a < s$, and added $0 = s \wedge a - t \wedge a$. □

We return to prove part (ii) of Lemma 9.3. To use Lemmas 9.4 and 9.5, observe first that any simple predictable process Y as in (5.6) can be written so that for each $i < j$, either

$$\begin{aligned} &\text{(i) } (t_i, t_{i+1}] = (t_j, t_{j+1}] \text{ and } A_i \cap A_j = \emptyset, \text{ or} \\ &\text{(ii) } t_{i+1} \leq t_j. \end{aligned} \tag{9.28}$$

To see this, suppose that initially

$$Y(t,x,\omega) = \sum_i \tilde{\xi}_i(\omega)\mathbf{1}_{(\tilde{t}_i,\tilde{t}_{i+1}]}(t)\mathbf{1}_{\tilde{A}_i}(x)$$

with $\{(\tilde{t}_i,\tilde{t}_{i+1}]\times\tilde{A}_i\}$ pairwise disjoint. Let $\{t_1 < \cdots < t_N\} = \{\tilde{t}_i\}$ be the ordered set of time points that appear in the sum, and reorder the sum:

$$Y(t,x,\omega) = \sum_{k=1}^{N} \mathbf{1}_{(t_k,t_{k+1}]}(t) \sum_{i:(\tilde{t}_i,\tilde{t}_{i+1}]\supset(t_k,t_{k+1}]} \tilde{\xi}_i(\omega)\mathbf{1}_{(\tilde{t}_i,\tilde{t}_{i+1}]}(t)\mathbf{1}_{\tilde{A}_i}(x).$$

Note that if $(\tilde{t}_i,\tilde{t}_{i+1}]$ and $(\tilde{t}_j,\tilde{t}_{j+1}]$ both contain $(t_k,t_{k+1}]$, the disjointness of $(\tilde{t}_i,\tilde{t}_{i+1}]\times\tilde{A}_i$ and $(\tilde{t}_j,\tilde{t}_{j+1}]\times\tilde{A}_j$ forces $\tilde{A}_i\cap\tilde{A}_j=\emptyset$. Relabel everything once more to get the representation

$$Y(t,x,\omega) = \sum_{i=1}^{n} \xi_i(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(t)\mathbf{1}_{A_i}(x)$$

subject to (9.28).

To prove that Z_t of (9.22) is a martingale, take $s<t$ and compute $E\big[(Y\cdot W)_t(G)\cdot(Y\cdot W)_t(H)\,|\,\mathcal{F}_s\big]$ by multiplying out inside the expectation and applying Lemmas 9.4 and 9.5 to the terms. This results in the martingale property of the following process:

$$\widetilde{Z}_t = (Y\cdot W)_t(G)\cdot(Y\cdot W)_t(H) - \sum_{i=1}^{n} \xi_i^2(t\wedge t_{i+1} - t\wedge t_i)\, m(A_i\cap G\cap H) \tag{9.29}$$

Now observe that the last sum is $\int_{(0,t]\times(G\cap H)} Y(s,x)^2\,ds\,dx$ because, by the disjointness of the sets $(t_i,t_{i+1}]\times A_i$,

$$Y(s,x,\omega)^2 = \sum_{i=1}^{n} \xi_i^2(\omega)\mathbf{1}_{(t_i,t_{i+1}]}(s)\mathbf{1}_{A_i}(x).$$

Thus $\widetilde{Z}_t = Z_t$ and we have checked part (ii).

Part (iii) follows by taking $G=H$ and $EZ_t=0$ in part (ii) □

We can now state the definition of the integral of $\mathcal{L}_2(W)$-integrands with respect to white noise.

Definition 9.6. Let W_t be a white noise martingale measure on $\mathbf{R}^d$ defined on a probability space $(\Omega,\mathcal{F},P)$ with respect to a complete filtration $\{\mathcal{F}_t\}$. For any measurable adapted process $Y\in\mathcal{L}_2(W)$, the *stochastic integral* $Y\cdot W$ is the finite martingale measure that satisfies, for any Borel set $B\subseteq\mathbf{R}^d$,

$$\lim_{n\to\infty} \|(Y\cdot W)_{\boldsymbol{\cdot}}(B) - (Y_n\cdot W)_{\boldsymbol{\cdot}}(B)\|_{\mathcal{M}_2^c} = 0 \tag{9.30}$$

for any sequence $Y_n \in \mathcal{S}_2$ of simple predictable processes such that $d_{\mathcal{L}_2(W)}(Y_n, Y) \to 0$. The process $Y \cdot W$ is unique in the sense that, for each B, up to indistinguishability the limit (9.30) identifies a unique martingale $(Y \cdot W)_{\bullet}(B)$.

Justification of the definition. For any fixed Borel $B \subseteq \mathbf{R}^d$, $M^{(n)} = (Y_n \cdot W)_{\bullet}(B)$ is a Cauchy sequence in the space $\mathcal{M}_2^c$ of continuous L^2 martingales by virtue of the estimate

$$\|M_t^{(m)} - M_t^{(n)}\|_{L^2(P)} = E[((Y_m - Y_n) \cdot W)_t(B)^2] \le \|Y_m - Y_n\|_{2,t}.$$

By completeness (Theorem 3.41) there exists a continuous limit martingale that we call $(Y \cdot W)_{\bullet}(B)$.

In general on a measure space, if $f_n \to f$ and $g_n \to g$ in L^2, then $f_n g_n \to fg$ in L^1. This applied to the two terms below separately implies the L^1 convergence

$$\begin{aligned} &(Y_n \cdot W)_t(G) \cdot (Y_n \cdot W)_t(H) - \int_{(0,t]\times(G\cap H)} Y_n(s,x)^2\,ds\,dx \\ &\quad\longrightarrow (Y \cdot W)_t(G) \cdot (Y \cdot W)_t(H) - \int_{(0,t]\times(G\cap H)} Y(s,x)^2\,ds\,dx. \end{aligned} \tag{9.31}$$

Consequently the limit process is also a martingale. Note that for ω in the full probability event

$$\Bigl\{\omega : \int_{(0,T]\times\mathbf{R}^d} Y(s,x,\omega)^2\,ds\,dx < \infty \ \forall T < \infty\Bigr\}$$

the function $t \mapsto \int_{(0,t]\times(G\cap H)} Y(s,x,\omega)^2\,ds\,dx$ is continuous simply by virtue of the dominated convergence theorem. Hence the limit martingale in (9.31) has a.s. continuous paths.

To conclude that the limit martingales $(Y \cdot W)_{\bullet}(B)$ form a finite martingale measure, it remains to check that for a fixed t, $B \mapsto (Y \cdot W)_t(B)$ is a finite L^2-valued measure. For a finite disjoint union $B = B_1 \cup \cdots \cup B_m$, additivity

$$(Y \cdot W)_t(B) = (Y \cdot W)_t(B_1) + \cdots + (Y \cdot W)_t(B_m)$$

is inherited from the additivity of the approximating variables $(Y_n \cdot W)_t(B)$. Let $B = \bigcup_{k\ge1} B_k$ be a countable disjoint union and $G_m = B \smallsetminus (B_1 \cup \cdots \cup B_m)$. The required property is $(Y \cdot W)_t(G_m) \to 0$ as $m \to \infty$ in L^2. It follows from the isometry and dominated convergence:

$$E[(Y \cdot W)_t(G_m)^2] = E\int_{(0,t]\times G_m} Y(s,x,\omega)^2\,ds\,dx \to 0.$$

Similarly we have the bound over all Borel subsets:

$$\sup_{B\subseteq\mathbf{R}^d} E[(Y\cdot W)_t(B)^2] \le E\int_{(0,t]\times\mathbf{R}^d} Y(s,x,\omega)^2\,ds\,dx < \infty.$$

The uniqueness claim is proved from estimate (9.23) as was done for Definition 5.11. ☐

In (9.11) we gave alternative notations for the integral. The martingale property of the process in (9.31) can be equivalently expressed in terms of the predictable bracket process as

$$\left\langle \int_{(0,\cdot]\times G} Y\,dW\,,\ \int_{(0,\cdot]\times H} Y\,dW \right\rangle_t = \int\limits_{(0,t]\times(G\cap H)} Y(s,x)^2\,ds\,dx. \tag{9.32}$$

For the next section on a stochastic heat equation we need an application of the Burkholder-Davis-Gundy inequality that we state as the next lemma.

Lemma 9.7. *Assumptions as in Definition 9.6. Let $B\in\mathcal{B}_{\mathbf{R}}$ and $p\in[2,\infty)$. Then there exists a constant $C_p<\infty$ that does not depend on either Y or B such that, for all $T\in[0,\infty)$,*

$$\begin{aligned} &E\left[\sup_{t\in[0,T]}\left|\int_{(0,t]\times B} Y(s,x)\,dW(s,x)\right|^p\right] \\ &\qquad\le C_p E\left[\left(\int_{(0,T]\times B} Y(s,x)^2\,ds\,dx\right)^{\frac{p}{2}}\right]. \end{aligned}\tag{9.33}$$

Proof. Immediate application of (6.28) to the martingale $(Y\cdot W)_{\boldsymbol{\cdot}}(B)$, together with the fact that $[M]=\langle M\rangle$ for a continuous L^2 martingale M (Proposition 3.34). ☐

9.2. Stochastic heat equation

In Chapter 7 we took a basic ODE $x'(t) = b(x(t))$, added noise to it, and studied the resulting SDE $dX = b(X)dt + \sigma(X)dB$. Here we do the same for a basic partial differential equation (PDE), namely the heat equation. We begin by introducing the heat equation in one space dimension.

The *heat equation* on the real line is the PDE

$$\rho_t = \tfrac{1}{2}\rho_{xx} \tag{9.34}$$

for an unknown real-valued function $\rho(t,x)$ on $\mathbf{R}_+ \times \mathbf{R}$. Subscripts are shorthands for partial derivatives: $\rho_t = \partial\rho/\partial t$ and $\rho_{xx} = \partial^2\rho/\partial x^2$. This equation is intimately tied to Brownian motion because of the role of the Gaussian kernel $p(t,x,y) = (2\pi t)^{-1/2}\exp[\frac{1}{2t}(x-y)^2]$ as the fundamental solution of the heat equation. Namely, thinking of y fixed, for $t > 0$ the function $\rho(t,x) = p(t,x,y)$ satisfies (9.34), and, as $t \searrow 0$, as a probability measure $\rho(t,x)dx$ converges weakly to δ_y, the pointmass at y. Hence for $t = 0$ it is natural to use the convention

$$\int_{\mathbf{R}} p(0,x,y)f(y)\,dy = \int_{\mathbf{R}} p(0,y,x)f(y)\,dy = f(x). \tag{9.35}$$

Consider the initial value problem

$$\rho_t = \tfrac{1}{2}\rho_{xx} \text{ on } (0,\infty)\times\mathbf{R}, \quad \rho(0,x) = f(x) \tag{9.36}$$

for the heat equation, with a given bounded continuous initial function f. Define the convolution

$$\rho(t,x) = (p(t) * f)(x) = \int_{\mathbf{R}} p(t,x,y)f(y)\,dy. \tag{9.37}$$

Then ρ is infinitely differentiable on $(0,\infty)\times\mathbf{R}$, continuous on $\mathbf{R}_+\times\mathbf{R}$, and solves (9.36). This solution ρ is unique among solutions that satisfy an exponential growth bound in space. (See Thm. 7 in Sect. 2.3 in [7]).

We add space-time white noise to the heat equation to get the *stochastic heat equation*

$$u_t = \tfrac{1}{2}u_{xx} + \sigma(u)\dot{W}. \tag{9.38}$$

By allowing the coefficient function σ we give ourselves some additional generality. Equation (9.38) is an example of a stochastic partial differential equation (SPDE). The solution u is a function $u(t,x,\omega)$ of time, space, and a sample point ω of the probability space on which the white noise W is defined. If the coefficient function σ is a constant, (9.38) is quite simple. When $\sigma(u)$ is not identically 0 or 1 equation (9.38) is sometimes called the stochastic heat equation with multiplicative noise.

The setting is now more complex than for SDEs. As with SDEs, we cannot expect our solution process u to be actually differentiable in the variables t and x. We first embark on a discussion of how to make rigorous mathematical sense of equation (9.38).

In the theory of partial differential equations there is a standard way of dealing with equations where derivatives cannot immediately be expected to exist in the classical (that is, calculus) sense. The original equation is replaced by an integral equation via these steps: (i) multiply the original equation by a smooth compactly supported test function $\psi \in C_c^\infty(\mathbf{R}_+ \times \mathbf{R})$, (ii) integrate over $[0,t] \times \mathbf{R}$, and (iii) move all derivatives onto ψ through a formal integration by parts. For (9.38) steps (i) and (ii) give

$$\int_{[0,t]\times\mathbf{R}} \psi(s,x)\, u_t(s,x)\, ds\, dx = \tfrac{1}{2} \int_{[0,t]\times\mathbf{R}} \psi(s,x) u_{xx}(s,x)\, ds\, dx + \int_{[0,t]\times\mathbf{R}} \psi(s,x)\, \sigma(u(s,x)) \dot{W}(s,x)\, ds\, dx.$$

Step (iii), and interpreting the integral against the "generalized function" $\dot{W}(s,x)$ as a white-noise stochastic integral, leads to the equation

$$\begin{aligned} &\int_{\mathbf{R}} \psi(t,x)\, u(t,x)\, dx - \int_{\mathbf{R}} \psi(0,x)\, u(0,x)\, dx - \int_{[0,t]\times\mathbf{R}} \psi_t(s,x)\, u(s,x)\, ds\, dx \\ &= \tfrac{1}{2} \int_{[0,t]\times\mathbf{R}} \psi_{xx}(s,x)\, u(s,x)\, ds\, dx + \int_{[0,t]\times\mathbf{R}} \psi(s,x)\, \sigma(u(s,x))\, dW(s,x). \end{aligned} \tag{9.39}$$

Compact support of ψ on $\mathbf{R}$ kills the boundary terms at $x = \pm\infty$. Note the nature of this activity: we are not "solving" (9.38) but rather working our way informally towards a reformulation of (9.38) that could conceivably be solved in a mathematically rigorous fashion. The virtue of (9.39) is that it makes sense for any measurable, adapted process $u(t,x,\omega)$ for which the last (stochastic) integral is well-defined. Compact support of ψ restricts this integral to a compact rectangle $[0,t] \times [-R,R]$ so under strict enough assumptions on σ this integral will present no problem. In p.d.e. terminology (9.39) is the *weak* form of equation (9.38).

Let us rewrite (9.39) once more to highlight the feature that ψ is acted on by a backward heat operator:

$$
\begin{aligned}
\int_{\mathbf{R}} \psi(t,x)\,u(t,x)\,dx \;&=\; \int_{\mathbf{R}} \psi(0,x)\,u(0,x)\,dx \\
&\quad + \int_{[0,t]\times\mathbf{R}} \big[\psi_t(s,x) + \tfrac{1}{2}\psi_{xx}(s,x)\big]\,u(s,x)\,ds\,dx \\
&\quad + \int_{[0,t]\times\mathbf{R}} \psi(s,x)\,\sigma(u(s,x))\,dW(s,x).
\end{aligned}
\tag{9.40}
$$

In (9.40) we have an equation for the unknown process u that is more meaningfully posed than (9.38). But (9.40) is not immediately conducive to the application of an iterative scheme for proving existence. So we set out to rewrite the equation once more, by making a judicious choice of the test function ψ. Since we have not made any precise assumptions yet, mathematical precision will continue to take a backseat.

Take a test function $\phi \in C_c^\infty(\mathbf{R})$ and define

$$
G(t,\phi,x) = \int_{\mathbf{R}} p(t,z,x)\phi(z)\,dz.
$$

Keeping $t > 0$ fixed, set $\psi(s,x) = G(t-s,\phi,x)$. Then $\psi(t,x) = \phi(x)$ and $\psi_t(s,x) + \frac{1}{2}\psi_{xx}(s,x) = 0$ while $s \in (0,t)$. With this ψ (9.40) turns into

$$
\begin{aligned}
\int_{\mathbf{R}} \phi(x)\,u(t,x)\,dx \;&=\; \int_{\mathbf{R}} G(t,\phi,x)\,u(0,x)\,dx \\
&\quad + \int_{[0,t]\times\mathbf{R}} G(t-s,\phi,x)\,\sigma(u(s,x))\,dW(s,x).
\end{aligned}
\tag{9.41}
$$

Now let ϕ shrink nicely down to a point mass at a particular x_0. For example, take $\phi_\varepsilon(x) = p(\varepsilon, x_0, x)$ and then let $\varepsilon \searrow 0$. Then

$$
G(t,\phi_\varepsilon,y) = \int p(t,x,y)\phi_\varepsilon(x)\,dx \to p(t,x_0,y).
$$

After the $\varepsilon \to 0$ limit equation (9.41) becomes

$$
\begin{aligned}
u(t,x_0) \;&=\; \int_{\mathbf{R}} p(t,x_0,x)\,u(0,x)\,dx \\
&\quad + \int_{[0,t]\times\mathbf{R}} p(t-s,x_0,x)\,\sigma(u(s,x))\,dW(s,x).
\end{aligned}
\tag{9.42}
$$

Of course taking the limit above rigorously would require some assumptions. But we will not worry about justifying the steps from (9.38) to (9.42). In fact, we haven't even given a rigorous definition for what (9.38) means. The fruitful thing to do is to take (9.42) as the definition of the solution to (9.38). Recall that $\{\mathcal{G}_t\}$ is the filtration on $\mathbf{R} \times \Omega$ obtained by augmenting $\{\mathcal{B}_{\mathbf{R}} \otimes \mathcal{F}_t\}$.

Definition 9.8. Let $u_0(x, \omega)$ be a given $\mathcal{G}_0$-measurable initial function. Let $u(t, x, \omega)$ be a measurable adapted process. Then u is a *mild solution* of the initial value problem

$$u_t = \tfrac{1}{2} u_{xx} + \sigma(u)\dot{W} \quad \text{on } (0,\infty) \times \mathbf{R}, \quad u(0,x) = u_0(x) \text{ for } x \in \mathbf{R} \tag{9.43}$$

if u satisfies the equation

$$\begin{aligned} u(t,x) \;&=\; \int_{\mathbf{R}} p(t,x,y)\, u_0(y)\, dy \\ &\qquad + \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\, \sigma(u(s,y))\, dW(s,y) \end{aligned} \tag{9.44}$$

in the sense that, for each $(t, x) \in (0, \infty) \times \mathbf{R}$, the equality holds a.s. Part of the definition is that the integrals above are well-defined. In particular, for each (t, x) the integrand $\xi^{t,x}(s, x, \omega) = \mathbf{1}_{[0,t)}(s)p(t - s, x, y)\, \sigma(u(s, y))$ is a member of $\mathcal{L}_2(W)$.

Note that the stochastic integral term on the right of (9.44) is the value $Y_t^{t,x}$ of the process

$$Y_r^{t,x} = \int_{[0,r]\times\mathbf{R}} \mathbf{1}_{[0,t)}(s)\, p(t-s,x,y)\, \sigma(u(s,y))\, dW(s,y), \qquad r \in \mathbf{R}_+$$

defined as in Definition 9.6. Exercise 9.13 asks you to verify that equation (9.44) implies the weak form (9.40).

Example 9.9. The case $\sigma(u) = 1$ is called the stochastic heat equation with additive noise:

$$u_t = \tfrac{1}{2} u_{xx} + \dot{W} \quad \text{on } (0,\infty) \times \mathbf{R}, \quad u(0,x) = u_0(x) \text{ for } x \in \mathbf{R}. \tag{9.45}$$

In this case (9.44) immediately gives us an explicit solution:

$$u(t,x) \;=\; \int_{\mathbf{R}} p(t,x,y)\, u_0(y)\, dy \;+\; \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\, dW(s,y). \tag{9.46}$$

In particular, the stochastic integral term has a deterministic L^2 integrand. This integral can be defined as an instance of the isonormal process in Section 9.1.

Returning to the general SPDE (9.43), our goal is to prove an existence and uniqueness theorem for the mild solution. The familiar Lipschitz assumption on the coefficient $\sigma(u)$ is again invoked.

Assumption 9.10. Assume the Borel function $\sigma : \mathbf{R} \to \mathbf{R}$ satisfies the Lipschitz condition

$$|\sigma(u) - \sigma(v)| \le L|u - v| \tag{9.47}$$

for a constant $L < \infty$ and all $u, v \in \mathbf{R}$.

Note that assumption (9.47) implies also the growth bound

$$|\sigma(u)| \le L'(1 + |u|) \tag{9.48}$$

for another constant L'.

We need to impose two types of assumptions on the initial function u_0, basically to make the two integrals on the right-hand side of (9.44) well defined. The first integral is an ordinary Lebesgue integal, evaluated ω by ω. It needs an assumption on the almost sure behavior of u_0. We also insert u_0 in the stochastic integral. This part would be well-defined with an L^p assumption with $p = 2$. In order to get a continuous solution process we strengthen this to $p > 8$.

Assumption 9.11. Let $u_0 : \mathbf{R} \times \Omega \to \mathbf{R}$ be a $\mathcal{G}_0$-measurable function with the following properties. For P-a.e. ω, $x \mapsto u_0(x, \omega)$ is a continuous function and there exists a constant $C(\omega) < \infty$ such that

$$|u_0(x, \omega)| \le C(\omega)e^{C(\omega)|x|} \quad \forall x \in \mathbf{R}. \tag{9.49}$$

Furthermore, for some $8 < p < \infty$ and constants $0 \le A, K < \infty$ we have the moment bound

$$E[|u_0(x)|^p] \le Ke^{A|x|} \quad \forall x \in \mathbf{R}. \tag{9.50}$$

Theorem 9.12. (a) Existence. *Assume Assumptions 9.10 and 9.11. Then there exists a continuous, measurable, adapted process $u(t, x, \omega)$ that satisfies Definition 9.8 of the mild solution of the stochastic heat equation (9.43). Furthermore, for each $T \in (0, \infty)$ there exists a constant $K(T) < \infty$ such that*

$$\forall t \in [0, T],\ x \in \mathbf{R} : \quad E[\,|u(t, x)|^p\,] \le K(T)e^{A|x|} \tag{9.51}$$

where A is the same constant as in assumption (9.50).

