Stochastic integration

P.J.C. Spreij

this version: January 19, 2018

Contents

1 Stochastic processes

In this section we review some fundamental facts from the general theory of stochastic processes.

1.1 General theory

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We will use a set of time instants I. When this set is not specified, it will be $[0, \infty)$, occasionally $[0, \infty]$, or $\mathbb{N}$. Let $(E, \mathcal{E})$ another measurable space. If the set E is endowed with a metric, the Borel σ-algebra that is generated by this metric will be denoted by $\mathcal{B}(E)$ or $\mathcal{B}E$. Mostly E will be $\mathbb{R}$ or $\mathbb{R}^d$ and $\mathcal{E}$ the ordinary Borel σ-algebra on it.

Definition 1.1 A random element of E is a map from Ω into E that is $\mathcal{F}/\mathcal{E}$-measurable. A stochastic process X with time set I is a collection $\{X_t, t \in I\}$ of random elements of E. For each ω the map $t \mapsto X_t(\omega)$ is called a (sample) path, trajectory or realization of X.

Since we will mainly encounter processes where $I = [0, \infty)$, we will discuss processes whose paths are continuous, or right-continuous, or *càdlàg*. The latter means that all paths are right-continuous functions with finite left limits at each $t > 0$. We will also encounter processes that satisfy these properties almost surely.

Often we have to specify in which sense two (stochastic) processes are the same. The following concepts are used.

Definition 1.2 Two *real valued or* $\mathbb{R}^d$ *valued* processes X and Y are called *indistinguishable* if the set $\{X_t = Y_t, \forall t \in I\}$ contains a set of probability one (hence the paths of indistinguishable processes are a.s. equal). They are called *modifications* of each other if $\mathbb{P}(X_t = Y_t) = 1$, for all $t \in I$. The processes are said to have the same finite dimensional distributions if for any n-tuple $(t_1, \ldots, t_n)$ with the $t_i \in I$ the laws of the random vectors $(X_{t_1}, \ldots, X_{t_n})$ and $(Y_{t_1}, \ldots, Y_{t_n})$ coincide.

Clearly, the first of these three concepts is the strongest, the last the weakest. Whereas the first two definitions only make sense for processes defined on the same probability space, for the last one this is not necessary.

Example 1.3 Let T be a nonnegative real random variable with a continuous distribution. Let $X = 0$ and Y be defined by $Y_t(\omega) = \mathbf{1}_{t=T(\omega)}$, $t \in [0, \infty)$. Then X is a modification of Y, whereas $\mathbb{P}(X_t = Y_t, \forall t \geq 0) = 0$.

Proposition 1.4 *Let Y be a modification of X and assume that all paths of X and Y are right-continuous. Then X and Y are indistinguishable.*

Proof Right-continuity allows us to write $\{X_t = Y_t, \forall t \geq 0\} = \{X_t = Y_t, \forall t \in [0, \infty) \cap \mathbb{Q}\}$. Since Y is a modification of X, the last set (is measurable and) has probability one. $\square$

Throughout the course we will need various measurability properties of stochastic processes. Viewing a process X as a map from $[0,\infty) \times \Omega$ into E, we call this process measurable if $X^{-1}(A)$ belongs to $\mathcal{B}[0,\infty) \times \mathcal{F}$ for all $A \in \mathcal{B}(E)$.