(b) Uniqueness. *Assume Assumption 9.10. Let u_0 be a random function such that the first integral on the right-hand side of (9.44) makes sense for a.e. ω. Suppose u and v are two measurable processes adapted to the filtration $\{\mathcal{G}_t\}$ defined above. Assume u and v satisfy moment bound (9.51)* *with $p = 2$ for some T, and equation (9.44) almost surely for each point $(t, x) \in [0, T] \times \mathbf{R}$. Then $u(t, x) = v(t, x)$ a.s. for each $(t, x) \in [0, T] \times \mathbf{R}$.*

At the heart of the existence proof is again an iteration. The scheme comes naturally from the definition (9.44). Let

$$u_0(t, x) = \int_{\mathbf{R}} p(t, x, y)\, u_0(y)\, dy \tag{9.52}$$

and for $n \geq 1$

$$\begin{aligned} u_n(t, x) \;&=\; \int_{\mathbf{R}} p(t, x, y)\, u_0(y)\, dy \\ &\quad + \int_{[0,t]\times\mathbf{R}} p(t - s, x, y)\, \sigma(u_{n-1}(s, y))\, dW(s, y). \end{aligned} \tag{9.53}$$

As it stands, $u_n(t, x)$ is defined for each (t, x) separately as an L^2 limit, according to the definition of the white noise integral from the previous section. This is not good enough. In order to insert u_n back into the stochastic integral for the next step of the iteration, u_n needs to have some regularity properties as a process indexed by (t, x). For this purpose we show that, at each stage, we can select a continuous version of the process u_n defined by the right-hand side of (9.53). We begin by sketching the simple argument for the first integral on the right-hand side of (9.44) and (9.53).

Lemma 9.13. *Let f be a continuous function on $\mathbf{R}$ and C a constant such that $|f(x)| \leq Ce^{C|x|}$ $\forall x \in \mathbf{R}$. Then the function*

$$v(t, x) = \int_{\mathbf{R}} p(t, x, y)\, f(y)\, dy$$

is continuous on $\mathbf{R}_+ \times \mathbf{R}$.

Proof. If $(t, x) \to (s, z)$ such that $s > 0$, it is straightforward to use dominated convergence to show $v(t, x) \to v(s, z)$. Suppose $(t, x) \to (0, z)$. Continuity holds along the line $t = 0$ by convention (9.35), so we may assume $t \to 0$ along positive values. Pick a small $\delta > 0$ and write

$$\begin{aligned} |v(t, x) - f(z)| \leq \int_{y:|y-z|<\delta} p(t, x, y)|f(y) - f(z)|\, dy \\ + \int_{y:|y-(x-z)|\geq\delta} \frac{e^{-y^2/t}}{\sqrt{2\pi t}} |f(y) - f(z)|\, dy. \end{aligned}$$

The first integral on the right can be made small by choice of δ, by the continuity of f. When $|x - z| < \delta/2$, the second integral can be bounded above by expanding the set over which the integral is taken to $\{y : |y| \geq \delta/2\}$. The derivative $(\partial/\partial t)p(t,0,y)$ shows that, for small enough t and all y outside $(-\delta/2, \delta/2)$, $p(t,0,y)$ decreases as $t \searrow 0$. Thus by dominated convergence the second intergral vanishes as $(t,x) \to (0,z)$. □

Before embarking on the real work we record some properties of the heat kernel. The following calculation will be used several times:

$$\text{(9.54)} \qquad \int_{[0,t]\times\mathbf{R}} p(s,x,y)^2\,ds\,dy = \int_0^t ds \int_{\mathbf{R}} \frac{e^{-y^2/s}}{2\pi s}\,dy = \sqrt{t/\pi}.$$

Note that shifting the integration variable y removes the variable x. Another useful formula is the Gaussian moment generating function

$$\text{(9.55)} \qquad \int_{\mathbf{R}} \frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} e^{\alpha y}\,dy = e^{\frac{1}{2}\sigma^2\alpha^2} \quad \text{for } \alpha \in \mathbf{R}.$$

Lemma 9.14. *There exists a constant C such that, for all $x \in \mathbf{R}$ and $0 < h, t < \infty$,*

$$\text{(9.56)} \qquad \int_{[0,t]\times\mathbf{R}} |p(s,x+h,y) - p(s,x,y)|^2\,ds\,dy \leq Ch.$$

and

$$\text{(9.57)} \qquad \int_{[0,t]\times\mathbf{R}} |p(s+h,x,y) - p(s,x,y)|^2\,ds\,dy \leq C\sqrt{h}.$$

Proof. Straightforward calculations. Expand the square and integrate the exponentials separately.

$$\begin{aligned}
&\int_{[0,t]\times\mathbf{R}} |p(s,x+h,y) - p(s,x,y)|^2\,ds\,dy \\
&= \int_0^t \frac{ds}{2\pi s} \int_{\mathbf{R}} \big(e^{-\frac{w^2}{s}} - 2e^{-\frac{w^2}{2s} - \frac{(w+h)^2}{2s}} + e^{-\frac{(w+h)^2}{s}}\big)\,dw \\
&= \int_0^t \frac{1}{\sqrt{\pi s}}(1 - e^{-\frac{h^2}{4s}})\,ds = \frac{h}{2\sqrt{\pi}} \int_{h^2/(4t)}^{\infty} u^{-3/2}(1 - e^{-u})\,du \\
&\leq \frac{h}{2\sqrt{\pi}} \int_0^{\infty} u^{-3/2}(1 - e^{-u})\,du = Ch.
\end{aligned}$$

Similarly

$$\begin{aligned}\int_{[0,t]\times\mathbf{R}} &|p(s+h,x,y)-p(s,x,y)|^2\,ds\,dy\\ &=\frac{1}{\sqrt{\pi}}\big(\sqrt{t+h}-\sqrt{t+h/2}+(\sqrt{2}-1)\sqrt{h}+\sqrt{t}-\sqrt{t+h/2}\big)\\ &\le C\sqrt{h}.\end{aligned}$$

□

We refine the previous lemma by inserting an exponentially growing factor inside the integral.

Lemma 9.15. *Let $0 < A, T < \infty$. There exists a constant $C = C(A,T) < \infty$ such that, for all $x \in \mathbf{R}$, $t \in [0,T]$ and $0 < h \le 1$,*

$$\int_{[0,t]\times\mathbf{R}} |p(s,x+h,y)-p(s,x,y)|^2\,e^{A|y|}\,ds\,dy \le Ce^{A|x|}h. \tag{9.58}$$

and

$$\int_{[0,t]\times\mathbf{R}} |p(s+h,x,y)-p(s,x,y)|^2\,e^{A|y|}\,ds\,dy \le Ce^{A|x|}\sqrt{h}. \tag{9.59}$$

Proof. Since $e^{A|y|} \le e^{Ay}+e^{-Ay}$, we can drop the absolute value from y by allowing A to be real, and then at the end add the two bounds for A and $-A$.

First some separate parts of the calculations. The next identity will find use several times later also.

$$\begin{aligned}\int_{[0,t]\times\mathbf{R}} p(s,x,y)^2\,e^{Ay}\,ds\,dy &= e^{Ax}\int_0^t \frac{ds}{2\sqrt{\pi s}}\int_{\mathbf{R}} \frac{e^{-y^2/s}}{\sqrt{\pi s}}\,e^{Ay}\,dy\\ &= e^{Ax}\int_0^t \frac{ds}{2\sqrt{\pi s}}\,e^{A^2 s/4}\end{aligned} \tag{9.60}$$

and

$$\begin{aligned}\int_{[0,t]\times\mathbf{R}} & p(s,x+h,y)p(s,x,y)\,e^{Ay}\,ds\,dy\\ &= e^{h^2/4s}\int_0^t \frac{ds}{2\sqrt{\pi s}}\int_{\mathbf{R}} \frac{e^{-\frac{1}{s}(y-x-\frac{h}{2})^2}}{\sqrt{\pi s}}\,e^{Ay}\,dy\\ &= e^{A(x+\frac{h}{2})}e^{h^2/4s}\int_0^t \frac{ds}{2\sqrt{\pi s}}\,e^{A^2 s/4}.\end{aligned} \tag{9.61}$$

Now expand the square and collect terms.

$$\int_{[0,t]\times\mathbf{R}} |p(s,x+h,y)-p(s,x,y)|^2\, e^{Ay}\,ds\,dy$$

(9.62)
$$= e^{Ax}\int_0^t \frac{ds}{2\sqrt{\pi s}} e^{A^2s/4}\Big(1-2e^{Ah/2}e^{h^2/4s}+e^{Ah}\Big)$$
$$= e^{Ax}\int_0^t \frac{ds}{2\sqrt{\pi s}} e^{A^2s/4}\Big((e^{Ah/2}-1)^2+2e^{Ah/2}(1-e^{h^2/4s})\Big)$$

Use the bound $e^{A^2s/4}\le e^{A^2t/4}$. Bound the first part inside the large parentheses with $|e^x-1|\le e^{(x\vee 0)}|x|$:

$$(e^{Ah/2}-1)^2\le e^{(A\vee 0)h}h^2.$$

Integrate the second part:

$$\int_0^t \frac{1}{\sqrt{\pi s}}(1-e^{-\frac{h^2}{4s}})\,ds = \frac{h}{2\sqrt{\pi}}\int_{h^2/(4t)}^\infty u^{-3/2}(1-e^{-u})\,du$$
$$\le \frac{h}{2\sqrt{\pi}}\int_0^\infty u^{-3/2}(1-e^{-u})\,du = Ch.$$

In the end we get

$$\int_{[0,t]\times\mathbf{R}} |p(s,x+h,y)-p(s,x,y)|^2\, e^{Ay}\,ds\,dy$$
$$\le Ce^{Ax}e^{A^2t/4}e^{(A\vee 0)h}(h^2\sqrt{t}+h)$$

where C is a numerical constant that does not depend on any of the parameters of the calculation. With h restricted to $(0,1]$ we have $h^2\le h$, and doubling this bound to account for $\pm A$ gives (9.58).

We proceed similarly for the second bound. First a piece of the calculation.

$$\int_{[0,t]\times\mathbf{R}} p(s+h,x,y)p(s,x,y)\, e^{Ay}\,ds\,dy$$

(9.63)
$$= \int_0^t \frac{ds}{2\sqrt{\pi(s+h/2)}}\int_{\mathbf{R}} \frac{e^{-(y-x)^2\frac{1}{2\frac{(s+h)s}{2s+h}}}}{\sqrt{2\pi\frac{(s+h)s}{2s+h}}}\,e^{Ay}\,dy$$
$$= e^{Ax}\int_0^t \frac{ds}{2\sqrt{\pi(s+h/2)}} e^{\frac{A^2}{4}\cdot\frac{(s+h)s}{s+h/2}}.$$

Then, using (9.60) and (9.63),

$$\begin{aligned}
&\int_{[0,t]\times\mathbf{R}} |p(s+h,x,y)-p(s,x,y)|^2\, e^{Ay}\, ds\, dy \\
&= \frac{e^{Ax}}{2\sqrt{\pi}} \int_0^t \Big(\frac{e^{\frac{A^2}{4}(s+h)}}{\sqrt{s+h}} - 2\frac{e^{\frac{A^2}{4}\cdot\frac{(s+h)s}{s+h/2}}}{\sqrt{s+h/2}} + \frac{e^{\frac{A^2}{4}s}}{\sqrt{s}} \Big)\, ds \\
&\le \frac{e^{Ax}}{2\sqrt{\pi}} \int_0^t \Big(\frac{e^{\frac{A^2}{4}(s+h)} - e^{\frac{A^2}{4}\cdot\frac{(s+h)s}{s+h/2}}}{\sqrt{s+h/2}} \\
&\qquad + \frac{e^{\frac{A^2}{4}s} - e^{\frac{A^2}{4}\cdot\frac{(s+h)s}{s+h/2}}}{\sqrt{s+h/2}} + e^{\frac{A^2}{4}s}\Big[\frac{1}{\sqrt{s}} - \frac{1}{\sqrt{s+h/2}}\Big] \Big)\, ds \\
&\le e^{Ax} C(A,t)\sqrt{h}.
\end{aligned}$$

Inside the integral in the middle member above, the first term is $O(h)$, the second negative, and the third $O(\sqrt{h})$. ∎

Now the lemma that will be crucial in propagating good properties along the iteration (9.53).

Lemma 9.16. *Fix $p > 8$. Let $\xi(t,x,\omega)$ be an adapted measurable process. Assume that for each $T \in (0,\infty)$ there exist nonnegative finite constants $K = K(T)$ and $A = A(T)$ such that*

$$\forall t \in [0,T],\ x \in \mathbf{R}: \quad E[\,|\xi(t,x)|^p\,] \le Ke^{A|x|}. \tag{9.64}$$

For each $(t,x) \in \mathbf{R}_+ \times \mathbf{R}$ let

$$\eta(t,x) = \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\, \xi(s,y)\, dW(s,y).$$

Then the process $\eta(t,x)$ is well-defined and has a continuous version. Furthermore, the moment bound is preserved with the same constant in the exponent but with a larger constant in front: there exists a constant $C(A,T)$ such that

$$\forall t \in [0,T],\ x \in \mathbf{R} \quad E[\,|\eta(t,x)|^p\,] \le C(A,T)Ke^{A|x|} \tag{9.65}$$

where A and K are the same constants in both (9.64) and (9.65).

Proof. We can think of $\eta(t,x) = \int_{[0,T]\times\mathbf{R}} Y^{t,x}(s,y)\, dW(s,y)$ for any $T \ge t$, with $Y^{t,x}(s,y,\omega) = \mathbf{1}_{[0,t)}(s)p(t-s,x,y)\xi(s,y)$. That $Y^{t,x} \in \mathcal{L}_2(W)$ follows from assumption (9.64) and property (9.60) of the heat kernel. Thus $Y^{t,x}$ is a legitimate integrand.

The continuity of η will come from the Kolmogorov-Centsov criterion, Theorem B.20. So we estimate moments of increments, first the space increment and then the time increment. Fix $T < \infty$, and restrict t to $[0,T]$ so that K and A in (9.64) can be regarded as fixed finite constants. Below apply first the Burkholder-Davis-Gundy inequality (9.33). Then inside the expectation apply Hölder's inequality with conjugate exponents $p/(p-2)$ and $p/2$, and note that the first integral can be taken outside the expectation.

$$
\begin{aligned}
&E|\eta(t,x)-\eta(t,z)|^p \\
&= E\biggl[\biggl|\int_{[0,t]\times\mathbf{R}} (p_{t-s,x,y}-p_{t-s,z,y})\,\xi(s,y)\,dW(s,y)\biggr|^p\biggr] \\
&\le CE\biggl[\biggl|\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^2\,|\xi(s,y)|^2\,ds\,dy\biggr|^{\frac{p}{2}}\biggr] \\
&= CE\biggl[\biggl|\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^{2-\frac{4}{p}}\,|p_{t-s,x,y}-p_{t-s,z,y}|^{\frac{4}{p}}\,|\xi_{s,y}|^2\,ds\,dy\biggr|^{\frac{p}{2}}\biggr] \\
&\le C\biggl(\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^2\,ds\,dy\biggr)^{\frac{p-2}{2}} \\
&\qquad\times\ E\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^2\,|\xi(s,y)|^p\,ds\,dy \\
&\le C\biggl(\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^2\,ds\,dy\biggr)^{\frac{p-2}{2}} \\
&\qquad\times\int_{[0,t]\times\mathbf{R}} |p_{t-s,x,y}-p_{t-s,z,y}|^2\,Ke^{A|y|}\,ds\,dy \\
&\le C|x-z|^{\frac{p}{2}}Ke^{A|x|} \le C|x-z|^{\frac{p}{2}}.
\end{aligned}
$$

In the next to last step we applied (9.56) and (9.58) restricted t, x and z to compact intervals to get a single constant C in front.

The time increment we treat in two pieces. Let $0 \le t < t+h \le T$.

$$
\begin{aligned}
&\eta(t+h,x)-\eta(t,x) \\
&= \int_{[0,t]\times\mathbf{R}} (p(t+h-s,x,y)-p(t-s,x,y))\,\xi(s,y)\,dW(s,y)
\end{aligned}
$$

$$+\int_{[t,t+h]\times\mathbf{R}} p(t+h-s,x,y)\,\xi(s,y)\,dW(s,y)$$

and so, from $|a+b|^p \le C(|a|^p+|b|^p)$ and inequality (9.33) again,

$$\begin{aligned}
&E|\eta(t+h,x)-\eta(t,x)|^p \\
&\le CE\bigg[\bigg|\int\limits_{[0,t]\times\mathbf{R}} |p_{t+h-s,x,y}-p_{t-s,x,y}|^2\,|\xi(s,y)|^2\,ds\,dy\bigg|^{\frac{p}{2}}\bigg] \\
&\quad + CE\bigg[\bigg|\int_{[t,t+h]\times\mathbf{R}} |p(t+h-s,x,y)|^2\,|\xi(s,y)|^2\,ds\,dy\bigg|^{\frac{p}{2}}\bigg].
\end{aligned}$$

Repeat the Hölder step from above and use assumption (9.64) again to get

$$\begin{aligned}
&E|\eta(t+h,x)-\eta(t,x)|^p \\
&\le C\bigg(\int\limits_{[0,t]\times\mathbf{R}} |p_{t+h-s,x,y}-p_{t-s,x,y}|^2\,ds\,dy\bigg)^{\frac{p-2}{2}} \\
&\qquad\qquad \times \int\limits_{[0,t]\times\mathbf{R}} |p_{t+h-s,x,y}-p_{t-s,x,y}|^2\,Ke^{A|y|}\,ds\,dy \\
&\quad + C\bigg(\int\limits_{[t,t+h]\times\mathbf{R}} |p(t+h-s,x,y)|^2\,ds\,dy\bigg)^{\frac{p-2}{2}} \\
&\qquad\qquad \times \int\limits_{[t,t+h]\times\mathbf{R}} |p(t+h-s,x,y)|^2\,Ke^{A|y|}\,ds\,dy \\
&\le Ch^{\frac{p-2}{4}}\cdot C(A,t)Ke^{|x|}\sqrt{h} \le Ch^{\frac{p}{4}}
\end{aligned}$$

We bounded the four integrals with (9.57), (9.59), (9.54) and (9.60), and then restricted t, x and z to compact intervals.

Combine the estimates to get, for (s,x) and (t,z) restricted to a compact subset of $\mathbf{R}_+\times\mathbf{R}$,

$$\begin{aligned}
E|\eta(s,x)-\eta(t,z)|^p &\le C\big(|s-t|^{\frac{p}{4}}+|x-z|^{\frac{p}{2}}\big) \\
&\le C|(s,x)-\eta(t,z)|^{\frac{p}{4}}.
\end{aligned}$$

In the last expression $|\ |$ stands for Euclidean distance on the plane. The restriction to a compact set allowed us to drop the higher exponent, at the price of adjusting the constant C. Note that for small increments the smaller exponent gives the larger upper bound, and this is the relevant one. Since dimension of space-time is $d=2$, the criterion from Theorem B.20 for the existence of a continuous version is $\frac{p}{4}>2 \Leftrightarrow p>8$.

To prove (9.65), use the Hölder argument from above and then (9.54) and (9.60), with t restricted to $[0, T]$.

$$\begin{aligned}
E|\eta(t,x)|^p &= E\biggl[\,\biggl|\int_{[0,t]\times\mathbf{R}} p_{t-s,x,y}\,\xi(s,y)\,dW(s,y)\biggr|^p\,\biggr] \\
&\le CE\biggl[\,\biggl|\int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,|\xi(s,y)|^2\,ds\,dy\biggr|^{\frac{p}{2}}\,\biggr] \\
&\le C\biggl(\int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,ds\,dy\biggr)^{\frac{p-2}{2}} \int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,Ke^{A|y|}\,ds\,dy \\
&\le C(A,T)Ke^{A|x|}.
\end{aligned}$$
□

Let us first discuss the consequences of the lemma for the mild solution u. Suppose we have a measurable adapted process u that satisfies the moment bound (9.64) and such that, for each fixed $(t, x) \in \mathbf{R}_+ \times \mathbf{R}$, (9.44) holds a.s. Property (9.48) transfers (9.64) to $\sigma(u(t, x))$, and then by Lemma 9.16 the stochastic integral on the right-hand side of (9.44) has a continuous version. Then the entire right-hand side of (9.44) has a continuous version, and hence so does the left-hand side, namely u itself. (Processes u and $\bar{u}$ are now versions of each other if $P\{u(t, x) = \bar{u}(t, x)\} = 1$ for each (t, x).) The stochastic integral is not changed by replacing u with another version of it. The reason is that the metric $d_{\mathcal{L}_2(W)}$ does not distinguish between versions and so versions have the same approximating simple predictable processes in Proposition 9.2. Consequently equation (9.44) continues to hold a.s. for the continuous version of u. The upshot is that in order to have a continuous solution u, we only need a measurable adapted solution u that satisfies moment bound (9.64).

We turn to the iteration (9.53). Assumptions (9.49) and (9.50) on u_0 and property (9.48) on $\sigma(u)$ guarantee the hypotheses of Lemmas 9.13 and 9.16 so that $u_0(t, x)$ is well-defined, has a continuous version, and satisfies again the locally uniform L^p bound (9.64). Repeating this, we define a sequence of continuous processes u_n by (9.53), each satisfying moment bound (9.64).

For the existence of the solution we develop a Gronwall type argument for the convergence of the iterates u_n, analogously to our earlier treatment of SDEs. Note the following convolution property (B.12) of the Gaussian kernel:

$$\int_{\mathbf{R}} p(s, y, x)p(s, x, y)\,dx = \frac{1}{\sqrt{2\pi \cdot 2s}}$$

Fix an compact time interval $[0, T]$. By restricting the treatment to $[0, T]$ we do not need to keep track of the t-dependence of various constants.

It is crucial to develop the moment bound (9.50) along the way in a manner that does not alter the constant A in the exponent. First we pass it to $u_0(t,x)$ defined by (9.52). Since $u_0(0,x) = u_0(x)$, take $t \in (0,T]$.

$$\begin{aligned} E|u_0(t,x)|^p &\le \int_{\mathbf{R}} p(t,x,y)\, Ke^{A|y|}\, dy \le Ke^{A|x|} \int_{\mathbf{R}} p(t,0,y)\, e^{A|y|}\, dy \\ &= Ke^{A|x|} E(e^{A\sqrt{t}|Z|}) = K_0 e^{A|x|} \end{aligned}$$

where Z denoted a $\mathcal{N}(0,1)$ random variable, and we defined a new constant $K_0 = K_0(A,T)$. Note that the constant A in $e^{A|x|}$ was preserved.

Next we control the initial difference. The constant C below can change from line to line and depend on (T,p), but not on (A,x).

$$\begin{aligned} &E|u_1(t,x) - u_0(t,x)|^p \\ &= E\biggl[\,\biggl| \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\, \sigma(u_0(s,y))\, dW(s,y) \biggr|^p\, \biggr] \\ &\le CE\biggl[\,\biggl| \int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\, |\sigma(u_0(s,y))|^2\, ds\, dy \biggr|^{\frac{p}{2}}\, \biggr] \\ &\le C\biggl(\int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\, ds\, dy \biggr)^{\frac{p-2}{2}} \int_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\, E|\sigma(u_0(s,y))|^p\, ds\, dy \\ &\le C \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2\, (1 + E|u_0(s,y)|^p)\, ds\, dy \\ &\le C \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2\, (1 + K_0 e^{A|y|})\, ds\, dy \\ &\le C(1 + C(A,T) K_0 e^{A|x|}) \le K_1 e^{A|x|}. \end{aligned}$$

On the last line we introduced yet another constant $K_1 = K_1(A,T)$.

For $n \ge 1$ we have the equation

$$\begin{aligned} &u_{n+1}(t,x) - u_n(t,x) \\ &\quad = \int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\, [\sigma(u_n(s,y)) - \sigma(u_{n-1}(s,y))]\, dW(s,y). \end{aligned}$$

For the induction argument introduce the sequence

$$a_n(t) = \sup_{s\in[0,t]} \sup_{x\in\mathbf{R}} e^{-A|x|} E|u_{n+1}(s,x) - u_n(s,x)|^p. \tag{9.66}$$

The calculation above gives the inequality

$$a_0(t) \le K_1 \quad \text{for } t \in [0,T] \tag{9.67}$$

which serves as the initial step of the induction. For the induction step we imitate yet again the calculations of Lemma 9.16, with restriction $t \in [0,T]$. The Lipschitz assumption (9.47) comes here into the argument. Again C can change from line to line and depend on (T,p) but not on (A,x).

$$\begin{aligned}
&E|u_{n+1}(t,x) - u_n(t,x)|^p \\
&= E\Biggl[\,\biggl|\int\limits_{[0,t]\times\mathbf{R}} p(t-s,x,y)\,[\sigma(u_n(s,y)) - \sigma(u_{n-1}(s,y))]\,dW(s,y)\biggr|^p\,\Biggr] \\
&\le CE\Biggl[\,\biggl|\int\limits_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,|\sigma(u_n(s,y)) - \sigma(u_{n-1}(s,y))|^2\,ds\,dy\biggr|^{\frac{p}{2}}\,\Biggr] \\
&\le CE\Biggl[\,\biggl|\int\limits_{[0,t]\times\mathbf{R}} p^{2-\frac{4}{p}}_{t-s,x,y}\,p^{\frac{4}{p}}_{t-s,x,y}\,|u_n(s,y) - u_{n-1}(s,y)|^2\,ds\,dy\biggr|^{\frac{p}{2}}\,\Biggr] \\
&\le C\biggl(\int\limits_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,ds\,dy\biggr)^{\frac{p-2}{2}} \\
&\qquad\qquad \times \int\limits_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,E|u_n(s,y) - u_{n-1}(s,y)|^p\,ds\,dy \\
&\le C \int\limits_{[0,t]\times\mathbf{R}} p^2_{t-s,x,y}\,e^{A|y|}\,a_{n-1}(s)\,ds\,dy \\
&= C\int_0^t ds\; a_{n-1}(s) \int_{\mathbf{R}} \frac{e^{-\frac{(y-x)^2}{t-s}}}{2\pi(t-s)}\,e^{A|y|}\,dy \\
&\le C\int_0^t ds\,\frac{a_{n-1}(s)}{\sqrt{t-s}}\;e^{A|x| + \frac{t-s}{4}A^2} \\
&\le e^{A|x|}\cdot K_2 \int_0^t \frac{a_{n-1}(s)}{\sqrt{t-s}}\,ds.
\end{aligned}$$

The next to last inequality came from the Gaussian moment generating function (9.55). We introduced a new constant $K_2 = K_2(A,T) = Ce^{TA^2/4}$. The development above can be summarized in the inequality

$$a_n(t) \le K_2 \int_0^t \frac{a_{n-1}(s)}{\sqrt{t-s}}\,ds \qquad t \in [0,T].$$

An application of Hölder's inequality gives

$$a_n(t) \le K_2 \Big(\int_0^t (t-s)^{-3/4}\, ds \Big)^{2/3} \Big(\int_0^t a_{n-1}(s)^3\, ds \Big)^{1/3}. \tag{9.68}$$

Define (yet again) a new constant and turn this into

$$a_n(t)^3 \le K_3 \int_0^t a_{n-1}(s)^3\, ds, \quad t \in [0,T].$$

As in (7.37)–(7.38), (9.67) together with the above inequality develop inductively into

$$a_n(t)^3 \le K_1^3 \frac{K_3^n t^n}{n!} \quad t \in [0,T].$$

We get this bound on L^p norms:

$$\begin{aligned} &\int_{[0,T]\times[-R,R]} \sum_{n\ge 0} E|u_{n+1}(t,x,\omega) - u_n(t,x,\omega)|^p \, dt\, dx \\ &\le 2RTe^{AR} \sum_{n\ge 0} \sup_{t\in[0,T]} \sup_{x\in\mathbf{R}} e^{-A|x|} E|u_{n+1}(t,x,\omega) - u_n(t,x,\omega)|^p \\ &\le 2RTe^{AR} \sum_{n\ge 0} a_n(T) \le C(A,T,R) \sum_{n\ge 0} \Big(\frac{K_3^n T^n}{n!} \Big)^{\frac13} < \infty \end{aligned}$$

where the convergence of the series can be seen for example from $n^n/n! \le e^n$.

From

$$E|u_{n+1}(t,x) - u_n(t,x)| \le e^{A|x|/p} a_n(t)$$

we deduce

$$E \sum_{n\ge 0} |u_{n+1}(t,x,\omega) - u_n(t,x,\omega)| < \infty,$$

from which is follows that

$$\forall\, (t,x) \,:\, \sum_{n\ge 0} |u_{n+1}(t,x,\omega) - u_n(t,x,\omega)| < \infty \text{ for } P\text{-a.e. } \omega.$$

Consequently there exists a limit function $u(t,x,\omega)$ such that

$$\forall\, (t,x) \,:\, u_n(t,x,\omega) \to u(t,x,\omega) \ P\text{-a.s.} \tag{9.69}$$

In particular, u is measurable and adapted to $\{\mathcal{G}_t\}$.

By (9.66) and the triangle inequality, for $t \in [0,T]$, $x \in \mathbf{R}$, and $n < m < \infty$,

$$\|u_n(t,x) - u_m(t,x)\|_{L^p(P)} \le e^{A|x|/p} \sum_{k=n}^{\infty} a_k(T)^{\frac1p} = e^{A|x|/p} K_{p,n}^{1/p} < \infty.$$

The last equality above defines the constant $K_{p,n}$ that vanishes as $n \to \infty$ by the convergence of the series. Since $u_m(t,x) \to u(t,x)$ a.s., we can let $m \to \infty$ on the left and apply Fatou's lemma to get

$$\|u_n(t,x) - u(t,x)\|_{L^p(P)} \le e^{A|x|/p} K_{p,n}^{1/p}. \tag{9.70}$$

Recall that the iteration preserved the L^p bound (9.64) for each u_n, with a fixed A but K that changed with n. The estimate above extends this bound to u and thereby verifies (9.51).

To summarize, we have a measurable, adapted limit process u with moment bound (9.51). From Lemma 9.16 we know that the stochastic integral on the right-hand side of (9.44) is well-defined. To conclude the proof that this process satisfies the definition of the mild solution, it only remains to show that the iterative scheme (9.53) turns, in the limit, into the definition (9.44) of the mild solution. It is enough to show L^2 convergence of the right-hand side of (9.53) to the right-hand side of (9.44), since we already know the a.s. convergence of the left-hand side. Utilizing the isometry, the Lipschitz assumption (9.47), and estimate (9.70),

$$\begin{aligned}
&E\biggl|\int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\,[\sigma(u_n(s,y)) - \sigma(u(s,y))]\,dW(s,y)\biggr|^2 \\
&= E\int_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2\,|\sigma(u_n(s,y)) - \sigma(u(s,y))|^2\,ds\,dy \\
&\le C\int_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2\,E|u_n(s,y) - u(s,y)|^2\,ds\,dy \\
&\le CK_{2,n}\int_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2\,e^{A|y|}\,ds\,dy \\
&\longrightarrow 0 \quad \text{as } n \to \infty.
\end{aligned}$$

This completes the existence proof.

We turn to uniqueness. Same steps as above, with the isometry and the Lipschitz assumption (9.47). Define

$$\alpha(t) = \sup_{s\in[0,t]}\sup_{x\in\mathbf{R}} e^{-A|x|}E|u(s,x) - v(s,x)|^2$$

$$\begin{aligned}
&E|u(t,x) - v(t,x)|^2 \\
&= E\biggl|\int_{[0,t]\times\mathbf{R}} p(t-s,x,y)\,[\sigma(u(s,y)) - \sigma(v(s,y))]\,dW(s,y)\biggr|^2
\end{aligned}$$

$$= E \int\limits_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2 \, |\sigma(u(s,y)) - \sigma(v(s,y))|^2 \, ds\, dy$$
$$\le C \int\limits_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2 \, E|u(s,y) - v(s,y)|^2 \, ds\, dy$$
$$\le C \int\limits_{[0,t]\times\mathbf{R}} p(t-s,x,y)^2 \, \alpha(s) e^{A|y|} \, ds\, dy.$$

As above, we develop this into

$$\alpha(t) \le C \int_0^t \frac{\alpha(s)}{\sqrt{t-s}} \, ds \le C \Big(\int_0^t \alpha(s)^3 \, ds \Big)^{1/3}$$

where we used again the Hölder trick of (9.68) and restricted to $t \in [0,T]$ to have a constant C independent of t. Gronwall's inequality (Lemma A.20) implies that $\alpha(t) \equiv 0$. This completes the proof of Theorem 9.12.

Exercises

In Exercises 9.2–9.3, W is a white noise based on a σ-finite measure space $(X, \mathcal{B}, \lambda)$.

Exercise 9.1. Let $(X, \mathcal{B}, \lambda)$ be a σ-finite measure space. Show that on some probability space there exists a mean zero Gaussian process $\{W(A) : A \in \mathcal{B},\ \lambda(A) < \infty\}$ with covariance $E[W(A)W(B)] = \lambda(A \cap B)$. In other words, *white noise* exists. This can be viewed as an application of Exercise 1.19.

Exercise 9.2. (a) Compute $E[(W(A \cup B) - W(A) - W(B))^2]$ and deduce the finite additivity of W.

(b) Show that $E[(W(A) - W(B))^2] = \lambda(A \triangle B)$.

(c) Show the a.s. countable additivity (9.3).

Exercise 9.3. For a set $U \in \mathcal{A}$ define the σ-algebra

$$\mathcal{G}_U = \sigma\{W(A) : A \in \mathcal{A}, A \subseteq U, \lambda(A) < \infty\}.$$

Show that the σ-algebras $\mathcal{G}_U$ and $\mathcal{G}_{U^c}$ are independent. *Hints.* By a π-λ argument it suffices to show the independence of any two vectors $(W(A_1), \dots, W(A_m))$ and $(W(B_1), \dots, W(B_n))$ where each $A_i \subseteq U$ and each $B_j \subseteq U^c$. Then use either the fact that the full vector $(W(A_1), \dots, W(A_m), W(B_1), \dots, W(B_n))$ is jointly Gaussian, or decompose $W(A_i)$'s and $W(B_j)$'s into sums of independent pieces.

Exercise 9.4. In this exercise you show that white noise $W(\,\cdot\,, \omega)$ is not a signed measure for a.e. fixed ω. Let W be a white noise on $\mathbf{R}^d$ based on Lebesgue measure m. Fix positive numbers $\{c_k\}_{k\in\mathbf{N}}$ such that $\sum c_k^2 = 1$

but $\sum c_k = \infty$. (For example, $c_k = \sqrt{6}/(\pi k)$.) Let $A = \bigcup_k A_k$ be a disjoint union of Borel sets with $m(A_k) = c_k^2$. Show that $\sum|W(A_k)| = \infty$ a.s. Thus $\sum W(A_k)$ does not converge absolutely and consequently $W(\,\cdot\,,\omega)$ cannot be a signed measure.

Exercise 9.5. Let W be white noise on $\mathbf{R}^d$ with respect to Lebesgue measure. For $c > 0$ show that $W^{(c)}(A,\omega) = c^{-d/2}W(cA,\omega)$ is another white noise on $\mathbf{R}^d$.

Exercise 9.6. Let W be white noise on $\mathbf{R}^d$ with respect to Lebesgue measure. If W had a density ξ, that is, if in some sense it would be true that $W(B) = \int_B \xi(x)\,dx$ for a stochastic process ξ, then presumably, again in some sense, $m(B)^{-1}W(B) \to \xi(x_0)$ as the set B shrinks down to the point x_0. However, show that this is not true even in the weakest possible sense, namely as a limit in distribution. Take $B_\varepsilon = [x_0 - \varepsilon/2, x_0 + \varepsilon/2]^d$ and show that $m(B_\varepsilon)^{-1}W(B_\varepsilon)$ is not even tight as $\varepsilon \searrow 0$. (For example, show that for any compact interval $[-b,b] \subseteq \mathbf{R}$, $P\{m(B)^{-1}W(B) \in [-b,b]\} \to 0$ as $\varepsilon \searrow 0$.)

Exercise 9.7. This exercise contains some of the details of the construction of the isonormal process $\{W(h) : h \in L^2(\lambda)\}$.

(a) Show linearity $W(\alpha g + \beta h) = \alpha W(g) + \beta W(h)$ for $\alpha, \beta \in \mathbf{R}$ and simple functions $g, h \in L^2(\lambda)$.

(b) Show that $W(h)$ is well-defined: that is, if $\{g_n\}$ and $\{h_n\}$ are two sequences of simple functions such that $g_n \to h$ and $h_n \to h$ in $L^2(\lambda)$, then $\lim_{n\to\infty} W(g_n) = \lim_{n\to\infty} W(h_n)$.

(c) Show that $\{W(h) : h \in L^2(\lambda)\}$ is a mean zero Gaussian process with covariance $E[W(g)W(h)] = \int gh\,d\lambda$. Note that you need to show that for any choice $h_1, \dots, h_n \in L^2(\lambda)$, the vector $(W(h_1), \dots, W(h_n))$ has a normal distribution.

Exercise 9.8. Let $\{B_t : t \in \mathbf{R}_+^d\}$ be Brownian sheet with index set $\mathbf{R}_+^d$. Show that for each compact set $K \subseteq \mathbf{R}_+^d$ there exists a constant $C = C(K) < \infty$ such that $E[(B_s - B_t)^2] \le C\sum_{i=1}^d |s_i - t_i|$ for $s, t \in K$. Use Theorem B.20 to show that Brownian sheet has a continuous version. (Exercise 1.12 points the way to finding bounds on high moments of $B_s - B_t$.)

Exercise 9.9. Expand the calculation in (9.9) to prove that for any A and B with $m(A) + m(B) < \infty$, $Z_t = W_t(A)W_t(B) - tm(A \cap B)$ is a martingale. You will need Exercise 9.3.

Exercise 9.10. Let $\mathcal{A}$ be a ring of subsets of a space X. Show that (i) $\emptyset \in \mathcal{A}$, (ii) $\mathcal{A}$ is closed under intersections, and (iii) $\mathcal{A}$ is an algebra iff $X \in \mathcal{A}$.

Exercise 9.11. Fill in the details of the proof of Lemma 9.4.

Exercise 9.12. (Fubini's theorem for stochastic integrals) Let $(U, \mathcal{U}, \mu)$ be a finite measure space, and $Y(t, x, u, \omega)$ a jointly measurable function on $\mathbf{R}_+ \times \mathbf{R}^d \times \Omega \times U$. Prove the statement

$$\begin{aligned}\int_U \mu(du) \int_{(0,t]\times B} Y(s,x,u)\, dW(s,x) \\ = \int_{(0,t]\times B} \Big(\int_U Y(s,x,u)\, \mu(du) \Big)\, dW(s,x)\end{aligned} \tag{9.71}$$

under some natural assumptions.

Exercise 9.13. Check that if process $u(t, x)$ satisfies definition (1.41) of a mild solution then it also satisfies the weak form (1.37). You will need (9.71).

Exercise 9.14. Since the existence proof of Theorem 9.12 is technical, a valuable way to understand the issues would be to rework parts of the proof with stronger assumptions. For example, reprove Lemma 9.16 to show that, if we assume

$$\sup_{x\in\mathbf{R}} E[\,|u_0(x)|^p\,] < \infty \tag{9.72}$$

for some $p > 8$, then

$$\forall T < \infty : \quad \sup_{t\in[0,T],\, x\in\mathbf{R}} E[\,|u(t,x)|^p\,] < \infty. \tag{9.73}$$

Appendix A

Analysis

Definition A.1. Let X be a space. A function $d : X \times X \to [0, \infty)$ is a *metric* if for all $x, y, z \in X$,

(i) $d(x, y) = 0$ if and only if $x = y$,

(ii) $d(x, y) = d(y, x)$, and

(iii) $d(x, y) \le d(x, z) + d(z, y)$.

The pair (X, d), or X alone, is called a *metric space*.

Convergence of a sequence $\{x_n\}$ to a point x in X means that the distance vanishes in the limit: $x_n \to x$ if $d(x_n, x) \to 0$. $\{x_n\}$ is a *Cauchy sequence* if $\sup_{m>n} d(x_m, x_n) \to 0$ as $n \to \infty$. *Completeness* of a metric space means that every Cauchy sequence in the space has a limit in the space. A countable set $\{y_k\}$ is *dense* in X if for every $x \in X$ and every $\varepsilon > 0$, there exists a k such that $d(x, y_k) < \varepsilon$. X is a *separable* metric space if it has a countable dense set. A complete, separable metric space is called a *Polish space*.

The Cartesian product $X^n = X \times X \times \cdots \times X$ is a metric space with metric $\bar{d}(\mathbf{x}, \mathbf{y}) = \sum_i d(x_i, y_i)$ defined for vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in X^n.

A *vector space* (also *linear space*) is a space X on which addition $x + y$ ($x, y \in X$) and scalar multiplication αx ($\alpha \in \mathbf{R}$, $x \in X$) are defined, and satisfy the usual algebraic properties that familiar vector spaces such as $\mathbf{R}^n$ have. For example, there must exist a zero vector $0 \in X$, and each $x \in X$ has an additive inverse $-x$ such that $x + (-x) = 0$.

Definition A.2. Let X be a vector space. A function $\|\cdot\| : X \to [0, \infty)$ is a *norm* if for all $x, y \in X$ and $\alpha \in \mathbf{R}$,

(i) $\|x\| = 0$ if and only if $x = 0$,
(ii) $\|\alpha x\| = |\alpha| \cdot \|x\|$, and
(iii) $\|x + y\| \leq \|x\| + \|y\|$.

X is a *normed space* if it has a norm. A metric on a normed space is defined by $d(x, y) = \|x - y\|$. A normed space that is complete as a metric space is called a *Banach Space*.

A basic extension result says that a uniformly continuous map (a synonym for "function") into a complete space can be extended to the closure of its domain.

Lemma A.3. *Let (X, d) and (Y, r) be metric spaces, and assume (Y, r) is complete. Let S be a subset of X, and $f : S \to Y$ a map that is uniformly continuous. Then there exists a unique continuous map $g : \bar{S} \to Y$ such that $g(x) = f(x)$ for $x \in S$. Furthermore g is also uniformly continuous.*

Proof. Fix $x \in \bar{S}$. There exists a sequence $x_n \in S$ that converges to x. We claim that $\{f(x_n)\}$ is a Cauchy sequence in Y. Let $\varepsilon > 0$. By uniform continuity, there exists $\eta > 0$ such that for all $z, w \in S$, if $d(z, w) \leq \eta$ then $r\big(f(z), f(w)\big) \leq \varepsilon/2$. Since $x_n \to x$, there exists $N < \infty$ such that $d(x_n, x) < \eta/2$ whenever $n \geq N$. Now if $m, n \geq N$,

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \eta/2 + < \eta/2 = \eta,$$

so by choice of η,

$$r\big(f(x_m), f(x_n)\big) \leq \varepsilon/2 < \varepsilon \quad \text{for } m, n \geq N. \tag{A.1}$$

By completeness of Y, the Cauchy sequence $\{f(x_n)\}$ converges to some point $y \in Y$. Let us check that for any other sequence $S \ni z_k \to x$, the limit of $f(z_m)$ is again y. Let again $\varepsilon > 0$, choose η and N as above, and choose K so that $d(z_k, x) < \eta/2$ whenever $k \geq K$. Let $n \geq N$ and $k \geq K$. By the triangle inequality

$$r\big(y, f(z_k)\big) \leq r\big(y, f(x_n)\big) + r\big(f(x_n), f(z_k)\big).$$

We can let $m \to \infty$ in (A.1), and since $f(x_m) \to y$ and the metric is itself a continuous funcion of each of its arguments, in the limit $r\big(y, f(x_n)\big) \leq \varepsilon/2$. Also,

$$d(z_k, x_n) \leq d(z_k, x) + d(x, x_n) < \eta/2 + < \eta/2 = \eta,$$

so $r\big(f(x_n), f(z_k)\big) \leq \varepsilon/2$. We have shown that $r\big(y, f(z_k)\big) \leq \varepsilon$ for $k \geq K$. In other words $f(z_k) \to y$.