Definition 1.5 A filtration $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ is a collection of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$. We put $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, t \geq 0)$. Given a stochastic process X we denote by $\mathcal{F}_t^X$ the smallest σ-algebra for which all X_s, with $s \leq t$, are measurable and $\mathbb{F}^X = \{\mathcal{F}_t^X, t \geq 0\}$.
Given a filtration $\mathbb{F}$ for $t \geq 0$ the σ-algebras $\mathcal{F}_{t+}$ and $\mathcal{F}_{t-}$ for $t > 0$ are defined as follows. $\mathcal{F}_{t+} = \bigcap_{h>0} \mathcal{F}_{t+h}$ and $\mathcal{F}_{t-} = \sigma(\mathcal{F}_{t-h}, h > 0)$. We will call a filtration $\mathbb{F}$ right-continuous if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$, left-continuous if $\mathcal{F}_t = \mathcal{F}_{t-}$ for all $t > 0$ and continuous if it is both left- and right-continuous. A filtration is said to satisfy the *usual conditions* if it is right-continuous and if $\mathcal{F}_0$ contains all $\mathcal{F}$-null sets. We use the notation $\mathbb{F}^+$ for the filtration $\{\mathcal{F}_{t+}, t \geq 0\}$.

Definition 1.6 Given a filtration $\mathbb{F}$ a process X is called $\mathbb{F}$-adapted, adapted to $\mathbb{F}$, or simply adapted, if for all $t \geq 0$ the random variable X_t is $\mathcal{F}_t$-measurable. Clearly, any process X is adapted to $\mathbb{F}^X$. A process X is called *progressive* or *progressively measurable*, if for all $t \geq 0$ the map $(s,\omega) \mapsto X_s(\omega)$ from $[0,t] \times \Omega$ into $\mathbb{R}$ is $\mathcal{B}[0,t] \otimes \mathcal{F}_t$-measurable. A progressive process is always adapted (Exercise 1.11) and measurable.

Proposition 1.7 *Let X be a process that is (left-) or right-continuous. Then it is measurable. Such a process is progressive if it is adapted.*

Proof We will prove the latter assertion only, the first one can be established by a similar argument. Assume that X is right-continuous and fix $t > 0$ (the proof for a left-continuous process is analogous). Then, since X is right-continuous, X is on $[0,t]$ the pointwise limit of the X^n defined as

$$X^n = X_0 \mathbf{1}_{\{0\}}(\cdot) + \sum_{k=1}^{2^n} \mathbf{1}_{((k-1)2^{-n}t, k2^{-n}t]}(\cdot) X_{k2^{-n}t}.$$

It is easy to see that all X^n are $\mathcal{B}[0,t] \times \mathcal{F}_t$-measurable, and so is their limit X. $\square$

1.2 Stopping times

Definition 1.8 A map $T : \Omega \to [0,\infty]$ is a called a random time, if T is a random variable. T is called a *stopping time* (w.r.t. the filtration $\mathbb{F}$, or an $\mathbb{F}$-stopping time) if for all $t \geq 0$ the set $\{T \leq t\} \in \mathcal{F}_t$.

If $T < \infty$, then T is called finite and if $\mathbb{P}(T < \infty) = 1$ it is called a.s. finite. Likewise we say that T is bounded if there is a $K \in [0,\infty)$ such that $T(\omega) \leq K$ and that T is a.s. bounded if for such a K we have $\mathbb{P}(T \leq K) = 1$. For a random time T and a process X one defines on the (measurable) set $\{T < \infty\}$ the function X_T by

$$X_T(\omega) = X_{T(\omega)}(\omega).$$

If next to the process X we also have a random variable X_∞, then X_T is defined on the whole set Ω.

The set of stopping times is closed under many operations.

Proposition 1.9 *Let $S, T, T_1, T_2, \ldots$ be stopping times. Then all random variables $S \vee T$, $S \wedge T$, $S + T$, $\sup_n T_n$ are stopping times. The random variable $\inf_n T_n$ is an $\mathbb{F}^+$-stopping time. If $a > 0$, then $T + a$ is a stopping time as well.*

Proof Exercise 1.5. □

Suppose that E is endowed with a metric d and that the $\mathcal{E}$ is the Borel σ-algebra on E. Let $D \subset E$ and X a process with values in E. The *hitting time* H_D is defined as $H_D = \inf\{t \geq 0 : X_t \in D\}$. Hitting times are stopping times under extra conditions.

Proposition 1.10 *If G is an open set and X a right-continuous $\mathbb{F}$-adapted process, then H_G is an $\mathbb{F}^+$-stopping time. Let F be a closed set and X a càdlàg process. Define $\tilde{H}_F = \inf\{t \geq 0 : X_t \in F \text{ or } X_{t-} \in F\}$, then $\tilde{H}_F$ is an $\mathbb{F}$-stopping time. If X is a continuous process, then H_F is an $\mathbb{F}$-stopping time.*

Proof Notice first that $\{H_G < t\} = \cup_{s<t}\{X_s \in G\}$. Since G is open and X is right-continuous, we may replace the latter union with $\cup_{s<t, s\in\mathbb{Q}}\{X_s \in G\}$. Hence $\{H_G < t\} \in \mathcal{F}_t$ and thus H_G is an $\mathbb{F}^+$-stopping time in view of Exercise 1.4.
Since F is closed it is the intersection $\cap_{n=1}^\infty F^n$ of the open sets $F^n = \{x \in E : d(x, F) < \frac{1}{n}\}$. The event $\{\tilde{H}_F \leq t\}$ can be written as the union of $\{X_t \in F\}$, $\{X_{t-} \in F\}$ and $\cap_{n\geq 1} \cup_{s<t, s\in\mathbb{Q}} \{X_s \in F^n\}$ by an argument similar to the one we used above. The result follows. □

Definition 1.11 For a stopping time T we define $\mathcal{F}_T$ as the collection $\{F \in \mathcal{F}_\infty : F \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}$. Since a constant random variable $T \equiv t_0$ is a stopping time, one has for this stopping time $\mathcal{F}_T = \mathcal{F}_{t_0}$. In this sense the notation $\mathcal{F}_T$ is unambiguous. Similarly, we have $\mathcal{F}_{T+} = \{F \in \mathcal{F} : F \cap \{T \leq t\} \in \mathcal{F}_{t+}, \forall t \geq 0\}$.

Proposition 1.12 *Let S and T be stopping times.*

(i) *For all $A \in \mathcal{F}_S$ it holds that $A \cap \{S \leq T\} \in \mathcal{F}_T$, and if X is $\mathcal{F}_S$-measurable, then $\mathbf{1}_{\{S\leq T\}}X$ is $\mathcal{F}_T$-measurable. In particular, if $S \leq T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

(ii) *$\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S\wedge T}$ and $\{S \leq T\} \in \mathcal{F}_S \cap \mathcal{F}_T$.*

Proof (i) Write $A \cap \{S \leq T\} \cap \{T \leq t\}$ as $A \cap \{S \leq t\} \cap \{S \wedge t \leq T \wedge t\} \cap \{T \leq t\}$. The intersection of the first two sets belongs to $\mathcal{F}_t$ if $A \in \mathcal{F}_S$, the fourth one obviously too. That the third set belongs to $\mathcal{F}_t$ is left as Exercise 1.6. The statement concerning X follows.

(ii) It follows from the first assertion that $\mathcal{F}_S \cap \mathcal{F}_T \supset \mathcal{F}_{S\wedge T}$. Let $A \in \mathcal{F}_S \cap \mathcal{F}_T$. Then $A \cap \{S \wedge T \leq t\} = (A \cap \{S \leq t\}) \cup (A \cap \{T \leq t\})$, and this obviously belongs to $\mathcal{F}_t$. From the previous assertions it follows that $\{S \leq T\} \in \mathcal{F}_T$. But then we also have $\{S > T\} \in \mathcal{F}_T$ and by symmetry $\{S < T\} \in \mathcal{F}_S$. Likewise we have for every $n \in \mathbb{N}$ that $\{S < T + \frac{1}{n}\} \in \mathcal{F}_S$. Taking intersections, we get $\{S \leq T\} \in \mathcal{F}_S$. $\square$

For a stopping time T and a stochastic process X we define the stopped process X^T by $X^T_t = X_{T\wedge t}$, for $t \geq 0$.