The above development shows that, given $x \in \bar{S}$, we can unambigously define $g(x) = \lim f(x_n)$ by using any sequence x_n in S that converges to x. By the continuity of f, if x happens to lie in S, then $g(x) = \lim f(x_n) = f(x)$, so g is an extension of f to $\bar{S}$.

To show the continuity of g, let $\varepsilon > 0$, pick η as above, and suppose $d(x,z) \le \eta/2$ for $x, z \in \bar{S}$. Pick sequences $x_n \to x$ and $z_n \to z$ from S. Pick n large enough so that $r\big(f(x_n), f(x)\big) < \varepsilon/4$, $r\big(f(z_n), f(z)\big) < \varepsilon/4$, $d(x_n, x) < \eta/4$ and $d(z_n, z) < \eta/4$. Then

$$d(x_n, z_n) \le d(x_n, x) + d(x, z) + d(z_n, z) < \eta$$

which implies $r\big(f(x_n), f(z_n)\big) \le \varepsilon/2$. Then

$$\begin{aligned} r\big(f(x), f(z)\big) &\le r\big(f(x), f(x_n)\big) + r\big(f(x_n), f(z_n)\big) + r\big(f(z_n), f(z)\big) \\ &\le \varepsilon. \end{aligned}$$

This shows the uniform continuity of g.

Uniqueness of g follows from above because any continuous extension h of f must satisfy $h(x) = \lim f(x_n) = g(x)$ whenever a sequence x_n from S converges to x. ☐

Without uniform continuity the extension might not be possible, as evidenced by a simple example such as $S = [0,1) \cup (1,2]$, $f \equiv 1$ on $[0,1)$, and $f \equiv 2$ on $(1,2]$.

One key application of the extension is the following situation.

Lemma A.4. *Let X, Y and S be as in the previous lemma. Assume in addition that they are all linear spaces, and that the metrics satisfy $d(x_0, x_1) = d(x_0 + z, x_1 + z)$ for all $x_0, x_1, z \in X$ and $r(y_0, y_1) = r(y_0 + w, y_1 + w)$ for all $y_0, y_1, w \in Y$. Let $I : S \to Y$ be a continuous linear map. Then there exists a linear map $T : \bar{S} \to Y$ which agrees with I on S.*

A.1. Continuous, cadlag and BV functions

Fix an interval $[a,b]$. The uniform norm and the uniform metric on functions on $[a,b]$ are defined by

$$\|f\|_\infty = \sup_{t\in[a,b]} |f(t)| \quad \text{and} \quad d_\infty(f,g) = \sup_{t\in[a,b]} |f(t) - g(t)|. \tag{A.2}$$

One needs to distinguish these from the L^∞ norms defined in (1.8).

We can impose this norm and metric on different spaces of functions on $[a,b]$. Continuous functions are the most familiar. Studying stochastic processes leads us also to consider *cadlag functions*, which are right-continuous and have left limits at each point in $[a,b]$. Both cadlag functions and continuous functions form a complete metric space under the uniform metric. This is proved in the next lemma.

Lemma A.5. (a) *Suppose $\{f_n\}$ is a Cauchy sequence of functions in the metric d_∞ on $[a,b]$. Then there exists a function f on $[a,b]$ such that $d_\infty(f_n, f) \to 0$.*

(b) *If all the functions* f_n *are cadlag on* $[a,b]$, *then so is the limit* f.

(c) *If all the functions* f_n *are continuous on* $[a,b]$, *then so is the limit* f.

Proof. As an instance of the general definition, a sequence of functions $\{f_n\}$ on $[a,b]$ is a *Cauchy sequence* in the metric d_∞ if for each $\varepsilon > 0$ there exists a finite N such that $d_\infty(f_n, f_m) \le \varepsilon$ for all $m, n \ge N$. For a fixed t, the sequence of numbers $\{f_n(t)\}$ is a Cauchy sequence, and by the completeness of the real number system $f_n(t)$ converges as $n \to \infty$. These pointwise limits define the function

$$f(t) = \lim_{n\to\infty} f_n(t).$$

It remains to show uniform convergence $f_n \to f$. Fix $\varepsilon > 0$, and use the uniform Cauchy property to pick N so that

$$|f_n(t) - f_m(t)| \le \varepsilon \quad \text{for all } m, n \ge N \text{ and } t \in [a,b].$$

With n and t fixed, let $m \to \infty$. In the limit we get

$$|f_n(t) - f(t)| \le \varepsilon \quad \text{for all } n \ge N \text{ and } t \in [a,b].$$

This shows the uniform convergence.

(b) Fix $t \in [a,b]$. We first show $f(s) \to f(t)$ as $s \searrow t$. (If $t = b$ no approach from the right is possible and there is nothing to prove.) Let $\varepsilon > 0$. Pick n so that

$$\sup_{x\in[a,b]} |f_n(x) - f(x)| \le \varepsilon/4.$$

By assumption f_n is cadlag so we may pick $\delta > 0$ so that $t \le s \le t + \delta$ implies $|f_n(t) - f_n(s)| \le \varepsilon/4$. The triangle inequality then shows that

$$|f(t) - f(s)| \le \varepsilon$$

for $s \in [t, t+\delta]$. We have shown that f is right-continuous.

To show that left limits exist for f, use the left limit of f_n to find $\eta > 0$ so that $|f_n(r) - f_n(s)| \le \varepsilon/4$ for $r, s \in [t-\eta, t)$. Then again by the triangle inequality, $|f(r) - f(s)| \le \varepsilon$ for $r, s \in [t-\eta, t)$. This implies that

$$0 \le \limsup_{s\nearrow t} f(s) - \liminf_{s\nearrow t} f(s) \le \varepsilon$$

and that both limsup and liminf are finite. Since $\varepsilon > 0$ was arbitrary, the existence of the limit $f(t-) = \lim_{s\nearrow t} f(s)$ follows.

(c) Right continuity of f follows from part (b), and left continuity is proved by the same argument. ∎

A function f defined on $\mathbf{R}^d$ (or some subset of it) is *locally Lipschitz* if for every compact set K there exists a constant L_K such that

$$|f(x) - f(y)| \le L_K|x - y|$$

for all $x, y \in K$ in the domain of f. In particular, a locally Lipschitz function on $\mathbf{R}$ is Lipschitz continuous on any compact interval $[a, b]$. A weaker form of continuity is local Hölder continuity with exponent $\gamma \in (0, 1]$:

$$|f(x) - f(y)| \le L_K|x - y|^\gamma$$

for $x, y \in K$ in the domain of f. If $\gamma = 1$ then we are back in Lipschitz continuity. Values $\gamma > 1$ are not of interest. (Why?)

Lemma A.6. *Suppose $g \in BV[0,T]$ and f is locally Lipschitz on $\mathbf{R}$, or some subset of it that contains the range of g. Then $f \circ g$ is BV.*

Proof. A BV function on $[0, T]$ is bounded because

$$|g(x)| \le |g(0)| + V_g(T).$$

Hence f is Lipschitz continuous on the range of g. With L denoting the Lipschitz constant of f on the range of g, for any partition $\{t_i\}$ of $[0, T]$,

$$\sum_i |f(g(t_{i+1})) - f(g(t_i))| \le L \sum_i |g(t_{i+1}) - g(t_i)| \le LV_g(T). \qquad \square$$

Lemma A.7. *Suppose f has left and right limits at all points in $[0, T]$. Let $\alpha > 0$. Define the set of jumps of magnitude at least α by*

$$U = \{t \in [0,T] : |f(t+) - f(t-)| \ge \alpha\} \tag{A.3}$$

with the interpretations $f(0-) = f(0)$ and $f(T+) = f(T)$. Then U is finite. Consequently f can have at most countably many jumps in $[0, T]$.

Proof. If U were infinite, it would have a limit point $s \in [0, T]$. This means that every interval $(s - \delta, s + \delta)$ contains a point of U, other than s itself. But since the limits $f(s\pm)$ both exist, we can pick δ small enough so that $|f(r) - f(t)| < \alpha/2$ for all pairs $r, t \in (s - \delta, s)$, and all pairs $r, t \in (s, s + \delta)$. Then also $|f(t+) - f(t-)| \le \alpha/2$ for all $t \in (s - \delta, s) \cup (s, s + \delta)$, and so these intervals cannot contain any point from U. $\square$

The lemma applies to monotone functions, BV functions, cadlag and caglad functions.

Lemma A.8. *Let f be a cadlag function on $[0, T]$ and define U as in (A.3). Then*

$$\varlimsup_{\delta \searrow 0} \sup\{|f(v) - f(u)| : 0 \le u < v \le T, v - u \le \delta, (u, v] \cap U = \emptyset\} \le \alpha.$$

Proof. This is proved by contradiction. Assume there exists a sequence $\delta_n \searrow 0$ and points $u_n, v_n \in [0,T]$ such that $0 < v_n - u_n \le \delta_n$, $(u_n, v_n] \cap U = \emptyset$, and

$$|f(v_n) - f(u_n)| > \alpha + \varepsilon \tag{A.4}$$

for some $\varepsilon > 0$. By compactness of $[0,T]$, we may pass to a subsequence (denoted by u_n, v_n again) such that $u_n \to s$ and $v_n \to s$ for some $s \in [0,T]$. One of the three cases below has to happen for infinitely many n.

Case 1. $u_n < v_n < s$. Passing to the limit along a subsequence for which this happens gives

$$f(v_n) - f(u_n) \to f(s-) - f(s-) = 0$$

by the existence of left limits for f. This contradicts (A.4).

Case 2. $u_n < s \le v_n$. By the cadlag property,

$$|f(v_n) - f(u_n)| \to |f(s) - f(s-)|.$$

(A.4) is again contradicted because $s \in (u_n, v_n]$ implies s cannot lie in U, so the jump at s must have magnitude strictly less than α.

Case 3. $s \le u_n < v_n$. This is like Case 1. Cadlag property gives

$$f(v_n) - f(u_n) \to f(s) - f(s) = 0. \qquad \square$$

In Section 1.1.9 we defined the total variation $V_f(t)$ of a function defined on $[0,t]$. Let us also define the *quadratic cross variation* of two functions f and g by

$$[f,g](t) = \lim_{\text{mesh}(\pi)\to 0} \sum_{i=0}^{m(\pi)-1} \big(f(s_{i+1}) - f(s_i)\big)\big(g(s_{i+1}) - g(s_i)\big)$$

if the limit exists as the mesh of the partition $\pi = \{0 = s_0 < \cdots < s_{m(\pi)} = t\}$ tends to zero. The *quadratic variation* of f is $[f] = [f,f]$.

In the next development we write down sums of the type $\sum_{\alpha\in\mathcal{A}} x(\alpha)$ where $\mathcal{A}$ is an arbitrary set and $x : \mathcal{A} \to \mathbf{R}$ a function. Such a sum can be defined as follows: the sum has a finite value c if for every $\varepsilon > 0$ there exists a finite set $B \subseteq \mathcal{A}$ such that if E is a finite set with $B \subseteq E \subseteq \mathcal{A}$ then

$$\left| \sum_{\alpha\in E} x(\alpha) \; - \; c \right| \le \varepsilon. \tag{A.5}$$

If $\sum_{\mathcal{A}} x(\alpha)$ has a finite value, $x(\alpha) \ne 0$ for at most countably many α-values. In the above condition, the set B must contain all α such that $|x(\alpha)| > 2\varepsilon$, for otherwise adding on one such term violates the inequality. In other words, the set $\{\alpha : |x(\alpha)| \ge \eta\}$ is finite for any $\eta > 0$.

If $x(\alpha) \ge 0$ always, then

$$\sum_{\alpha\in\mathcal{A}} x(\alpha) = \sup\biggl\{\sum_{\alpha\in B} x(\alpha) : B \text{ is a finite subset of } \mathcal{A}\biggr\}$$

gives a value in $[0, \infty]$ which agrees with the definition above if it is finite. As for familiar series, absolute convergence implies convergence. In other words if

$$\sum_{\alpha\in\mathcal{A}} |x(\alpha)| < \infty,$$

then the sum $\sum_{\mathcal{A}} x(\alpha)$ has a finite value (Exercise A.2).

Lemma A.9. *Let f be a function with left and right limits on $[0, T]$. Then*

$$\sum_{s\in[0,T]} |f(s+) - f(s-)| \le V_f(T),$$

where we interpret $f(0-) = f(0)$ and $f(T+) = f(T)$. The sum is actually over a countable set because f has at most countably many jumps.

Proof. If f has unbounded variation there is nothing to prove because the right-hand side of the inequality is infinite. Suppose $V_f(T) < \infty$. If the conclusion of the lemma fails, there exists a finite set $\{s_1 < s_2 < \cdots < s_m\}$ of jumps such that

$$\sum_{i=1}^{m} |f(s_i+) - f(s_i-)| > V_f(T).$$

Pick disjoint intervals $(a_i, b_i) \ni s_i$ for each i. (If $s_1 = 0$ take $a_1 = 0$, and if $s_m = T$ take $b_m = T$.) Then

$$V_f(T) \ge \sum_{i=1}^{m} |f(b_i) - f(a_i)|$$

Let $a_i \nearrow s_i$ and $b_i \searrow s_i$ for each i, except for a_1 in case $s_1 = 0$ and for b_m in case $s_m = T$. Then the right-hand side above converges to

$$\sum_{i=1}^{m} |f(s_i+) - f(s_i-)|,$$

contradicting the earlier inequality. $\square$

Lemma A.10. *Let f and g be a real-valued cadlag functions on $[0, T]$, and assume $f \in BV[0, T]$. Then*

$$[f, g](T) = \sum_{s\in(0,T]} \bigl(f(s) - f(s-)\bigr)\bigl(g(s) - g(s-)\bigr) \tag{A.6}$$

and the sum above converges absolutely.

Proof. As a cadlag function g is bounded. (If it were not, we could pick s_n so that $|g(s_n)| \nearrow \infty$. By compactness, a subsequence $s_{n_k} \to t \in [0,T]$. But $g(t\pm)$ exist and are finite.) Then by the previous lemma

$$\sum_{s\in(0,T]} |f(s)-f(s-)|\cdot|g(s)-g(s-)| \le 2\|g\|_\infty \sum_{s\in(0,T]} |f(s)-f(s-)| \\ \le 2\|g\|_\infty V_f(T) < \infty.$$

This checks that the sum in (A.6) converges absolutely. Hence it has a finite value, and we can approximate it with finite sums. Let $\varepsilon > 0$. Let U_α be the set defined in (A.3) for g. For small enough $\alpha > 0$,

$$\Biggl| \sum_{s\in U_\alpha} \bigl(f(s)-f(s-)\bigr)\bigl(g(s)-g(s-)\bigr) \\ - \sum_{s\in(0,T]} \bigl(f(s)-f(s-)\bigr)\bigl(g(s)-g(s-)\bigr)\Biggr| \le \varepsilon.$$

Shrink α further so that $2\alpha V_f(T) < \varepsilon$.

For such $\alpha > 0$, let $U_\alpha = \{u_1 < u_2 < \cdots < u_n\}$. Let $\delta = \frac{1}{2}\min\{u_{k+1} - u_k\}$ be half the minimum distance between two of these jumps. Consider partitions $\pi = \{0 = s_0 < \cdots < s_{m(\pi)} = t\}$ with $\operatorname{mesh}(\pi) \le \delta$. Let $i(k)$ be the index such that $u_k \in (s_{i(k)}, s_{i(k)+1}]$, $1 \le k \le n$. By the choice of δ, a partition interval $(s_i, s_{i+1}]$ can contain at most one u_k. Each u_k lies in some $(s_i, s_{i+1}]$ because these intervals cover $(0,T]$ and for a cadlag function 0 is not a discontinuity. Let $I = \{0,\dots,m(\pi)-1\} \setminus \{i(1),\dots,i(n)\}$ be the complementary set of indices.

By Lemma A.8 we can further shrink δ so that if $\operatorname{mesh}(\pi) \le \delta$ then $|g(s_{i+1}) - g(s_i)| \le 2\alpha$ for $i \in I$. Then

$$\sum_{i=0}^{m(\pi)-1} \bigl(f(s_{i+1})-f(s_i)\bigr)\bigl(g(s_{i+1})-g(s_i)\bigr) \\ = \sum_{k=1}^{n} \bigl(f(s_{i(k)+1})-f(s_{i(k)})\bigr)\bigl(g(s_{i(k)+1})-g(s_{i(k)})\bigr) \\ + \sum_{i\in I} \bigl(f(s_{i+1})-f(s_i)\bigr)\bigl(g(s_{i+1})-g(s_i)\bigr) \\ \le \sum_{k=1}^{n} \bigl(f(s_{i(k)+1})-f(s_{i(k)})\bigr)\bigl(g(s_{i(k)+1})-g(s_{i(k)})\bigr) + 2\alpha V_f(T).$$

As $\text{mesh}(\pi) \to 0$, $s_{i(k)} < u_k$ and $s_{i(k)+1} \ge u_k$, while both converge to u_k. By the cadlag property the sum on the last line above converges to

$$\sum_{k=1}^{n} \big(f(u_k) - f(u_k-)\big)\big(g(u_k) - g(u_k-)\big).$$

Combining this with the choice of α made above gives

$$\begin{aligned}\varlimsup_{\text{mesh}(\pi)\to 0} \sum_{i=0}^{m(\pi)-1} &\big(f(s_{i+1}) - f(s_i)\big)\big(g(s_{i+1}) - g(s_i)\big)\\ &\le \sum_{s\in(0,T]} \big(f(s) - f(s-)\big)\big(g(s) - g(s-)\big) + 2\varepsilon.\end{aligned}$$

Reversing inequalities in the argument gives

$$\begin{aligned}\varliminf_{\text{mesh}(\pi)\to 0} \sum_{i=0}^{m(\pi)-1} &\big(f(s_{i+1}) - f(s_i)\big)\big(g(s_{i+1}) - g(s_i)\big)\\ &\ge \sum_{s\in(0,T]} \big(f(s) - f(s-)\big)\big(g(s) - g(s-)\big) - 2\varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary the proof is complete. ∎

Corollary A.11. *If f is a cadlag BV-function on $[0,T]$, then $[f]_T$ is finite and given by*

$$[f](T) = \sum_{s\in(0,T]} \big(f(s) - f(s-)\big)^2.$$

Lemma A.12. *Let f_n, f, g_n and g be cadlag functions on $[0,T]$ such that $f_n \to f$ uniformly and $g_n \to g$ uniformly on $[0,T]$. Suppose $f_n \in BV[0,T]$ for each n, and $C_0 \equiv \sup_n V_{f_n}(T) < \infty$. Then $[f_n, g_n]_T \to [f,g]_T$ as $n \to \infty$.*

Proof. The function f is also BV, because for any partition,

$$\sum_i |f(s_{i+1}) - f(s_i)| = \lim_{n\to\infty} \sum_i |f_n(s_{i+1}) - f_n(s_i)| \le C_0.$$

Consequently by Lemma A.10 we have

$$[f,g]_T = \sum_{s\in(0,T]} \big(f(s) - f(s-)\big)\big(g(s) - g(s-)\big)$$

and

$$[f_n,g_n]_T = \sum_{s\in(0,T]} \big(f_n(s) - f_n(s-)\big)\big(g_n(s) - g_n(s-)\big)$$

and all these sums converge absolutely. Pick $\delta > 0$. Let

$$U_n(\delta) = \{s \in (0,T] : |g_n(s) - g_n(s-)| \ge \delta\}$$

be the set of jumps of g of magnitude at least δ, and $U(\delta)$ the same for g. By the absolute convergence of the sum for $[f,g]$ we can pick $\delta > 0$ so that

$$\Bigl|[f,g]_T - \sum_{s\in U(\delta)} \bigl(f(s)-f(s-)\bigr)\bigl(g(s)-g(s-)\bigr)\Bigr| \le \varepsilon.$$

Shrink δ further so that $\delta < \varepsilon/C_0$. Then for any finite $H \supseteq U_n(\delta)$,

$$\begin{aligned}&\Bigl|\, [f_n,g_n]_T - \sum_{s\in H} \bigl(f_n(s)-f_n(s-)\bigr)\bigl(g_n(s)-g_n(s-)\bigr)\Bigr| \\ &\le \sum_{s\notin H} |f_n(s)-f_n(s-)|\cdot|g_n(s)-g_n(s-)| \\ &\le \delta \sum_{s\in(0,T]} |f_n(s)-f_n(s-)| \le \delta C_0 \le \varepsilon.\end{aligned}$$

We claim that for large enough n, $U_n(\delta) \subseteq U(\delta)$. Since a cadlag function has only finitely many jumps with magnitude above any given positive quantity, there exists a small $\alpha > 0$ such that g has no jumps s such that $|g(s)-g(s-)| \in (\delta-\alpha,\delta)$. If n is large enough so that $\sup_s|g_n(s)-g(s)| < \alpha/4$, then for each s

$$\bigl|\bigl(g_n(s)-g_n(s-)\bigr) \,-\, \bigl(g(s)-g(s-)\bigr)\bigr| \le \alpha/2.$$

Now if $s \in U_n(\delta)$, $|g_n(s)-g_n(s-)| \ge \delta$ and the above inequality imply $|g(s)-g(s-)| \ge \delta - \alpha/2$. This jump cannot fall in the forbidden range $(\delta-\alpha,\delta)$, so in fact it must satisfy $|g(s)-g(s-)| \ge \delta$ and then $s \in U(\delta)$.