Proposition 1.13 *Let T be a stopping time, X a progressive process. Then $Y := X_T \mathbf{1}_{T<\infty}$ is $\mathcal{F}_T$-measurable and the stopped process X^T is progressive too.*

Proof First we show that X^T is progressive. Let $t > 0$. The map $\phi : ([0,t] \times \Omega, \mathcal{B}[0,t] \times \mathcal{F}_t) \to ([0,t] \times \Omega, \mathcal{B}[0,t] \times \mathcal{F}_t)$ defined by $\phi : (s,\omega) \to (T(\omega) \wedge s, \omega)$ is measurable and so is the composition $(s,\omega) \mapsto X(\phi(s,\omega)) = X_{T(\omega)\wedge s}(\omega)$, which shows that X^T is progressive. Fubini's theorem then says that the section map $\omega \mapsto X_{T(\omega)\wedge t}(\omega)$ is $\mathcal{F}_t$-measurable. To show that Y is $\mathcal{F}_T$-measurable we have to show that $\{Y \in B\} \cap \{T \leq t\} = \{X_{T\wedge t} \in B\} \cap \{T \leq t\} \in \mathcal{F}_t$. This has by now become obvious. $\square$

1.3 Exercises

1.1 Let X be a measurable process on $[0,\infty)$. Then the maps $t \mapsto X_t(\omega)$ are Borel-measurable. If $\mathbb{E}\,|X_t| < \infty$ for all t, then also $t \mapsto \mathbb{E}\,X_t$ is measurable and if $\int_0^T \mathbb{E}\,|X_t|\,\mathrm{d}t < \infty$, then $\int_0^T \mathbb{E}\,X_t\,\mathrm{d}t = \mathbb{E}\int_0^T X_t\,\mathrm{d}t$. Prove these statements. Show also that the process $\int_0^{\cdot} X_s\,\mathrm{d}s$ is progressive if X is progressive.

1.2 Let X be a càdlàg adapted process and A the event that X is continuous on an interval $[0,t)$. Then $A \in \mathcal{F}_t$.

1.3 Let X be a measurable process and T a random time. Then $X_T \mathbf{1}_{T<\infty}$ is a random variable.

1.4 A random time T is an $\mathbb{F}^+$-stopping time iff for all $t > 0$ one has $\{T < t\} \in \mathcal{F}_t$.

1.5 Prove Proposition 1.9.

1.6 Let S, T be stopping times bounded by a constant t_0. Then $\sigma(S) \subset \mathcal{F}_{t_0}$ and $\{S \leq T\} \in \mathcal{F}_{t_0}$.

1.7 Let T be a stopping time and S a random time such that $S \geq T$. If S is $\mathcal{F}_T$-measurable, then S is a stopping time too.

1.8 Let $\mathcal{F} = \mathcal{F}_\infty$, S and T be stopping times and X an integrable random variable. Consider the martingale defined by $X_t = \mathbb{E}\,[X|\mathcal{F}_t]$, $t \in [0, \infty]$.

(a) Show that $X_t^T = \mathbb{E}\,[X_T|\mathcal{F}_t]$.

(b) Show that $\mathbb{E}\,[\mathbb{E}\,[X|\mathcal{F}_T]|\mathcal{F}_S] = \mathbb{E}\,[X|\mathcal{F}_{S\wedge T}]$. (It follows that taking conditional expectation w.r.t. $\mathcal{F}_T$ and $\mathcal{F}_S$ are commutative operations. This is in general not true for arbitrary σ-algebras, see Exercise 1.9.)