Now we can complete the argument. Take n large enough so that $U(\delta) \supset U_n(\delta)$, and take $H = U(\delta)$ in the estimate above. Putting the estimates together gives

$$\begin{aligned}\bigl|\,[f,g]-[f_n,g_n]\,\bigr| &\le 2\varepsilon \\ &\quad + \Bigl|\sum_{s\in U(\delta)} \bigl(f(s)-f(s-)\bigr)\bigl(g(s)-g(s-)\bigr) \\ &\qquad - \sum_{s\in U(\delta)} \bigl(f_n(s)-f_n(s-)\bigr)\bigl(g_n(s)-g_n(s-)\bigr)\Bigr|.\end{aligned}$$

As $U(\delta)$ is a fixed finite set, the difference of two sums over U tends to zero as $n\to\infty$. Since $\varepsilon > 0$ was arbitrary, the proof is complete. ☐

We regard vectors as column vectors. T denotes transposition. So if $\mathbf{x} = [x_1,\dots,x_d]^T$ and $\mathbf{y} = [y_1,\dots,y_d]^T$ are elements of $\mathbf{R}^d$ and $A = (a_{i,j})$ is a $d\times d$ matrix, then

$$\mathbf{x}^T A\mathbf{y} = \sum_{1\le i,j\le d} x_i a_{i,j} y_j.$$

The Euclidean norm is $|\mathbf{x}| = (x_1^2 + \cdots + x_d^2)^{1/2}$, and we apply this also to matrices in the form

$$|A| = \Big\{ \sum_{1\le i,j\le d} a_{i,j}^2 \Big\}^{1/2}.$$

The Schwarz inequality says $|\mathbf{x}^T\mathbf{y}| \le |\mathbf{x}|\cdot|\mathbf{y}|$. This extends to

$$|\mathbf{x}^T A\mathbf{y}| \le |\mathbf{x}|\cdot|A|\cdot|\mathbf{y}|. \tag{A.7}$$

Lemma A.13. *Let $g_1,\dots,g_d$ be cadlag functions on $[0,T]$, and form the $\mathbf{R}^d$-valued cadlag function $\mathbf{g} = (g_1,\dots,g_d)^T$ with coordinates $g_1,\dots,g_d$. A cadlag function on a bounded interval is bounded, so there exists a closed, bounded set $K\subseteq\mathbf{R}^d$ such that $\mathbf{g}(s)\in K$ for all $s\in[0,T]$.*

Let ϕ be a continuous function on $[0,T]^2\times K^2$ such that the function

$$\gamma(s,t,\mathbf{x},\mathbf{y}) = \begin{cases} \dfrac{\phi(s,t,\mathbf{x},\mathbf{y})}{|t-s|+|\mathbf{y}-\mathbf{x}|^2}, & s\ne t \text{ or } \mathbf{x}\ne\mathbf{y} \\ 0, & s=t \text{ and } \mathbf{x}=\mathbf{y} \end{cases} \tag{A.8}$$

is also continuous on $[0,T]^2\times K^2$. Let $\pi^\ell = \{0=t_0^\ell < t_1^\ell < \cdots < t_{m(\ell)}^\ell = T\}$ be a sequence of partitions of $[0,T]$ such that $\operatorname{mesh}(\pi^\ell)\to 0$ as $\ell\to\infty$, and

$$C_0 = \sup_\ell \sum_{i=0}^{m(\ell)-1} \big|\mathbf{g}(t_{i+1}^\ell) - \mathbf{g}(t_i^\ell)\big|^2 < \infty. \tag{A.9}$$

Then

$$\lim_{\ell\to\infty} \sum_{i=0}^{m(\ell)-1} \phi\big(t_i^\ell, t_{i+1}^\ell, \mathbf{g}(t_i^\ell), \mathbf{g}(t_{i+1}^\ell)\big) = \sum_{s\in(0,T]} \phi\big(s,s,\mathbf{g}(s-),\mathbf{g}(s)\big). \tag{A.10}$$

The limit on the right is a finite, absolutely convergent sum.

Proof. To begin, note that $[0,T]^2\times K^2$ is a compact set, so any continuous function on $[0,T]^2\times K^2$ is bounded and uniformly continuous.

We claim that

$$\sum_{s\in(0,T]} \big|\mathbf{g}(s)-\mathbf{g}(s-)\big|^2 \le C_0 \tag{A.11}$$

where C_0 is the constant defined in (A.9). This follows the reasoning of Lemma A.9. Consider any finite set of points $s_1<\cdots<s_n$ in $(0,T]$. For each ℓ, pick indices $i(k)$ such that $s_k\in(t_{i(k)}^\ell, t_{i(k)+1}^\ell]$, $1\le k\le n$. For large enough ℓ all the $i(k)$'s are distinct, and then by (A.9)

$$\sum_{k=1}^n \big|\mathbf{g}(t_{i(k)+1}^\ell) - \mathbf{g}(t_{i(k)}^\ell)\big|^2 \le C_0.$$

As $\ell \to \infty$ the inequality above becomes

$$\sum_{k=1}^{n} \big|\mathbf{g}(s_k) - \mathbf{g}(s_k-)\big|^2 \le C_0$$

by the cadlag property, because for each ℓ, $t^\ell_{i(k)} < s_k \le t^\ell_{i(k)+1}$, while both extremes converge to s_k. The sum on the left-hand side of (A.11) is by definition the supremum of sums over finite sets, hence the inequality in (A.11) follows.

By continuity of γ there exists a constant C_1 such that

$$|\phi(s, t, \mathbf{x}, \mathbf{y})| \le C_1\big(|t - s| + |\mathbf{y} - \mathbf{x}|^2\big) \tag{A.12}$$

for all $s, t \in [0, T]$ and $\mathbf{x}, \mathbf{y} \in K$. From (A.12) and (A.11) we get the bound

$$\sum_{s\in(0,T]} \big|\phi\big(s, s, \mathbf{g}(s-), \mathbf{g}(s)\big)\big| \le C_0 C_1 < \infty.$$

This absolute convergence implies that the sum on the right-hand side of (A.10) can be approximated by finite sums. Given $\varepsilon > 0$, pick $\alpha > 0$ small enough so that

$$\bigg|\sum_{s\in U_\alpha} \phi\big(s, s, \mathbf{g}(s-), \mathbf{g}(s)\big) \; - \sum_{s\in(0,T]} \phi\big(s, s, \mathbf{g}(s-), \mathbf{g}(s)\big)\bigg| \le \varepsilon$$

where

$$U_\alpha = \{s \in (0, T] : |\mathbf{g}(s) - \mathbf{g}(s-)| \ge \alpha\}$$

is the set of jumps of magnitude at least α.

Since γ is uniformly continuous on $[0, T]^2 \times K^2$ and vanishes on the set $\{(u, u, \mathbf{z}, \mathbf{z}) : u \in [0, T], \mathbf{z} \in K\}$, we can shrink α further so that

$$|\gamma(s, t, \mathbf{x}, \mathbf{y})| \le \varepsilon/(T + C_0) \tag{A.13}$$

whenever $|t - s| \le 2\alpha$ and $|\mathbf{y} - \mathbf{x}| \le 2\alpha$. Given this α, let I^ℓ be the set of indices $0 \le i \le m(\ell)-1$ such that $(t^\ell_i, t^\ell_{i+1}] \cap U_\alpha = \emptyset$. By Lemma A.8 and the assumption $\text{mesh}(\pi^\ell) \to 0$ we can fix ℓ_0 so that for $\ell \ge \ell_0$, $\text{mesh}(\pi^\ell) < 2\alpha$ and

$$|\mathbf{g}(t^\ell_{i+1}) - \mathbf{g}(t^\ell_i)| \le 2\alpha \text{ for all } i \in I^\ell. \tag{A.14}$$

Note that the proof of Lemma A.8 applies to vector-valued functions.

Let $J^\ell = \{0, \dots, m(\ell) - 1\} \setminus I^\ell$ be the complementary set of indices i such that $(t^\ell_i, t^\ell_{i+1}]$ contains a point of U_α. Now we can bound the difference in (A.10). Consider $\ell \ge \ell_0$.

$$\Bigl|\sum_{i=0}^{m(\ell)-1} \phi\bigl(t_i^\ell, t_{i+1}^\ell, \mathbf{g}(t_i^\ell), \mathbf{g}(t_{i+1}^\ell)\bigr) \;-\; \sum_{s\in(0,T]} \phi\bigl(s, s, \mathbf{g}(s-), \mathbf{g}(s)\bigr)\Bigr|$$
$$\le \sum_{i\in I^\ell} \bigl|\phi\bigl(t_i^\ell, t_{i+1}^\ell, \mathbf{g}(t_i^\ell), \mathbf{g}(t_{i+1}^\ell)\bigr)\bigr|$$
$$+ \Bigl|\sum_{i\in J^\ell} \phi\bigl(t_i^\ell, t_{i+1}^\ell, \mathbf{g}(t_i^\ell), \mathbf{g}(t_{i+1}^\ell)\bigr) \;-\; \sum_{s\in U_\alpha} \phi\bigl(s, s, \mathbf{g}(s-), \mathbf{g}(s)\bigr)\Bigr| + \varepsilon.$$

The first sum after the inequality above is bounded above by ε, by (A.8), (A.9), (A.13) and (A.14). The difference of two sums in absolute values vanishes as $\ell \to \infty$, because for large enough ℓ each interval $(t_i^\ell, t_{i+1}^\ell]$ for $i \in J^\ell$ contains a unique $s \in U_\alpha$, and as $\ell \to \infty$,

$$t_i^\ell \to s, \quad t_{i+1}^\ell \to s, \quad \mathbf{g}(t_i^\ell) \to \mathbf{g}(s-) \quad \text{and} \quad \mathbf{g}(t_{i+1}^\ell) \to \mathbf{g}(s)$$

by the cadlag property. (Note that U_α is finite by Lemma A.7 and for large enough ℓ index set J_ℓ has exactly one term for each $s \in U_\alpha$.) We conclude

$$\limsup_{\ell\to\infty} \Bigl|\sum_{i=0}^{m(\ell)-1} \phi\bigl(t_i^\ell, t_{i+1}^\ell, \mathbf{g}(t_i^\ell), \mathbf{g}(t_{i+1}^\ell)\bigr) \;-\; \sum_{s\in(0,T]} \phi\bigl(s, s, \mathbf{g}(s-), \mathbf{g}(s)\bigr)\Bigr| \le 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. ☐

A.2. Differentiation and integration

For an open set $G \subseteq \mathbf{R}^d$, $C^2(G)$ is the space of functions $f : G \to \mathbf{R}$ whose partial derivatives up to second order exist and are continuous. We use subscript notation for partial derivatives, as in

$$f_{x_1} = \frac{\partial f}{\partial x_1} \quad \text{and} \quad f_{x_1,x_2} = \frac{\partial^2 f}{\partial x_1 \partial x_2}.$$

These will always be applied to functions with continuous partial derivatives so the order of differentiation does not matter. The *gradient* Df is the column vector of first-order partial derivatives:

$$Df(\mathbf{x}) = \bigl[f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \dots, f_{x_d}(\mathbf{x})\bigr]^T.$$

The *Hessian matrix* $D^2 f$ is the $d \times d$ matrix of second-order partial derivatives:

$$D^2 f(\mathbf{x}) = \begin{bmatrix} f_{x_1,x_1}(\mathbf{x}) & f_{x_1,x_2}(\mathbf{x}) & \cdots & f_{x_1,x_d}(\mathbf{x}) \\ f_{x_2,x_1}(\mathbf{x}) & f_{x_2,x_2}(\mathbf{x}) & \cdots & f_{x_2,x_d}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_d,x_1}(\mathbf{x}) & f_{x_d,x_2}(\mathbf{x}) & \cdots & f_{x_d,x_d}(\mathbf{x}) \end{bmatrix}.$$

We prove a version of Taylor's theorem. Let us write $f \in C^{1,2}([0,T]\times G)$ if the partial derivatives f_t, f_{x_i} and f_{x_i,x_j} exist and are continuous on $(0,T)\times G$, and they extend as continuous functions to $[0,T]\times G$. The continuity of f_t is actually not needed for the next theorem. But this hypothesis will be present in the application to the proof of Itô's formula, so we assume it here already.

Theorem A.14. (a) *Let (a,b) be an open interval in $\mathbf{R}$, $f \in C^{1,2}([0,T]\times(a,b))$, $s,t \in [0,T]$ and $x,y \in (a,b)$. Then there exists a point τ between s and t and a point θ between x and y such that*

$$\begin{aligned} f(t,y) = f(s,x) &+ f_x(s,x)(y-x) + f_t(\tau,y)(t-s) \\ &+ \tfrac{1}{2} f_{xx}(s,\theta)(y-x)^2. \end{aligned} \tag{A.15}$$

(b) *Let G be an open convex set in $\mathbf{R}^d$, $f \in C^{1,2}([0,T]\times G)$, $s,t \in [0,T]$ and $\mathbf{x},\mathbf{y} \in G$. Then there exists a point τ between s and t and $\theta \in [0,1]$ such that, with $\xi = \theta\mathbf{x} + (1-\theta)\mathbf{y}$,*

$$\begin{aligned} f(t,\mathbf{y}) = f(s,\mathbf{x}) &+ Df(s,\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + f_t(\tau,\mathbf{y})(t-s) \\ &+ \tfrac{1}{2}(\mathbf{y}-\mathbf{x})^T D^2 f(s,\xi)(\mathbf{y}-\mathbf{x}). \end{aligned} \tag{A.16}$$

Proof. Part (a). Check by integration by parts that

$$\psi(s,x,y) = \int_x^y (y-z) f_{xx}(s,z)\,dz$$

satisfies

$$\psi(s,x,y) = f(s,y) - f(s,x) - f_x(s,x)(y-x).$$

By the mean value theorem there exists a point τ between s and t such that

$$f(t,y) - f(s,y) = f_t(\tau,y)(t-s).$$

By the intermediate value theorem there exists a point θ between x and y such that

$$\psi(s,x,y) = f_{xx}(s,\theta)\int_x^y (y-z)\,dz = \tfrac{1}{2} f_{xx}(s,\theta)(y-x)^2.$$

The application of the intermediate value theorem goes like this. Let $f_{xx}(s,u)$ and $f_{xx}(s,v)$ be the minimum and maximum of $f_{xx}(s,\cdot\,)$ in $[x,y]$ (or $[y,x]$ if $y<x$). Then

$$f_{xx}(s,u) \le \frac{\psi(s,x,y)}{\frac{1}{2}(y-x)^2} \le f_{xx}(s,v).$$

The intermediate value theorem gives a point θ between a and b such that

$$f_{xx}(s,\theta) = \frac{\psi(s,x,y)}{\frac{1}{2}(y-x)^2}.$$

The continuity of f_{xx} is needed here. Now

$$\begin{aligned} f(t,y) &= f(s,x) + f_x(s,x)(y-x) + f_t(\tau,y)(t-s) + \psi(s,x,y) \\ &= f(s,x) + f_x(s,x)(y-x) + f_t(\tau,y)(t-s) + \tfrac{1}{2} f_{xx}(s,\theta)(y-x)^2. \end{aligned}$$

Part (b). Apply part (a) to the function $g(t,r) = f\big(t, \mathbf{x} + r(\mathbf{y}-\mathbf{x})\big)$ for $(t,r) \in [0,T] \times (-\varepsilon, 1+\varepsilon)$ for a small enough $\varepsilon > 0$ so that $\mathbf{x} + r(\mathbf{y}-\mathbf{x}) \in G$ for $-\varepsilon \le r \le 1+\varepsilon$. □

The following generalization of the dominated convergence theorem is sometimes useful. It is proved by applying Fatou's lemma to $g_n \pm f_n$.

Theorem A.15. *Let f_n, f, g_n, and g be measurable functions such that $f_n \to f$, $|f_n| \le g_n \in L^1(\mu)$ and $\int g_n \, d\mu \to \int g \, d\mu < \infty$. Then*

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Lemma A.16. *Let $(X, \mathcal{B}, \nu)$ be a general measure space. Assume $g_n \to g$ in $L^p(\nu)$ for some $1 \le p < \infty$. Then for any measurable set $B \in \mathcal{B}$,*

$$\int_B |g_n|^p \, d\nu \to \int_B |g|^p \, d\nu.$$

Proof. Let $\widetilde{\nu}$ denote the measure restricted to the subspace B, and $\widetilde{g}_n$ and $\widetilde{g}$ denote functions restricted to this space. Since

$$\int_B |\widetilde{g}_n - \widetilde{g}|^p \, d\widetilde{\nu} = \int_B |g_n - g|^p \, d\nu \le \int_X |g_n - g|^p \, d\nu$$

we have $L^p(\widetilde{\nu})$ convergence $\widetilde{g}_n \to \widetilde{g}$. L^p norms (as all norms) are subject to the triangle inequality, and so

$$\big| \, \|\widetilde{g}_n\|_{L^p(\widetilde{\nu})} - \|\widetilde{g}\|_{L^p(\widetilde{\nu})} \big| \le \|\widetilde{g}_n - \widetilde{g}\|_{L^p(\widetilde{\nu})} \to 0.$$

Consequently

$$\int_B |g_n|^p \, d\nu = \|\widetilde{g}_n\|^p_{L^p(\widetilde{\nu})} \to \|\widetilde{g}\|^p_{L^p(\widetilde{\nu})} = \int_B |g|^p \, d\nu.$$

□

A common technical tool is approximation of general functions by functions of some convenient type. Here is one such result that we shall use later.

Lemma A.17. *Let μ be a σ-finite Borel measure on $\mathbf{R}$, $1 \le p < \infty$, and $f \in L^p(\mu)$. Let $\varepsilon > 0$. Then there is a continuous function g supported on a bounded interval, and a step function h of the form*

$$h(t) = \sum_{i=1}^{m-1} \alpha_i \mathbf{1}_{(s_i, s_{i+1}]}(t) \tag{A.17}$$

for some points $-\infty < s_1 < s_2 < \cdots < s_m < \infty$ and reals α_i, such that

$$\int_{\mathbf{R}} |f - g|^p \, d\mu < \varepsilon \quad \text{and} \quad \int_{\mathbf{R}} |f - h|^p \, d\mu < \varepsilon.$$

If there exists a constant C such that $|f| \le C$, then g and h can also be selected so that $|g| \le C$ and $|h| \le C$.

When the underlying measure is Lebesgue measure, one often writes $L^p(\mathbf{R})$ for the function spaces.

Proposition A.18 (L^p continuity)**.** *Let $f \in L^p(\mathbf{R})$. Then*

$$\lim_{h \to 0} \int_{\mathbf{R}} |f(t) - f(t+h)|^p \, dt = 0.$$

Proof. Check that the property is true for a step function of the type (A.17). Then approximate an arbitrary $f \in L^p(\mathbf{R})$ with a step function. □

Proposition A.19. *Let T be an invertible linear transformation on $\mathbf{R}^n$ and f a Borel or Lebesgue measurable function on $\mathbf{R}^n$. Then if f is either in $L^1(\mathbf{R}^n)$ or nonnegative,*

$$\int_{\mathbf{R}^n} f(x) \, dx = |\det T| \int_{\mathbf{R}^n} f(T(x)) \, dx. \tag{A.18}$$

The next inequality is a basic tool for deriving estimates. We use it in the chapter on SDEs.

Lemma A.20. (Gronwall's inequality) *Let g be an integrable Borel function on $[a, b]$, f a nondecreasing function on $[a, b]$, and assume that there exists a constant B such that*

$$g(t) \le f(t) + B \int_a^t g(s) \, ds, \quad a \le t \le b.$$

Then

$$g(t) \le f(t) e^{B(t-a)}, \quad a \le t \le b.$$

Proof. The integral $\int_a^t g(s)\,ds$ of an integrable function g is an absolutely continuous (AC) function of the upper limit t. At Lebesgue-almost every t it is differentiable and the derivative equals $g(t)$. Consequently the equation

$$\frac{d}{dt}\left(e^{-Bt}\int_a^t g(s)\,ds\right) = -Be^{-Bt}\int_a^t g(s)\,ds + e^{-Bt}g(t)$$

is valid for Lebesgue-almost every t. Hence by the hypothesis

$$\begin{aligned}\frac{d}{dt}\left(e^{-Bt}\int_a^t g(s)\,ds\right) &= e^{-Bt}\left(g(t) - B\int_a^t g(s)\,ds\right)\\ &\le f(t)e^{-Bt} \quad \text{for almost every } t.\end{aligned}$$

An absolutely continuous function is the integral of its almost everywhere existing derivative. Integrating above and using the monotonicity of f give

$$e^{-Bt}\int_a^t g(s)\,ds \le \int_a^t f(s)e^{-Bs}\,ds \le \frac{f(t)}{B}\big(e^{-Ba} - e^{-Bt}\big)$$

from which

$$\int_a^t g(s)\,ds \le \frac{f(t)}{B}\big(e^{B(t-a)} - 1\big).$$

Using the assumption once more,

$$g(t) \le f(t) + B\cdot\frac{f(t)}{B}\big(e^{B(t-a)} - 1\big) = f(t)e^{B(t-a)}. \qquad \square$$

Exercises

Exercise A.1. Let f be a cadlag function on $[a,b]$. Show that f is bounded, that is, $\sup_{t\in[a,b]}|f(t)| < \infty$. *Hint.* Suppose $|f(t_j)| \to \infty$ for some sequence $t_j \in [a,b]$. By compactness some subsequence converges: $t_{j_k} \to t \in [a,b]$. But cadlag says that $f(t-)$ exists as a limit among real numbers and $f(t+) = f(t)$.

Exercise A.2. Let $\mathcal{A}$ be a set and $x : \mathcal{A} \to \mathbf{R}$ a function. Suppose

$$c_1 \equiv \sup\left\{\sum_{\alpha\in B}|x(\alpha)| : B \text{ is a finite subset of } \mathcal{A}\right\} < \infty.$$

Show that then the sum $\sum_{\alpha\in\mathcal{A}} x(\alpha)$ has a finite value in the sense of the definition stated around equation (A.5).

Hint. Pick finite sets B_k such that $\sum_{B_k}|x(\alpha)| > c_1 - 1/k$. Show that the sequence $a_k = \sum_{B_k} x(\alpha)$ is a Cauchy sequence. Show that $c = \lim a_k$ is the value of the sum.

Appendix B

Probability

B.1. General matters

B.1.1. Measures and σ-algebras. It is sometimes convenient to work with classes of sets that are simpler than σ-algebras. A collection $\mathcal{A}$ of subsets of a space Ω is an *algebra* if

(i) $\Omega \in \mathcal{A}$.

(ii) $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$.

(iii) $A \cup B \in \mathcal{A}$ whenever $A \in \mathcal{A}$ and $B \in \mathcal{A}$.

A collection $\mathcal{S}$ of subsets of Ω is a *semialgebra* if it has these properties:

(i) $\emptyset \in \mathcal{S}$.

(ii) If $A, B \in \mathcal{S}$ then also $A \cap B \in \mathcal{S}$.

(iii) If $A \in \mathcal{S}$, then A^c is a finite disjoint union of elements of $\mathcal{S}$.

In applications, one sometimes needs to generate an algebra with a semialgebra, which is particularly simple.

Lemma B.1. *Let $\mathcal{S}$ be a semialgebra, and $\mathcal{A}$ the algebra generated by $\mathcal{S}$, in other words the intersection of all algebras that contain $\mathcal{S}$. Then $\mathcal{A}$ is the collection of all finite disjoint unions of members of $\mathcal{S}$.*

Proof. Let $\mathcal{B}$ be the collection of all finite disjoint unions of members of $\mathcal{S}$. Since any algebra containing $\mathcal{S}$ must contain $\mathcal{B}$, it suffices to verify that $\mathcal{B}$ is an algebra. By hypothesis $\emptyset \in \mathcal{S}$ and $\Omega = \emptyset^c$ is a finite disjoint union of members of $\mathcal{S}$, hence a member of $\mathcal{B}$. Since $A \cup B = (A^c \cap B^c)^c$ (deMorgan's law), it suffices to show that $\mathcal{B}$ is closed under intersections and complements.

Let $A = \bigcup_{1\le i\le m} S_i$ and $B = \bigcup_{1\le j\le n} T_j$ be finite disjoint unions of members of $\mathcal{S}$. Then $A \cap B = \bigcup_{i,j} A_i \cap B_j$ is again a finite disjoint union of members of $\mathcal{S}$. By the properties of a semialgebra, we can write $S_i^c = \bigcup_{1\le k\le p(i)} R_{i,k}$ as a finite disjoint union of members of $\mathcal{S}$. Then

$$A^c = \bigcap_{1\le i\le m} S_i^c = \bigcup_{(k(1),\dots,k(m))} \bigcap_{1\le i\le m} R_{i,k(i)}.$$

The last union above is over m-tuples $(k(1),\dots,k(m))$ such that $1 \le k(i) \le p(i)$. Each $\bigcap_{1\le i\le m} R_{i,k(i)}$ is an element of $\mathcal{S}$, and for distinct m-tuples these are disjoint because $k \ne \ell$ implies $R_{i,k} \cap R_{i,\ell} = \emptyset$. Thus $A^c \in \mathcal{B}$ too. $\square$

Algebras are sufficiently rich to provide approximation with error of arbitrarily small measure. The operation $\triangle$ is the symmetric difference defined by $A\triangle B = (A \setminus B) \cup (B \setminus A)$. The next lemma is proved by showing that the collection of sets that can be approximated as claimed form a σ-algebra.

Lemma B.2. *Suppose μ is a finite measure on the σ-algebra $\sigma(\mathcal{A})$ generated by an algebra $\mathcal{A}$. Show that for every $B \in \sigma(\mathcal{A})$ and $\varepsilon > 0$ there exists $A \in \mathcal{A}$ such that $\mu(A\triangle B) < \varepsilon$.*

A particularly useful tool for basic work in measure theory is Dynkin's π-λ theorem. Let $\mathcal{L}$ and $\mathcal{R}$ be collections of subsets of a space Ω. $\mathcal{R}$ is a π-system if it is closed under intersections, in other words if $A, B \in \mathcal{R}$, then $A \cap B \in \mathcal{R}$. $\mathcal{L}$ is a λ-system if it has the following three properties:

(1) $\Omega \in \mathcal{L}$.

(2) If $A, B \in \mathcal{L}$ and $A \subseteq B$ then $B \setminus A \in \mathcal{L}$.

(3) If $\{A_n : 1 \le n < \infty\} \subseteq \mathcal{L}$ and $A_n \nearrow A$ then $A \in \mathcal{L}$.

Theorem B.3 (Dynkin's π-λ theorem)**.** *If $\mathcal{R}$ is a π-system and $\mathcal{L}$ is a λ-system that contains $\mathcal{R}$, then $\mathcal{L}$ contains the σ-algebra $\sigma(\mathcal{R})$ generated by $\mathcal{R}$.*

For a proof, see the Appendix in [5]. The π-λ theorem has the following version for functions.

Theorem B.4. *Let $\mathcal{R}$ be a π-system on a space X such that $X = \bigcup B_i$ for some pairwise disjoint sequence $B_i \in \mathcal{R}$. Let $\mathcal{H}$ be a linear space of bounded functions on X. Assume that $\mathbf{1}_B \in \mathcal{H}$ for all $B \in \mathcal{R}$, and assume that $\mathcal{H}$ is closed under bounded, increasing pointwise limits. The second statement means that if $f_1 \le f_2 \le f_3 \le \cdots$ are elements of $\mathcal{H}$ and $\sup_{n,x} f_n(x) \le c$ for some constant c, then $f(x) = \lim_{n\to\infty} f_n(x)$ is a function in $\mathcal{H}$. It follows from these assumptions that $\mathcal{H}$ contains all bounded $\sigma(\mathcal{R})$-measurable functions.*

Proof. $\mathcal{L} = \{A : \mathbf{1}_A \in \mathcal{H}\}$ is a λ-system containing $\mathcal{R}$. Note that $X \in \mathcal{L}$ follows because by assumption $\mathbf{1}_X = \lim_{n\to\infty} \sum_{i=1}^n \mathbf{1}_{B_i}$ is an increasing limit of functions in $\mathcal{H}$. By the π-λ theorem indicator functions of all $\sigma(\mathcal{R})$-measurable sets lie in $\mathcal{H}$. By linearity all $\sigma(\mathcal{R})$-measurable simple functions lie in $\mathcal{H}$. A bounded, nonnegative $\sigma(\mathcal{R})$-measurable function is an increasing limit of nonnegative simple functions, and hence in $\mathcal{H}$. And again by linearity, all bounded $\sigma(\mathcal{R})$-measurable functions lie in $\mathcal{H}$. $\square$

Lemma B.5. *Let $\mathcal{R}$ be a π-system on a space X. Let μ and ν be two (possibly infinite) measures on $\sigma(\mathcal{R})$, the σ-algebra generated by $\mathcal{R}$. Assume that μ and ν agree on $\mathcal{R}$. Assume further that there is a countable collection of pairwise disjoint sets $\{R_i\} \subseteq \mathcal{R}$ such that $X = \bigcup R_i$ and $\mu(R_i) = \nu(R_i) < \infty$. Then $\mu = \nu$ on $\sigma(\mathcal{R})$.*

Proof. It suffices to check that $\mu(A) = \nu(A)$ for all $A \in \sigma(\mathcal{R})$ that lie inside some R_i. Then for a general $B \in \sigma(\mathcal{R})$,

$$\mu(B) = \sum_i \mu(B \cap R_i) = \sum_i \nu(B \cap R_i) = \nu(B).$$

Inside a fixed R_j, let

$$\mathcal{D} = \{A \in \sigma(\mathcal{R}) : A \subseteq R_j,\ \mu(A) = \nu(A)\}.$$

$\mathcal{D}$ is a λ-system. Checking property (2) uses the fact that R_j has finite measure under μ and ν so we can subtract: if $A \subseteq B$ and both lie in $\mathcal{D}$, then

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A),$$

so $B \setminus A \in \mathcal{D}$. By hypothesis $\mathcal{D}$ contains the π-system $\{A \in \mathcal{R} : A \subseteq R_j\}$. By the π-λ-theorem $\mathcal{D}$ contains all the $\sigma(\mathcal{R})$-sets that are contained in R_j. $\square$

Lemma B.6. *Let ν and μ be two finite Borel measures on a metric space (S, d). Assume that*

$$\int_S f\, d\mu = \int_S f\, d\nu \tag{B.1}$$

for all bounded continuous functions f on S. Then $\mu = \nu$.

Proof. Given a closed set $F \subseteq S$,

$$f_n(x) = \frac{1}{1 + n\,\mathrm{dist}(x, F)} \tag{B.2}$$

defines a bounded continuous function which converges to $\mathbf{1}_F(x)$ as $n \to \infty$. The quantity in the denominator is the distance from the point x to the set F, defined by

$$\mathrm{dist}(x, F) = \inf\{d(x, y) : y \in F\}.$$

For a closed set F, $\mathrm{dist}(x, F) = 0$ iff $x \in F$. Letting $n \to \infty$ in (B.1) with $f = f_n$ gives $\mu(F) = \nu(F)$. Apply Lemma B.5 to the class of closed sets. $\square$

B.1.2. Inequalities and limits. There is a handful of inequalities that are used constantly in probability. We already stated the conditional version of Jensen's inequality in Theorem 1.26. Of course the reader should realize that as a special case we get the unconditioned version: $f(EX) \le Ef(X)$ for convex f for which the expectations are well-defined.

Proposition B.7. (i) (*Markov's inequality*) *For a random variable* $X \ge 0$ *and a real number* $b > 0$,

$$P\{X \ge b\} \le b^{-1}EX. \tag{B.3}$$

(ii) (*Chebyshev's inequality*) *For a random variable* X *with finite mean and variance and a real* $b > 0$,

$$P\{|X - EX| \ge b\} \le b^{-2}\operatorname{Var}(X). \tag{B.4}$$

Proof. Chebyshev's is a special case of Markov's, and Markov's is quickly proved:

$$P\{X \ge b\} = E\mathbf{1}\{X \ge b\} \le b^{-1}E\big(X\mathbf{1}\{X \ge b\}\big) \le b^{-1}EX. \qquad \square$$

Let $\{A_n\}$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, P)$. We say A_n *happens infinitely often at* ω if $\omega \in A_n$ for infinitely many n. Equivalently,

$$\{\omega : A_n \text{ happens infinitely often}\} = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n.$$

The complementary event is

$$\{\omega : A_n \text{ happens only finitely many times}\} = \bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} A_n^c.$$

Often it is convenient to use the random variable

$$N_0(\omega) = \inf\Big\{m : \omega \in \bigcap_{n=m}^{\infty} A_n^c\Big\}$$

with $N_0(\omega) = \infty$ if A_n happens infinitely often at ω.

Lemma B.8. (*Borel-Cantelli Lemma*) *Let* $\{A_n\}$ *be a sequence of events in a probability space* $(\Omega, \mathcal{F}, P)$ *such that* $\sum_n P(A_n) < \infty$. *Then*

$$P\{A_n \textit{ happens infinitely often}\} = 0,$$

or equivalently, $P\{N_0 < \infty\} = 1$.

Proof. Since the tail of a convergent series can be made arbitrarily small, we have

$$P\Big(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\Big) \le \sum_{n=m}^{\infty} P(A_n) \to 0 \quad \text{as } m \to \infty. \qquad \square$$

The typical application of the Borel-Cantelli lemma is some variant of this idea.

Lemma B.9. *Let* $\{X_n\}$ *be a sequence of random variables on* $(\Omega, \mathcal{F}, P)$. *Suppose*

$$\sum_n P\{|X_n| \geq \varepsilon\} < \infty \quad \textit{for every } \varepsilon > 0.$$

Then $X_n \to 0$ *almost surely.*

Proof. Translating the familiar ε-definition of convergence into an event gives

$$\{\omega : X_n(\omega) \to 0\} = \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} \{\omega : |X_n(\omega)| \leq 1/k\}.$$

For each $k \geq 1$, the hypothesis and the Borel-Cantelli lemma give

$$P\Big(\bigcup_{m\geq 1} \bigcap_{n\geq m} \{|X_n| \leq 1/k\} \Big) = 1 - P\Big(\bigcap_{m\geq 1} \bigcup_{n\geq m} \{|X_n| > 1/k\} \Big) = 1.$$

A countable intesection of events of probability one has also probability one. □

Here is an example of the so-called *diagonal trick* for constructing a single subsequence that satisfies countably many requirements.

Lemma B.10. *For each* $j \in \mathbf{N}$ *let* Y^j *and* $\{X_n^j\}$ *be random variables such that* $X_n^j \to Y^j$ *in probability as* $n \to \infty$. *Then there exists a single subsequence* $\{n_k\}$ *such that, simultaneously for all* $j \in \mathbf{N}$, $X_{n_k}^j \to Y^j$ *almost surely as* $k \to \infty$.

Proof. Applying part (iii) of Theorem 1.20 as it stands gives for each j separately a subsequence $n_{k,j}$ such that $X_{n_{k,j}}^j \to Y^j$ almost surely as $k \to \infty$. To get one subsequence that works for all j we need a further argument.

Begin by choosing for $j = 1$ a subsequence $\{n(k,1)\}_{k\in\mathbf{N}}$ such that $X_{n(k,1)}^1 \to Y^1$ almost surely as $k \to \infty$. Along the subsequence $n(k,1)$ the variables $X_{n(k,1)}^2$ still converge to Y^2 in probability. Consequently we can choose a further subsequence $\{n(k,2)\} \subseteq \{n(k,1)\}$ such that $X_{n(k,2)}^2 \to Y^2$ almost surely as $k \to \infty$.

Continue inductively, producing nested subsequences

$$\{n(k,1)\} \supseteq \{n(k,2)\} \supseteq \cdots \supseteq \{n(k,j)\} \supseteq \{n(k,j+1)\} \supseteq \cdots$$

such that for each j, $X_{n(k,j)}^j \to Y^j$ almost surely as $k \to \infty$. Now set $n_k = n(k,k)$. Then for each j, from $k = j$ onwards n_k is a subsequence of $n(k,j)$. Consequently $X_{n_k}^j \to Y^j$ almost surely as $k \to \infty$. □

Here is a convergence theorem for random series.

Theorem B.11. *Let $\{X_k\}_{k\in\mathbf{N}}$ be independent random variables such that $E(X_k) = 0$ and $\sum \text{Var}(X_k) < \infty$. Then the random series $\sum_{k=1}^n X_k$ converges almost surely to a finite limit as $n \to \infty$.*

The basic convergence theorems of integration theory often work if almost everywhere convergence is replaced by the weaker types of convergence that are common in probability. As an example, here is the dominated convergence theorem under convergence in probability.

Theorem B.12. *Let X_n be random variables on $(\Omega, \mathcal{F}, P)$, and assume $X_n \to X$ in probability. Assume there exists a random variable $Y \geq 0$ such that $|X_n| \leq Y$ almost surely for each n, and $EY < \infty$. Then $EX_n \to EX$.*

Proof. It suffices to show that every subsequence $\{n_k\}$ has a further subsubsequence $\{n_{k_j}\}$ such that $EX_{n_{k_j}} \to EX$ as $j \to \infty$. So let $\{n_k\}$ be given. Convergence in probability $X_{n_k} \to X$ implies almost sure convergence $X_{n_{k_j}} \to X$ along some subsubsequence $\{n_{k_j}\}$. The standard dominated convergence theorem now gives $EX_{n_{k_j}} \to EX$. □

B.1.3. More about conditional expectations. When we discuss conditional expectations, $\mathcal{A}$ is a sub-σ-field of $\mathcal{F}$ on the probability space $(\Omega, \mathcal{F}, P)$. It is sometimes convenient to know that, just like expectations, conditional expectations are well-defined for nonnegative random variables without integrability assumptions.

Lemma B.13. *Let X be a $[0, \infty]$-valued random variable on $(\Omega, \mathcal{F}, P)$. Then there exists an a.s. unique $\mathcal{A}$-measurable $[0, \infty]$-valued random variable $E(X|\mathcal{A})$ with the property that*

$$E[\mathbf{1}_A X] = E\big[\mathbf{1}_A E(X|\mathcal{A})\big] \qquad \text{for all } A \in \mathcal{A}.$$

Since X is not assumed integrable, the identity above may be $\infty = \infty$ for some choices of A.

Proof. *Existence.* The sequence $E(X \wedge n|\mathcal{A})$ is almost surely monotone nondecreasing, and so the almost sure $\mathcal{A}$-measurable $[0, \infty]$-valued limit $Y = \lim_{n\to\infty} E(X \wedge n|\mathcal{A})$ exists. By the monotone convergence theorem

$$E\big[\mathbf{1}_A Y\big] = \lim_{n\to\infty} E\big[\mathbf{1}_A\, E(X \wedge n|\mathcal{A})\big] = \lim_{n\to\infty} E\big[\mathbf{1}_A \cdot (X \wedge n)\big] = E\big[\mathbf{1}_A\, X\big].$$

Consequently the integral identity $E[\mathbf{1}_A Y] = E[\mathbf{1}_A X]$ holds for all $A \in \mathcal{A}$.

Uniqueness. Suppose Y and $\tilde{Y}$ are two nonnegative $\mathcal{A}$-measurable random variables that satisfy the integral identity. Two calculations combine to yield the uniqueness $P(Y \neq \tilde{Y}) = 0$. First for any $n \in \mathbf{N}$

$$\begin{aligned} E\big[\mathbf{1}_{\{\tilde{Y}<Y\le n\}}(Y-\tilde{Y})\big] &= E[\mathbf{1}_{\{\tilde{Y}<Y\le n\}}Y] - E[\mathbf{1}_{\{\tilde{Y}<Y\le n\}}\tilde{Y}] \\ &= E[\mathbf{1}_{\{\tilde{Y}<Y\le n\}}X] - E[\mathbf{1}_{\{\tilde{Y}<Y\le n\}}X] = 0. \end{aligned}$$

It is of course essential that the last subtraction is between two finite numbers, which follows from the fact that the previous expectations are $[0,n]$-valued. Union over n and reversing the roles of Y and $\tilde{Y}$ give $P(\tilde{Y} < Y < \infty) + P(Y < \tilde{Y} < \infty) = 0$.

To handle infinite values, compute again first for $n \in \mathbf{N}$:

$$n \ge E[\mathbf{1}_{\{Y=\infty,\tilde{Y}\le n\}}\tilde{Y}] = E[\mathbf{1}_{\{Y=\infty,\tilde{Y}\le n\}}X] = E[\mathbf{1}_{\{Y=\infty,\tilde{Y}\le n\}}Y].$$

Finiteness of the last expectation forces $P(Y = \infty, \tilde{Y} \le n) = 0$, and this extends to $P(\tilde{Y} < \infty = Y) + P(Y < \infty = \tilde{Y}) = 0$. □

Conditional expectations satisfy some of the same convergence theorems as ordinary expectations.

Theorem B.14. *The hypotheses and conclusions below are all "almost sure". In parts* (i) *and* (ii) *the random variables are not necessarily integrable.*

(i) (*Monotone Convergence Theorem*) *Suppose* $0 \le X_n \le X_{n+1}$ *for all* n *and* $X_n \nearrow X$. *Then* $E(X_n|\mathcal{A}) \nearrow E(X|\mathcal{A})$.

(ii) (*Fatou's Lemma*) *If* $X_n \ge 0$ *for all* n, *then*

$$E\big(\varliminf X_n|\mathcal{A}\big) \le \varliminf E(X_n|\mathcal{A}).$$

(iii) (*Dominated Convergence Theorem*) *Suppose* $X_n \to X$, $|X_n| \le Y$ *for all* n, *and* $Y \in L^1(P)$. *Then* $E(X_n|\mathcal{A}) \to E(X|\mathcal{A})$.

Proof. Part (i). By monotonicity the almost sure limit $\lim_{n\to\infty} E(X_n|\mathcal{A})$ exists. By the ordinary monotone convergence theorem this limit satisfies the defining property of $E(X|\mathcal{A})$: for $A \in \mathcal{A}$,

$$E\big[\mathbf{1}_A \cdot \lim_{n\to\infty} E(X_n|\mathcal{A})\big] = \lim_{n\to\infty} E\big[\mathbf{1}_A\, E(X_n|\mathcal{A})\big] = \lim_{n\to\infty} E\big[\mathbf{1}_A\, X_n\big] = E\big[\mathbf{1}_A\, X\big].$$

Part (ii). The sequence $Y_k = \inf_{m\ge k} X_m$ increases up to $\varliminf X_n$. Thus by part (i),

$$E\big(\varliminf X_n|\mathcal{A}\big) = \lim_{n\to\infty} E\big(\inf_{k\ge n} X_k|\mathcal{A}\big) \le \varliminf_{n\to\infty} E(X_n|\mathcal{A}).$$

Part (iii). Using part (ii),

$$\begin{aligned} E(X|\mathcal{A}) + E(Y|\mathcal{A}) &= E\big(\varliminf\{X_n+Y\}|\mathcal{A}\big) \le \varliminf E\big(X_n+Y|\mathcal{A}\big) \\ &= \varliminf E(X_n|\mathcal{A}) + E(Y|\mathcal{A}). \end{aligned}$$

This gives $\underline{\lim}\, E(X_n|\mathcal{A}) \geq E(X|\mathcal{A})$. Apply this to $-X_n$ to get $\overline{\lim}\, E(X_n|\mathcal{A}) \leq E(X|\mathcal{A})$. ∎

We defined uniform integrability for a sequence in Section 1.2.2. The idea is the same for a more general collection of random variables.

Definition B.15. Let $\{X_\alpha : \alpha \in A\}$ be a collection of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. They are *uniformly integrable* if

$$\lim_{M\to\infty} \sup_{\alpha\in A} E\big[\,|X_\alpha| \cdot \mathbf{1}\{|X_\alpha| \geq M\}\big] = 0.$$

Equivalently, the following two conditions are satisfied. (i) $\sup_\alpha E|X_\alpha| < \infty$. (ii) Given $\varepsilon > 0$, there exists a $\delta > 0$ such that for every event B such that $P(B) \leq \delta$,

$$\sup_{\alpha\in A} \int_B |X_\alpha|\, dP \leq \varepsilon.$$

Proof of the equivalence of the two formulations of uniform integrability can be found for example in [4]. The next lemma is a great exercise, or a proof can be looked up in Section 4.5 of [5].

Lemma B.16. *Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$. Then the collection of random variables*

$$\{E(X|\mathcal{A}) : \ \mathcal{A} \text{ is a sub-}\sigma\text{-field of } \mathcal{F}\}$$

is uniformly integrable.