1.9 Let X_1, X_2, X_3 be *iid* random variables, defined on the same probability space. Let $\mathcal{G} = \sigma(X_1 + X_2)$, $\mathcal{H} = \sigma(X_2 + X_3)$ and $X = X_1 + X_2 + X_3$. Show that $X_{HG} := \mathbb{E}[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}] = \frac{1}{2}(X_1 + X_2)$ and $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \frac{1}{2}(X_2 + X_3)$. (Taking conditional expectation w.r.t. $\mathcal{G}$ and $\mathcal{H}$ does not commute). What happens eventually to iterated conditional expectations $X_{HGH} = \mathbb{E}\,[X_{HG}|\mathcal{H}]$, $X_{HGHG} = \mathbb{E}\,[X_{HGH}|\mathcal{G}]$, etc.?

1.10 Let S and T be stopping times such that $S \leq T$ and even $S < T$ on the set $\{S < \infty\}$. Then $\mathcal{F}_{S+} \subset \mathcal{F}_T$.

1.11 Show that a progressive process is adapted and measurable.

1.12 Show that $\mathbb{F}^+$ is a right-continuous filtration.

1.13 Show that $\mathbb{F}^X$ is a left-continuous filtration if X is a left-continuous process.

1.14 Let X be a process that is adapted to a filtration $\mathbb{F}$. Let Y be a modification of X. Show that also Y is adapted to $\mathbb{F}$ if $\mathbb{F}$ satisfies the usual conditions.

1.15 Show that Proposition 1.13 can be refined as follows. Under the stated assumptions the process X^T is even progressive w.r.t. the filtration $\{\mathcal{F}_{t\wedge T} : t \geq 0\}$.

2 Martingales

In this section we review some properties of martingales and supermartingales. Throughout the section we work with a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F}$. Familiarity with the theory of martingales in discrete time is assumed.

2.1 Generalities

We start with the definition of martingales, submartingales and supermartingales in continuous time.

Definition 2.1 A real-valued process X is called a supermartingale, if it is adapted, if all X_t are integrable and $\mathbb{E}\,[X_t|\mathcal{F}_s] \leq X_s$ a.s. for all $t \geq s$. If $-X$ is a supermartingale, we call X a submartingale and X will be called a martingale if it is both a submartingale and a supermartingale.

The number of upcrossings of a stochastic process X over an interval $[a, b]$ when time runs through a finite set F is denoted by $U([a, b]; F)$. Then we define the number of upcrossings over the interval $[a, b]$ when time runs through a set I by $U([a, b]; I) := \sup\{U([a, b]; F) : F \subset I, F \text{ finite}\}$. Note that $U([a, b]; I)$ is $\mathcal{F}_t$-measurable if X is adapted and $I \subset [0, t]$ is countable. If I is uncountable, $I = [0, t]$ say, but we also know that X has right-continuous paths, then $U([a, b]; I)$ is still $\mathcal{F}_t$-measurable as the paths of X are determined by their values at the countably many rational time points. Fundamental properties of (sub)martingales are collected in the following proposition.

Proposition 2.2 *Let X be a submartingale with right-continuous paths. Then*

(i) For all $\lambda > 0$ and $0 \leq s \leq t$ one has Doob's supremal inequality

$$\lambda \mathbb{P}(\sup_{s \leq u \leq t} X_u \geq \lambda) \leq \mathbb{E}\, X_t^+ \tag{2.1}$$

and

$$\lambda \mathbb{P}(\inf_{s \leq u \leq t} X_u \leq -\lambda) \leq \mathbb{E}\, X_t^+ - \mathbb{E}\, X_s. \tag{2.2}$$

(ii) If X is a nonnegative submartingale and $p > 1$, then one has Doob's L^p inequality

$$|| \sup_{s \leq u \leq t} X_u ||_p \leq \frac{p}{p-1} ||X_t||_p, \tag{2.3}$$

where for random variables ξ we put $||\xi||_p = (\mathbb{E}\,|\xi|^p)^{1/p}$. In particular, if M is a right-continuous martingale with $\mathbb{E}\,M_t^2 < \infty$ for all $t > 0$, then

$$\mathbb{E}\,(\sup_{s \le u \le t} M_u^2) \le 4\,\mathbb{E}\,M_t^2. \tag{2.4}$$

(iii) For the number of upcrossings $U([a,b];[s,t])$ it holds that

$$\mathbb{E}\,U([a,b];[s,t]) \le \frac{\mathbb{E}\,X_t^+ + |a|}{b-a}.$$

(iv) Almost every path of X is bounded on compact intervals and has no discontinuities of the second kind (left limits exist a.s. at all $t > 0$). For almost each path the set of points at which this path is discontinuous is at most countable.

Proof The proofs of these results are essentially the same as in the discrete time case. The basic argument to justify this claim is to consider X restricted to a finite set F in the interval $[s,t]$. With X restricted to such a set, the above inequalities in (i) and (ii) are valid as we know it from discrete time theory. Take then a sequence of such F, whose union is a dense subset of $[s,t]$, then the inequalities keep on being valid. By right-continuity of X we can extend the validity of the inequalities to the whole interval $[s,t]$. The same reasoning applies to (iii).

For (iv) we argue as follows. Combining the inequalities in (i), we see that $\mathbb{P}(\sup_{s \le u \le t} |X_u| = \infty) = 0$, hence almost all sample paths are bounded on intervals $[s,t]$. We have to show that almost all paths of X admit left limits everywhere. This follows as soon as we can show that for all n the set $\{\liminf_{s \uparrow t} X_s < \limsup_{s \uparrow t} X_s, \text{for some } t \in [0,n]\}$ has zero probability. This set is contained in $\cup_{a,b \in \mathbb{Q}} \{U([a,b],[0,n]) = \infty\}$. But by (iii) this set has zero probability. It is a *fact* from analysis that the set of discontinuities of the first kind of a function is at most countable, which yields the last assertion. □

The last assertion of this proposition describes a regularity property of the sample paths of a right-continuous submartingale. The next theorem gives a sufficient and necessary condition that justifies the fact that we mostly restrict our attention to càdlàg submartingales. This condition is trivially satisfied for martingales. We state the theorem without proof.

Theorem 2.3 *Let X be a submartingale and let the filtration $\mathbb{F}$ satisfy the usual conditions. Suppose that the function $t \mapsto \mathbb{E}\,X_t$ is right-continuous on $[0,\infty)$. Then there exists a modification Y of X that has càdlàg paths and that is also a submartingale w.r.t. $\mathbb{F}$.*

2.2 Limit theorems and optional sampling

The following two theorems are the fundamental convergence theorems.

Theorem 2.4 *Let X be a right-continuous submartingale with $\sup_{t\geq 0}\mathbb{E}\,X_t^+ < \infty$. Then there exists a $\mathcal{F}_\infty$-measurable random variable X_∞ with $\mathbb{E}\,|X_\infty| < \infty$ such that $X_t \stackrel{\text{a.s.}}{\to} X_\infty$. If moreover X is uniformly integrable, then we also have $X_t \stackrel{L^1}{\to} X_\infty$ and $\mathbb{E}\,[X_\infty|\mathcal{F}_t] \geq X_t$ a.s. for all $t \geq 0$.*

Theorem 2.5 *Let X be a right-continuous martingale. Then X is uniformly integrable iff there exists an integrable random variable Z such that $\mathbb{E}\,[Z|\mathcal{F}_t] = X_t$ a.s. for all $t \geq 0$. In this case we have $X_t \stackrel{\text{a.s.}}{\to} X_\infty$ and $X_t \stackrel{L^1}{\to} X_\infty$, where $X_\infty = \mathbb{E}\,[Z|\mathcal{F}_\infty]$.*

Proof The proofs of these theorems are like in the discrete time case. □

We will frequently use the optional sampling theorem (Theorem 2.7) for submartingales and martingales. In the proof of this theorem we use the following lemma.