Lemma B.17. *Suppose $X_n \to X$ in L^1 on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{A}$ be a sub-σ-field of $\mathcal{F}$. Then there exists a subsequence $\{n_j\}$ such that $E[X_{n_j}|\mathcal{A}] \to E[X|\mathcal{A}]$ a.s.*

Proof. We have

$$\lim_{n\to\infty} E\big(E[\,|X_n - X|\,|\mathcal{A}]\,\big) = \lim_{n\to\infty} E\big(\,|X_n - X|\,\big) = 0,$$

and since

$$|\,E[X_n|\mathcal{A}] - E[X|\mathcal{A}]\,| \leq E[\,|X_n - X|\,|\mathcal{A}],$$

we conclude that $E[X_n|\mathcal{A}] \to E[X|\mathcal{A}]$ in L^1. L^1 convergence implies a.s. convergence along some subsequence. ∎

Lemma B.18. *Let X be a random d-vector and $\mathcal{A}$ a sub-σ-field on the probability space $(\Omega, \mathcal{F}, P)$. Let*

$$\phi(\theta) = \int_{\mathbf{R}^d} e^{i\theta^T\mathbf{x}}\,\mu(d\mathbf{x}) \quad (\theta \in \mathbf{R}^d)$$

be the characteristic function (Fourier transform) of a probability distribution μ *on* $\mathbf{R}^d$*. Assume*

$$E\big[\exp\{i\theta^T X\}\mathbf{1}_A\big] = \phi(\theta)P(A)$$

for all $\theta \in \mathbf{R}^d$ *and* $A \in \mathcal{A}$*. Then* X *has distribution* μ *and is independent of* $\mathcal{A}$*.*

Proof. Taking $A = \Omega$ above shows that X has distribution μ. Fix A such that $P(A) > 0$, and define the probability measure ν_A on $\mathbf{R}^d$ by

$$\nu_A(B) = \frac{1}{P(A)} E\big[\mathbf{1}_B(X)\mathbf{1}_A\big], \quad B \in \mathcal{B}_{\mathbf{R}^d}.$$

By hypothesis, the characteristic function of ν_A is ϕ. Hence $\nu_A = \mu$, which is the same as saying that

$$P\big(\{X \in B\} \cap A\big) = \nu_A(B)P(A) = \mu(B)P(A).$$

Since $B \in \mathcal{B}_{\mathbf{R}^d}$ and $A \in \mathcal{A}$ are arbitrary, the independence of X and $\mathcal{A}$ follows. □

In Chapter 1, page 27, we defined conditional probabilities $P(B|\mathcal{A})$ as special cases of conditional expectations: $P(B|\mathcal{A}))(\omega) = E(\mathbf{1}_B|\mathcal{A})(\omega)$. The question we address here is whether this definition can give us actual conditional probability measures so that conditional expectations could then be defined by integration. This point is not needed often in this text, but does appear in the proof of Theorem 7.15.

It turns out that when the underlying spaces are complete, separable metric spaces (such spaces are called Polish spaces), this can be arranged. This is good news because all the spaces we usually work with are Polish, such as $\mathbf{R}^d$ with its usual metric, $C_{\mathbf{R}^d}[0,T]$ with the uniform metric, and $C_{\mathbf{R}^d}[0,\infty)$ with the metric of uniform convergence on compact time intervals.

Theorem B.19. (a) *Let* $(\Omega, \mathcal{F}, P)$ *consist of a Polish space* Ω *with its Borel* σ*-algebra* $\mathcal{F}$ *and a Borel probability measure* P*. Let* $\mathcal{A}$ *be a sub-*σ*-algebra of* $\mathcal{F}$*. Then there exists a mapping* $\omega \mapsto P^\omega$ *from* Ω *into the space of Borel probability measures on* Ω *such that, for each* $B \in \mathcal{F}$*,* $\omega \mapsto P^\omega(B)$ *is* $\mathcal{A}$*-measurable and* $P^\omega(B)$ *is a version of* $P(B|\mathcal{A})(\omega)$*.*

That $P^\omega(B)$ is a version of $P(B|\mathcal{A})(\omega)$ means that $P^\omega(B)$ satisfies the definition of $P(B|\mathcal{A})(\omega)$. The term "version" is used because a conditional expectation is only defined up to events of measure zero.

P^ω is called the conditional probability measure of P, given $\mathcal{A}$. It can be used to define conditional expectations by integration: for any $X \in L^1(P)$,

$$E(X|\mathcal{A})(\omega) = \int_\Omega X(\widetilde{\omega})\, P^\omega(d\widetilde{\omega})$$

defines a version of the conditional expectation.

When Ω is Polish, the space of probability measures $\mathcal{M}_1(\Omega)$ with its so-called weak topology is also Polish. Hence $\mathcal{M}_1(\Omega)$ has its own Borel σ-algebra. The mapping $\omega \mapsto P^\omega$ from Ω into $\mathcal{M}_1(\Omega)$ is Borel measurable.

B.1.4. Continuity of stochastic processes. The extremely useful Kolmogorov-Centsov criterion establishes path-level continuity of a stochastic process from moment estimates on increments. We prove the result for a process X indexed by the d-dimensional unit cube $[0,1]^d$ and with values in a general complete separable metric space (S, ρ). The unit cube can be replaced by any bounded rectangle in $\mathbf{R}^d$ by a simple affine transformation $s \mapsto As + h$ of the index.

Define

$$D_n = \big\{(k_1, \dots, k_d)2^{-n} : k_1, \dots, k_d \in \{0, 1, \dots, 2^n\}\big\}$$

and then the set $D = \bigcup_{n\geq 0} D_n$ of *dyadic rational points* in $[0,1]^d$. The trick for getting a continuous version of a process is to use Lemma A.3 to rebuild the process from its values on D.

Theorem B.20. *Let (S, ρ) be a complete separable metric space.*

(a) *Suppose $\{X_s : s \in D\}$ is an S-valued stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$ with the following property: there exist constants $K < \infty$ and $\alpha, \beta > 0$ such that*

$$E\big[\rho(X_s, X_r)^\beta\big] \le K|s-r|^{d+\alpha} \quad \textit{for all } r, s \in D. \tag{B.5}$$

Then there exists an S-valued process $\{Y_s : s \in [0,1]^d\}$ on $(\Omega, \mathcal{F}, P)$ such that the path $s \mapsto Y_s(\omega)$ is continuous for each $\omega \in \Omega$ and $P\{Y_s = X_s\} = 1$ for each $s \in D$. Furthermore, Y is Hölder continuous with index γ for each $\gamma \in (0, \alpha/\beta)$: for each $\omega \in \Omega$ there exists a constant $C(\gamma, \omega) < \infty$ such that

$$\rho(Y_s(\omega), Y_r(\omega)) \le C(\gamma, \omega)|s-r|^\gamma \quad \textit{for all } r, s \in [0,1]^d. \tag{B.6}$$

(b) *Suppose $\{X_s : s \in [0,1]^d\}$ is an S-valued stochastic process that satisfies the moment bound (B.5) for all $r, s \in [0,1]^d$. Then X has a continuous version that is Hölder continuous with exponent γ for each $\gamma \in (0, \alpha/\beta)$.*

Proof. For $n \ge 0$ let $\eta_n = \max\{\rho(X_r, X_s) : r, s \in D_n,\ |s-r| = 2^{-n}\}$ denote the maximal nearest-neighbor increment among points in D_n. The idea of the proof is to use nearest-neighbor increments to control all increments by building paths through dyadic points. Pick $\gamma \in (0, \alpha/\beta)$. First, noting that D_n has $(2^n+1)^d$ points and each point in D_n has at most $2d$ nearest neighbors,

$$E(\eta_n^\beta) \le \sum_{r,s\in D_n:|s-r|=2^{-n}} E\big[\rho(X_s,X_r)^\beta\,\big] \le 2d(2^n+1)^d K 2^{-n(d+\alpha)}$$
$$\le C(d)2^{-n\alpha}.$$

Consequently

$$E\Big[\sum_{n\ge 0}(2^{n\gamma}\eta_n)^\beta\Big] = \sum_{n\ge 0} 2^{n\gamma\beta} E\eta_n^\beta \le C(d)\sum_{n\ge 0} 2^{n(\gamma\beta-\alpha)} < \infty.$$

A nonnegative random variable with finite expectation must be finite almost surely. So there exists an event Ω_0 such that $P(\Omega_0)=1$ and $\sum_n (2^{n\gamma}\eta_n)^\beta < \infty$ for $\omega\in\Omega_0$. In particular, for each $\omega\in\Omega_0$ there exists a constant $C(\omega)<\infty$ such that

$$\eta_n(\omega) \le C(\omega)2^{-n\gamma} \quad \text{for all } n\ge 0. \tag{B.7}$$

This constant depends on γ we ignore that in the notation.

Now we construct the paths. Suppose first $d=1$ so our dyadic rationals are points in the interval $[0,1]$. The simple but key observation is that given $r<s$ in D such that $s-r\le 2^{-m}$, the interval $(r,s]$ can be partitioned as $r=s_0<s_1<\cdots<s_M=s$ so that each pair (s_{i-1},s_i) is a nearest-neighbor pair in a particular D_n for some $n\ge m$, and no D_n contributes more than two such pairs .

To see this, start by finding the smallest $m_0\ge m$ such that $(r,s]$ contains an interval of the type $(k2^{-m_0},(k+1)2^{-m_0}]$. At most two such intervals fit inside $(r,s]$ because otherwise we could have taken m_0-1. Remove these level m_0 intervals from $(r,s]$. What remains are two intervals $(r,k_0 2^{-m_0}]$ and $(\ell_0 2^{-m_0},s]$ both of length strictly less than 2^{-m_0}. Now the process continues separately on the left and right piece. On the left, next look for the smallest $m_1>m_0$ such that for some k_1

$$(k_1-1)2^{-m_1} < r \le k_1 2^{-m_1} < k_0 2^{-m_0}.$$

This choice ensures that $(k_1 2^{-m_1}, k_0 2^{-m_0})$ is a nearest-neighbor pair in D_{m_1}. Remove the interval $(k_1 2^{-m_1}, k_0 2^{-m_0}]$, be left with $(r, k_1 2^{-m_1}]$, and continue in this manner. Note that if $r<k_1 2^{-m_1}$ then r cannot lie in D_{m_1}. Since r lies in some D_{n_0} eventually this process stops. The same process is performed on the right.

Extend this to d dimensions by connecting each coordinate in turn. In conclusion, if $r,s\in D$ satisfy $|s-r|\le 2^{-m}$ we can write a finite sum $s-r=\sum(s_i-s_{i-1})$ such that each (s_{i-1},s_i) is a nearest-neighbor pair from a particular D_n for some $n\ge m$, and no D_n contributes more than $2d$ such pairs. By the triangle inequality and by (B.7), for $\omega \in \Omega_0$

$$\rho(X_r(\omega), X_s(\omega)) \le \sum_{n \ge m} 2dC(\omega)2^{-n\gamma} \le C(\omega)2^{-m\gamma}.$$

We utilized above the common convention of letting a constant change from one step to the next when its precise value is of no consequence.

Now given any $r, s \in D$ pick $m \ge 0$ so that $2^{-m-1} \le |r - s| \le 2^{-m}$, and use the previous display to write, still for $\omega \in \Omega_0$,

$$\begin{aligned} \rho(X_r(\omega), X_s(\omega)) &\le C(\omega)2^{-m\gamma} \le C(\omega)2^{\gamma}(2^{-m-1})^{\gamma} \\ &\le C(\omega)|s - r|^{\gamma}. \end{aligned} \tag{B.8}$$

This shows that for $\omega \in \Omega_0$, $X_{\bullet}(\omega)$ is Hölder continuous with exponent γ, and so in particular uniformly continuous. To define Y fix some point $x_0 \in S$, given $s \in [0,1]^d$ find $D \ni s_j \to s$ and define

$$Y_s(\omega) = \begin{cases} \lim_{j\to\infty} X_{s_j}(\omega), & \omega \in \Omega_0 \\ x_0, & \omega \notin \Omega_0. \end{cases}$$

The proof of Lemma A.3 shows that the definition of $Y_s(\omega)$ does not depend on the sequence s_j chosen and $s \mapsto Y_s(\omega)$ is continuous for $\omega \in \Omega_0$. For $\omega \notin \Omega_0$ this path is continuous by virtue of being a constant. Also, $Y_s(\omega) = X_s(\omega)$ for $\omega \in \Omega_0$ and $s \in D$. The Hölder continuity of Y follows from (B.8) by letting s and r converge to points in $[0,1]^d$. This completes the proof of part (a) of the theorem.

For part (b) construct Y as in part (a) from the restriction $\{X_s : s \in D\}$. To see that $P\{Y_s = X_s\} = 1$ also for $s \notin D$, pick again a sequence $D \ni s_j \to s$, $\varepsilon > 0$ and write

$$\begin{aligned} P\{\rho(X_s, Y_s) \ge \varepsilon\} \le P\{\rho(X_s, X_{s_j}) \ge \varepsilon/3\} &+ P\{\rho(X_{s_j}, Y_{s_j}) \ge \varepsilon/3\} \\ &+ P\{\rho(Y_{s_j}, Y_s) \ge \varepsilon/3\}. \end{aligned}$$

As $j \to \infty$, on the right the first probability vanishes by the moment assumption and Chebyshev's inequality:

$$P\{\rho(X_s, X_{s_j}) \ge \varepsilon/3\} \le 3^{\beta}\varepsilon^{-\beta}K|s - s_j|^{d+\alpha} \to 0.$$

The last probability vanishes as $j \to \infty$ because $\rho(Y_{s_j}(\omega), Y_s(\omega)) \to 0$ for all ω by the continuity of Y. The middle probability equals zero because $P\{Y_{s_j} = X_{s_j}\} = 1$.

In conclusion, $P\{\rho(X_s, Y_s) \ge \varepsilon\} = 0$ for every $\varepsilon > 0$. This implies that with probability 1, $\rho(X_s, Y_s) = 0$ or equivalently $X_s = Y_s$. ∎

B.2. Construction of Brownian motion

We construct here a one-dimensional standard Brownian motion by constructing its probability distribution on the "canonical" path space $C = C_{\mathbf{R}}[0,\infty)$. Let $B_t(\omega) = \omega(t)$ be the coordinate projections on C, and $\mathcal{F}_t^B = \sigma\{B_s : 0 \le s \le t\}$ the filtration generated by the coordinate process.

Theorem B.21. *There exists a Borel probability measure P^0 on $C = C_{\mathbf{R}}[0,\infty)$ such that the process $B = \{B_t : 0 \le t < \infty\}$ on the probability space $(C, \mathcal{B}_C, P^0)$ is a standard one-dimensional Brownian motion with respect to the filtration $\{\mathcal{F}_t^B\}$.*

The construction gives us the following regularity property of paths. Fix $0 < \gamma < \frac{1}{2}$. For P^0-almost every $\omega \in C$,

$$\sup_{0\le s<t\le T} \frac{|B_t(\omega) - B_s(\omega)|}{|t-s|^\gamma} < \infty \quad \text{for all } T < \infty. \tag{B.9}$$

The proof relies on the Kolmogorov Extension Theorem 1.28. We do not directly construct the measure on C. Instead, we first construct the process on positive dyadic rational time points

$$\mathbf{Q}_2 = \Big\{ \frac{k}{2^n} : k, n \in \mathbf{N} \Big\}.$$

Then we apply an important theorem, the Kolmogorov-Centsov criterion, to show that the process has a unique continuous extension from $\mathbf{Q}_2$ to $[0,\infty)$ that satisfies the Hölder property. The distribution of this extension will be the Wiener measure on C.

Let

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\Big\{ -\frac{x^2}{2t} \Big\} \tag{B.10}$$

be the density of the normal distribution with mean zero and variance t. This function is also called the *Gaussian kernel*. For an increasing n-tuple of positive times $0 < t_1 < t_2 < \cdots < t_n$, let $\mathbf{t} = (t_1, t_2, \ldots, t_n)$. We shall write $\mathbf{x} = (x_1, \ldots, x_n)$ for vectors in $\mathbf{R}^n$, and abbreviate $d\mathbf{x} = dx_1\,dx_2 \cdots dx_n$ to denote integration with respect to Lebesgue measure on $\mathbf{R}^n$. Define a probability measure $\mu_{\mathbf{t}}$ on $\mathbf{R}^n$ by

$$\mu_{\mathbf{t}}(A) = \int_{\mathbf{R}^n} \mathbf{1}_A(\mathbf{x})\, p_{t_1}(x_1) \prod_{i=2}^n p_{t_i - t_{i-1}}(x_i - x_{i-1})\, d\mathbf{x} \tag{B.11}$$

for $A \in \mathcal{B}_{\mathbf{R}^n}$. Before proceeding further, we check that this definition is the right one, namely that $\mu_{\mathbf{t}}$ is the distribution we want for the vector $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$.

Lemma B.22. *If a one-dimensional standard Brownian motion B exists, then for $A \in \mathcal{B}_{\mathbf{R}^n}$*

$$P\big\{(B_{t_1}, B_{t_2}, \ldots, B_{t_n}) \in A\big\} = \mu_{\mathbf{t}}(A).$$

Proof. Define the nonsingular linear map

$$T(y_1, y_2, y_3, \ldots, y_n) = (y_1, y_1 + y_2, y_1 + y_2 + y_3, \ldots, y_1 + y_2 + \cdots + y_n)$$

with inverse

$$T^{-1}(x_1, x_2, x_3, \ldots, x_n) = (x_1, x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1}).$$

In the next calculation, use the fact that the Brownian increments are independent and $B_{t_i} - B_{t_{i-1}}$ has density $p_{t_i - t_{i-1}}$.

$$\begin{aligned}
&P\big\{(B_{t_1}, B_{t_2}, B_{t_3}, \ldots, B_{t_n}) \in A\big\} \\
&= P\big\{T(B_{t_1}, B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{n-1}}) \in A\big\} \\
&= P\big\{(B_{t_1}, B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_n} - B_{t_{n-1}}) \in T^{-1}A\big\} \\
&= \int_{\mathbf{R}^n} \mathbf{1}_A(T\mathbf{y}) p_{t_1}(y_1) p_{t_2 - t_1}(y_2) \cdots p_{t_n - t_{n-1}}(y_n)\, d\mathbf{y}.
\end{aligned}$$

Now change variables through $x = Ty \Leftrightarrow y = T^{-1}x$. T has determinant one, so by (A.18) the integral above becomes

$$\int_{\mathbf{R}^n} \mathbf{1}_A(\mathbf{x}) p_{t_1}(x_1) p_{t_2 - t_1}(x_2 - x_1) \cdots p_{t_n - t_{n-1}}(x_n - x_{n-1})\, d\mathbf{x} = \mu_{\mathbf{t}}(A).$$

□

To apply Kolmogorov's Extension Theorem we need to define consistent finite-dimensional distributions. For an n-tuple $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ of distinct elements from $\mathbf{Q}_2$ let π be the permutation that orders it. In other words, π is the bijection on $\{1, 2, \ldots, n\}$ determined by

$$s_{\pi(1)} < s_{\pi(2)} < \cdots < s_{\pi(n)}.$$

(In this section all n-tuples of time points have distinct entries.) Define the probability measure $Q_{\mathbf{s}}$ on $\mathbf{R}^n$ by $Q_{\mathbf{s}} = \mu_{\pi\mathbf{s}} \circ \pi$, or in terms of integrals of bounded Borel functions,

$$\int_{\mathbf{R}^n} f\, dQ_{\mathbf{s}} = \int_{\mathbf{R}^n} f(\pi^{-1}(\mathbf{x}))\, d\mu_{\pi\mathbf{s}}.$$

Let us convince ourselves again that this definition is the one we want. $Q_{\mathbf{s}}$ should represent the distribution of the vector $(B_{s_1}, B_{s_2}, \ldots, B_{s_n})$, and indeed this follows from Lemma B.22:

$$\begin{aligned}\int_{\mathbf{R}^n} f(\pi^{-1}(\mathbf{x}))\,d\mu_{\pi\mathbf{s}} &= E\big[(f\circ\pi^{-1})(B_{s_{\pi(1)}}, B_{s_{\pi(2)}}, \dots, B_{s_{\pi(n)}})\big]\\ &= E\big[f(B_{s_1}, B_{s_2}, \dots, B_{s_n})\big].\end{aligned}$$

For the second equality above, apply to $y_i = B_{s_i}$ the identity

$$\pi^{-1}(y_{\pi(1)}, y_{\pi(2)}, \dots, y_{\pi(n)}) = (y_1, y_2, \dots, y_n)$$

which is a consequence of the way we defined the action of a permutation on a vector in Section 1.2.4.

Let us check the consistency properties (i) and (ii) required for the Extension Theorem 1.28. Suppose $\mathbf{t} = \rho\mathbf{s}$ for two distinct n-tuples $\mathbf{s}$ and $\mathbf{t}$ from $\mathbf{Q}_2$ and a permutation ρ. If π orders t then $\pi\circ\rho$ orders $\mathbf{s}$, because $t_{\pi(1)} < t_{\pi(2)}$ implies $s_{\rho(\pi(1))} < s_{\rho(\pi(2))}$. One must avoid confusion over how the action of permutations is composed:

$$\pi(\rho\mathbf{s}) = \big(s_{\rho(\pi(1))}, s_{\rho(\pi(2))}, \dots, s_{\rho(\pi(n))}\big)$$

because $\big(\pi(\rho\mathbf{s})\big)_i = (\rho\mathbf{s})_{\pi(i)} = \mathbf{s}_{\rho(\pi(i))}$. Then

$$Q_{\mathbf{s}}\circ\rho^{-1} = \big(\mu_{\pi(\rho\mathbf{s})}\circ(\pi\circ\rho)\big)\circ\rho^{-1} = \mu_{\pi\mathbf{t}}\circ\pi = Q_{\mathbf{t}}.$$

This checks (i). Property (ii) will follow from this lemma.

Lemma B.23. *Let $\mathbf{t} = (t_1, \dots, t_n)$ be an ordered n-tuple , and let $\hat{\mathbf{t}} = (t_1, \dots, t_{j-1}, t_{j+1}, \dots, t_n)$ be the $(n-1)$-tuple obtained by removing t_j from $\mathbf{t}$. Then for $A \in \mathcal{B}_{\mathbf{R}^{j-1}}$ and $B \in \mathcal{B}_{\mathbf{R}^{n-j}}$,*

$$\mu_{\mathbf{t}}(A\times\mathbf{R}\times B) = \mu_{\hat{\mathbf{t}}}(A\times B).$$

Proof. A basic analytic property of Gaussian densities is the convolution identity

$$(p_s * p_t)(x) = p_{s+t}(x), \tag{B.12}$$

where the *convolution* of two densities f and g is in general defined by

$$(f*g)(x) = \int_{\mathbf{R}} f(y)g(x-y)\,dy.$$

We leave checking this to the reader. The corresponding probabilistic property is that the sum of independent normal variables is again normal, and the means and variances add.

The conclusion of the lemma follows from a calculation. Abbreviate $\mathbf{x}' = (x_1, \dots, x_{j-1})$, $\mathbf{x}'' = (x_{j+1}, \dots, x_n)$ and $\hat{\mathbf{x}} = (\mathbf{x}', \mathbf{x}'')$. It is also convenient to use $t_0 = 0$ and $x_0 = 0$.

$$\mu_{\mathbf{t}}(A \times \mathbf{R} \times B) = \int_A d\mathbf{x}' \int_{\mathbf{R}} dx_j \int_B d\mathbf{x}'' \prod_{i=1}^{n} p_{t_i - t_{i-1}}(x_i - x_{i-1})$$
$$= \int_{A\times B} d\hat{\mathbf{x}} \prod_{i \neq j, j+1} p_{t_i - t_{i-1}}(x_i - x_{i-1})$$
$$\times \int_{\mathbf{R}} dx_j \, p_{t_j - t_{j-1}}(x_j - x_{j-1}) p_{t_{j+1} - t_j}(x_{j+1} - x_j).$$

After a change of variables $y = x_j - x_{j-1}$, the interior integral becomes

$$\int_{\mathbf{R}} dy \, p_{t_j - t_{j-1}}(y) p_{t_{j+1} - t_j}(x_{j+1} - x_{j-1} - y) = p_{t_j - t_{j-1}} * p_{t_{j+1} - t_j}(x_{j+1} - x_{j-1})$$
$$= p_{t_{j+1} - t_{j-1}}(x_{j+1} - x_{j-1}).$$

Substituting this back up gives

$$\int_{A\times B} d\hat{\mathbf{x}} \prod_{i \neq j, j+1}^{n} p_{t_i - t_{i-1}}(x_i - x_{i-1}) \cdot p_{t_{j+1} - t_{j-1}}(x_{j+1} - x_{j-1})$$
$$= \mu_{\hat{\mathbf{t}}}(A \times B). \qquad \square$$

As $A \in \mathcal{B}_{\mathbf{R}^{j-1}}$ and $B \in \mathcal{B}_{\mathbf{R}^{n-j}}$ range over these σ-algebras, the class of product sets $A \times B$ generates $\mathcal{B}_{\mathbf{R}^{n-1}}$ and is also closed under intersections. Consequently by Lemma B.5, the conclusion of the above lemma generalizes to

$$\text{(B.13)} \qquad \mu_{\mathbf{t}}\{\mathbf{x} \in \mathbf{R}^n : \hat{\mathbf{x}} \in G\} = \mu_{\hat{\mathbf{t}}}(G) \quad \text{for all } G \in \mathcal{B}_{\mathbf{R}^{n-1}}.$$

We are ready to check requirement (ii) of Theorem 1.28. Let $\mathbf{t} = (t_1, t_2, \dots, t_{n-1}, t_n)$ be an n-tuple, $\mathbf{s} = (t_1, t_2, \dots, t_{n-1})$ and $A \in \mathcal{B}_{\mathbf{R}^{n-1}}$. We need to show $Q_{\mathbf{s}}(A) = Q_{\mathbf{t}}(A \times \mathbf{R})$. Suppose π orders $\mathbf{t}$. Let j be the index such that $\pi(j) = n$. Then $\mathbf{s}$ is ordered by the permutation

$$\sigma(i) = \begin{cases} \pi(i), & i = 1, \dots, j-1 \\ \pi(i+1), & i = j, \dots, n-1. \end{cases}$$

A few observations before we compute: σ is a bijection on $\{1, 2, \dots, n-1\}$ as it should. Also

$$\sigma(x_{\pi^{-1}(1)}, x_{\pi^{-1}(2)}, \dots, x_{\pi^{-1}(n-1)}) = (x_{\pi^{-1}(\sigma(1))}, x_{\pi^{-1}(\sigma(2))}, \dots, x_{\pi^{-1}(\sigma(n-1))})$$
$$= (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) = \hat{\mathbf{x}}$$

where the notation $\hat{\mathbf{x}}$ is used as in the proof of the previous lemma. And lastly, using $\widehat{\pi\mathbf{t}}$ to denote the omission of the jth coordinate,

$$\begin{aligned}\widehat{\pi\mathbf{t}} &= (t_{\pi(1)}, \dots, t_{\pi(j-1)}, t_{\pi(j+1)}, \dots, t_{\pi(n)})\\ &= (t_{\sigma(1)}, \dots, t_{\sigma(j-1)}, t_{\sigma(j)}, \dots, t_{\sigma(n-1)})\\ &= \sigma\mathbf{s}.\end{aligned}$$

Equation (B.13) says that the distribution of $\hat{\mathbf{x}}$ under $\mu_{\pi\mathbf{t}}$ is $\mu_{\widehat{\pi\mathbf{t}}}$. From these ingredients we get

$$\begin{aligned}Q_{\mathbf{t}}(A \times \mathbf{R}) &= \mu_{\pi\mathbf{t}}(\pi(A \times \mathbf{R}))\\ &= \mu_{\pi\mathbf{t}}\{\mathbf{x} \in \mathbf{R}^n : (x_{\pi^{-1}(1)}, x_{\pi^{-1}(2)}, \dots, x_{\pi^{-1}(n-1)}) \in A\}\\ &= \mu_{\pi\mathbf{t}}\{\mathbf{x} \in \mathbf{R}^n : \hat{\mathbf{x}} \in \sigma A\}\\ &= \mu_{\widehat{\pi\mathbf{t}}}(\sigma A) = \mu_{\sigma\mathbf{s}}(\sigma A) = Q_{\mathbf{s}}(A).\end{aligned}$$

We have checked the hypotheses of Kolmogorov's Extension Theorem for the family $\{Q_{\mathbf{t}}\}$.

Let $\Omega_2 = \mathbf{R}^{\mathbf{Q}_2}$ with its product σ-algebra $\mathcal{G}_2 = \mathcal{B}(\mathbf{R})^{\otimes \mathbf{Q}_2}$. Write $\xi = (\xi_t : t \in \mathbf{Q}_2)$ for a generic element of Ω_2. By Theorem 1.28 there is a probability measure Q on $(\Omega_2, \mathcal{G}_2)$ with marginals

$$Q\big\{\xi \in \Omega_2 : (\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_n}) \in B\big\} = Q_{\mathbf{t}}(B)$$

for n-tuples $\mathbf{t} = (t_1, t_2, \dots, t_n)$ of distinct elements from $\mathbf{Q}_2$ and $B \in \mathcal{B}_{\mathbf{R}^n}$. We write E^Q for expectation under the measure Q.

So far we have not included the time origin in the index set $\mathbf{Q}_2$. This was for reasons of convenience. Because B_0 has no density, to include $t_0 = 0$ would have required two formulas for $\mu_{\mathbf{t}}$ in (B.11), one for n-tuples with zero, the other for n-tuples without zero. Now define $\mathbf{Q}_2^0 = \mathbf{Q}_2 \cup \{0\}$. On Ω_2 define random variables $\{X_q : q \in \mathbf{Q}_2^0\}$ by

$$X_0(\xi) = 0, \text{ and } X_q(\xi) = \xi_q \text{ for } q \in \mathbf{Q}_2. \tag{B.14}$$

The second step of our construction of Brownian motion is the proof that the process $\{X_q\}$ is uniformly continuous on bounded index sets. This comes from an application of the Kolmogorov-Centsov criterion, Theorem B.20.

Lemma B.24. *Let $\{X_q\}$ be the process defined by (B.14) on the probability space $(\Omega_2, \mathcal{G}_2, Q)$, where Q is the probability measure whose existence came from Kolmogorov's Extension Theorem. Let $0 < \gamma < \frac{1}{2}$. Then there is an event Γ such that $Q(\Gamma) = 1$ with this property: for all $\xi \in \Gamma$ and $T < \infty$, there exists a finite constant $C_T(\xi)$ such that*

$$|X_s(\xi) - X_r(\xi)| \le C_T(\xi)|s - r|^\gamma \quad \textit{for all } r, s \in \mathbf{Q}_2^0 \cap [0, T]. \tag{B.15}$$

In particular, for $\xi \in \Gamma$ the function $q \mapsto X_q(\xi)$ is uniformly continuous on $\mathbf{Q}_2^0 \cap [0,T]$ for every $T < \infty$.

Proof. We apply Theorem B.20 to an interval $[0,T]$. Though Theorem B.20 was written for the unit cube $[0,1]^d$ it applies obviously to any closed bounded interval or rectangle, via a suitable transformation of the index s.

We need to check the hypothesis (B.5). Due to the definition of the finite-dimensional distributions of Q, this reduces to computing a moment of the Gaussian distribution. Fix an integer $m \geq 2$ large enough so that $\frac{1}{2} - \frac{1}{2m} > \gamma$. Let $0 \leq q < r$ be indices in $\mathbf{Q}_2^0$. In the next calculation, note that after changing variables in the dy_2-integral it no longer depends on y_1, and the y_1-variable can be integrated away.

$$\begin{aligned}
E^Q\big[(X_r - X_q)^{2m}\big] &= \iint_{\mathbf{R}^2} (y_2 - y_1)^{2m} p_q(y_1) p_{r-q}(y_2 - y_1)\, dy_1\, dy_2 \\
&= \int_{\mathbf{R}} dy_1\, p_q(y_1) \int_{\mathbf{R}} dy_2\, (y_2 - y_1)^{2m} p_{r-q}(y_2 - y_1) \\
&= \int_{\mathbf{R}} dy_1\, p_q(y_1) \int_{\mathbf{R}} dx\, x^{2m} p_{r-q}(x) = \int_{\mathbf{R}} dx\, x^{2m} p_{r-q}(x) \\
&= \frac{1}{\sqrt{2\pi(r-q)}} \int_{\mathbf{R}} x^{2m} \exp\Big\{-\frac{x^2}{2(r-q)}\Big\}\, dx \\
&= (r-q)^m \int_{\mathbf{R}} z^{2m} \exp\Big\{-\frac{z^2}{2}\Big\}\, dz = C_m |r-q|^m,
\end{aligned}$$

where $C_m = 1 \cdot 3 \cdot 5 \cdots (2m-1)$, the product of the odd integers less than $2m$. We have verified the hypothesis (B.5) for the values $\alpha = m - 1$ and $\beta = 2m$, and by choice of m,

$$0 < \gamma < \alpha/\beta = \tfrac{1}{2} - \tfrac{1}{2m}.$$

Theorem B.20 now implies the following. For each $T < \infty$ there exists an event $\Gamma_T \subseteq \Omega_2$ such that $Q(\Gamma_T) = 1$ and for every $\xi \in \Gamma_T$ there exists a finite constant $C(\xi)$ such that

$$|X_r(\xi) - X_q(\xi)| \leq C(\xi)|r-q|^\gamma \tag{B.16}$$

for any $q, r \in [0,T] \cap \mathbf{Q}_2^0$. Take

$$\Gamma = \bigcap_{T=1}^{\infty} \Gamma_T.$$

Then $Q(\Gamma) = 1$ and each $\xi \in \Gamma$ has the required property. ☐

With the local Hölder property (B.15) in hand, we can now extend the definition of the process X to the entire time line $[0, \infty)$ while preserving the local Hölder property, as was done for part (b) in Theorem B.20. The value $X_t(\xi)$ for any $t \notin \mathbf{Q}_2^0$ can be defined by

$$X_t(\xi) = \lim_{i\to\infty} X_{q_i}(\xi)$$

for any sequence $\{q_i\}$ from $\mathbf{Q}_2^0$ such that $q_i \to t$. This tells us that the random variables $\{X_t : 0 \le t < \infty\}$ are measurable on Γ. To have a continuous process X_t defined on all of Ω_2 set

$$X_t(\xi) = 0 \quad \text{for } \xi \notin \Gamma \text{ and all } t \ge 0.$$

The process $\{X_t\}$ is a one-dimensional standard Brownian motion. The initial value $X_0 = 0$ has been built into the definition, and we have the path continuity. It remains to check finite-dimensional distributions. Fix $0 < t_1 < t_2 < \cdots < t_n$. Pick points $0 < q_1^k < q_2^k < \cdots < q_n^k$ from $\mathbf{Q}_2^0$ so that $q_i^k \to t_i$ as $k \to \infty$, for $1 \le i \le n$. Pick them so that for some $\delta > 0$, $\delta \le q_i^k - q_{i-1}^k \le \delta^{-1}$ for all i and k. Let ϕ be a bounded continuous function on $\mathbf{R}^n$. Then by the path-continuity (use again $t_0 = 0$ and $x_0 = 0$)

$$\begin{aligned}\int_{\Omega_2} \phi(X_{t_1}, X_{t_2}, \dots, X_{t_n})\, dQ &= \lim_{k\to\infty} \int_{\Omega_2} \phi(X_{q_1^k}, X_{q_2^k}, \dots, X_{q_n^k})\, dQ \\ &= \lim_{k\to\infty} \int_{\mathbf{R}^n} \phi(\mathbf{x}) \prod_{i=1}^n p_{q_i^k - q_{i-1}^k}(x_i - x_{i-1})\, d\mathbf{x} \\ &= \int_{\mathbf{R}^n} \phi(\mathbf{x}) \prod_{i=1}^n p_{t_i - t_{i-1}}(x_i - x_{i-1})\, d\mathbf{x} = \int_{\mathbf{R}^n} \phi(\mathbf{x})\, \mu_{\mathbf{t}}(d\mathbf{x}).\end{aligned}$$

The second-last equality above is a consequence of dominated convergence. An $L^1(\mathbf{R}^n)$ integrable bound can be gotten by

$$p_{q_i^k - q_{i-1}^k}(x_i - x_{i-1}) = \frac{\exp\Bigl\{-\frac{(x_i - x_{i-1})^2}{2(q_i^k - q_{i-1}^k)}\Bigr\}}{\sqrt{2\pi(q_i^k - q_{i-1}^k)}} \le \frac{\exp\Bigl\{-\frac{(x_i - x_{i-1})^2}{2\delta^{-1}}\Bigr\}}{\sqrt{2\pi\delta}}.$$

Comparison with Lemma B.22 shows that $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ has the distribution that Brownian motion should have. An application of Lemma B.6 is also needed here, to guarantee that it is enough to check continuous functions ϕ. It follows that $\{X_t\}$ has independent increments, because this property is built into the definition of the distribution $\mu_{\mathbf{t}}$.

To complete the construction of Brownian motion and finish the proof of Theorem 2.27, we define the measure P^0 on C as the distribution of the process $X = \{X_t\}$:

$$P^0(B) = Q\{\xi \in \Omega_2 : X(\xi) \in A\} \quad \text{for } A \in \mathcal{B}_C.$$

For this, X has to be a measurable map from Ω_2 into C. This follows from Exercise 1.8(b) because $\mathcal{B}_C$ is generated by the projections $\pi_t : \omega \mapsto$

$\omega(t)$. The compositions of projections with X are precisely the coordinates $\pi_t \circ X = X_t$ which are measurable functions on Ω_2. The coordinate process $B = \{B_t\}$ on C has the same distribution as X by the definition of P^0:

$$P^0\{B \in H\} = P^0(H) = Q\{X \in H\} \quad \text{for } H \in \mathcal{B}_C.$$

We also need to check that B has the correct relationship with the filtration $\{\mathcal{F}_t^B\}$, as required by Definition 2.26. B is adapted to $\{\mathcal{F}_t^B\}$ by construction. The independence of $B_t - B_s$ and $\mathcal{F}_s^B$ follows from two facts: B inherits independent increments from X (independence of increments is a property of finite-dimensional distributions), and the increments $\{B_v - B_u : 0 \le u < v \le s\}$ generate $\mathcal{F}_s^B$. This completes the proof of Theorem 2.27.

Hölder continuity (2.34) follows from the equality in distribution of the processes B and X, and because the Hölder property of X was built into the construction through the Kolmogorov-Centsov theorem.

Bibliography

[1] R. F. Bass. The measurability of hitting times. *Electron. Commun. Probab.*, 15:99–105, 2010.

[2] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[3] K. L. Chung and R. J. Williams. *Introduction to stochastic integration.* Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, second edition, 1990.

[4] R. M. Dudley. *Real analysis and probability.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.

[5] R. Durrett. *Probability: theory and examples.* Duxbury Advanced Series. Brooks/Cole–Thomson, Belmont, CA, third edition, 2004.

[6] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[7] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 1998.

[8] G. B. Folland. *Real analysis: Modern techniques and their applications.* Pure and Applied Mathematics. John Wiley & Sons Inc., New York, second edition, 1999.

[9] K. Itô and H. P. McKean, Jr. *Diffusion processes and their sample paths.* Springer-Verlag, Berlin, 1974. Second printing, corrected, Die Grundlehren der mathematischen Wissenschaften, Band 125.

[10] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[11] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991.

[12] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.

[13] S. Resnick. *Adventures in stochastic processes.* Birkhäuser Boston Inc., Boston, MA, 1992.

[14] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[15] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991.

[16] K. R. Stromberg. *Introduction to classical real analysis*. Wadsworth International, Belmont, Calif., 1981. Wadsworth International Mathematics Series.

[17] J. B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.

Notation and Conventions

Mathematicians tend to use parentheses ( ), braces { } and square brackets [ ] interchangeably in many places. But there are situations where the delimiters have specific technical meanings and they cannot be used carelessly. One example is (a, b) for an ordered pair or an open interval, and $[a, b]$ for a closed interval. Then there are situations where one type is conventionally preferred and deviations are discouraged: for example $f(x)$ for a function value and $\{x : f(x) \in B\}$ for sets. An example were all three forms are acceptable is the probability of the event $X \in B$: $P(X \in B)$, $P[X \in B]$ or $P\{X \in B\}$.

The time variable of a stochastic process can be in two places: X_t or $X(t)$. No distinction is meant between the two. For example B_t and $B(t)$ both denote Brownian motion at time t.

With expectations there is the question of the meaning of EX^2. Does the square come before or after the expectation? The notation in this book follows these conventions:

$$EX^2 = E(X^2) \quad \text{while} \quad E[X]^2 = E(X)^2 = \big(E[X]\big)^2.$$

When integrating a function of time and space over $[0, t] \times B$, there is a choice between

$$\iint_{[0,t]\times B} f(s, x)\, ds\, dx \quad \text{and} \quad \int_{[0,t]\times B} f(s, x)\, ds\, dx.$$

While the former is more familiar from calculus, we opt for the latter to avoid proliferating integral signs. Since $ds\,dx$ can be taken to mean product measure on $[0,t] \times B$ (Fubini's theorem), $\int$ makes just as much sense as $\iint$.

Vectors are sometimes boldface: $\mathbf{u} = (u_1, u_2, \dots, u_n)$. In multivariate integrals we can abbreviate $d\mathbf{u} = du_{1,n} = du_1\,du_2 \cdots du_n$.

A dot can be a placeholder for a variable. For example, a stochastic process $\{X_t\}$ indexed by time can be abbreviated as $X = \{X_t\}$ or as $X_{\boldsymbol{\cdot}} = \{X_t\}$. The latter will be used when omission of the dot may lead to ambiguity.

$B_r(x)$ is the open ball of radius r centered at point x in a metric space.

$C^2(D)$ is a function space introduced on page 208.

$C^{1,2}([0,T] \times D)$ is a function space introduced on page 212.

$\delta_{i,j} = \mathbf{1}\{i = j\} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$ is sometimes called the Dirac delta.

$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$ is the point mass at x, or the probability measure that puts all its mass on the point x.

$\triangle$ denotes the symmetric difference of two sets:

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cap B^c) \cup (B \cap A^c).$$

$\Delta f(x) = f_{x_1,x_1}(x) + \cdots + f_{x_d,x_d}(x)$ is the Laplace operator.

$f(t-) = \lim_{s \nearrow t,\, s<t} f(s)$ is the left limit of f at t.

$\Delta Z(t) = Z(t) - Z(t-)$ is the jump at t of the cadlag process Z.

iff is short for *if and only if*.

$\mathcal{L}_2 = \mathcal{L}_2(M, \mathcal{P})$ is the space of predictable processes locally L^2 under μ_M.

$\mathcal{L}(M, \mathcal{P})$ is the space of predictable processes locally in $\mathcal{L}_2(M, \mathcal{P})$, Definition 5.21.

$\mathcal{L}_2(W)$ is the space of locally square-integrable adapted processes that can be integrated by white noise, page 320.

m is Lebesgue measure, Section 1.1.2.

$\mathcal{M}_1(\Omega)$ is the space of Borel probability measures on the metric space Ω.

$\mathcal{M}_2$ is the space of cadlag L^2-martingales.

$\mathcal{M}_{2,\mathrm{loc}}$ is the space of cadlag, local L^2-martingales.

$\mu_M(A) = E \int_{[0,\infty)} \mathbf{1}_A(t,\omega)\, d[M]_t(\omega)$ is the Doléans measure of a square-integrable cadlag martingale M.

$\mathbf{N} = \{1, 2, 3, \dots\}$ is the set of positive integers.

$p(t, x, y) = p_t(y - x) = \dfrac{1}{\sqrt{2\pi t}} \exp\Big\{-\dfrac{(y-x)^2}{2t}\Big\}$ is the Gaussian kernel;

if $x, y \in \mathbf{R}^d$ then $\quad p(t, x, y) = (2\pi t)^{-\frac{d}{2}} \exp\Big\{-\dfrac{1}{2t}|y - x|^2\Big\}$.

$\mathcal{P}$ is the predictable σ-field.

$\mathbf{R}_+ = [0, \infty)$ is the set of nonnegative real numbers.

$\mathcal{S}_2$ is the space of simple predictable processes.

V_F is the total variation of a function F, Section 1.1.9.

$X^*_T(\omega) = \sup\limits_{0 \le t \le T} |X_t(\omega)|$.

$X^\tau_t = X_{\tau \wedge t}$ defines the stopped process X^τ.

Z_- is the left limit process $Z_-(t) = Z(t-)$.

$\mathbf{Z}_+ = \{0, 1, 2, \dots\}$ is the set of nonnegative integers.

Index