Lemma 2.6 *Let $(\mathcal{G}_n)$ be a decreasing sequence of σ-algebras and let (Y_n) be a sequence of integrable random variables such that Y_n is a $\mathcal{G}_n$-measurable random variable for all n and such that*

$$\mathbb{E}\,[Y_m|\mathcal{G}_n] \geq Y_n, \forall n \geq m.$$

If the sequence of expectations $\mathbb{E}\,Y_n$ is bounded from below, then the collection $\{Y_n, n \geq 1\}$ is uniformly integrable.

Proof Consider the chain of (in)equalities (where $n \geq m$)

$$\begin{aligned}
\mathbb{E}\,\mathbf{1}_{|Y_n|>\lambda}|Y_n| &= \mathbb{E}\,\mathbf{1}_{Y_n>\lambda}Y_n - \mathbb{E}\,\mathbf{1}_{Y_n<-\lambda}Y_n \\
&= -\mathbb{E}\,Y_n + \mathbb{E}\,\mathbf{1}_{Y_n>\lambda}Y_n + \mathbb{E}\,\mathbf{1}_{Y_n\geq-\lambda}Y_n \\
&\leq -\mathbb{E}\,Y_n + \mathbb{E}\,\mathbf{1}_{Y_n>\lambda}Y_m + \mathbb{E}\,\mathbf{1}_{Y_n\geq-\lambda}Y_m \\
&\leq -\mathbb{E}\,Y_n + \mathbb{E}\,Y_m + \mathbb{E}\,\mathbf{1}_{|Y_n|>\lambda}|Y_m|.
\end{aligned} \tag{2.5}$$

Since the sequence of expectations $(\mathbb{E}\,Y_n)$ has a limit and hence is Cauchy, we choose for given $\varepsilon > 0$ the integer m such that for all $n > m$ we have $-\mathbb{E}\,Y_n + \mathbb{E}\,Y_m < \varepsilon$. By the conditional version of Jensen's inequality we have $\mathbb{E}\,Y_n^+ \leq \mathbb{E}\,Y_1^+$. Since $\mathbb{E}\,Y_n \geq l$ for some finite l, we can conclude that $\mathbb{E}\,|Y_n| = 2\mathbb{E}\,Y_n^+ - \mathbb{E}\,Y_n \leq 2\mathbb{E}\,Y_1^+ - l$. This implies that $\mathbb{P}(|Y_n| > \lambda) \leq \frac{2\mathbb{E}\,Y_1^+ - l}{\lambda}$. Hence we can make the expression in (2.5) arbitrarily small for all n big enough and we can do this uniformly in n (fill in the details). □

Here is the *optional sampling theorem* for right-continuous submartingales.

Theorem 2.7 *Let X be a right-continuous submartingale with a last element X_∞ (i.e. $\mathbb{E}\,[X_\infty|\mathcal{F}_t] \geq X_t$, a.s. for every $t \geq 0$) and let S and T be two stopping times such that $S \leq T$. Then $X_S \leq \mathbb{E}\,[X_T|\mathcal{F}_S]$ a.s.*

Proof Put $T_n = 2^{-n}[2^n T + 1]$ and $S_n = 2^{-n}[2^n S + 1]$. Here $[x]$ denotes the largest integer less than or equal to x. Then the T_n and S_n are stopping times relative to the filtration $\{\mathcal{F}_{k2^{-n}}, k \geq 0\}$ (Exercise 2.5), $T_n \geq S_n$, and they form two non-increasing sequences with limits T, respectively S. Notice that all T_n and S_n are at most countably valued. We can apply the optional sampling theorem for discrete time submartingales (see Theorem B.4) to get $\mathbb{E}\,[X_{T_n}|\mathcal{F}_{S_n}] \geq X_{S_n}$, from which we obtain that for each $A \in \mathcal{F}_S \subset \mathcal{F}_{S_n}$ it holds that

$$\mathbb{E}\,\mathbf{1}_A X_{T_n} \geq \mathbb{E}\,\mathbf{1}_A X_{S_n}. \tag{2.6}$$

It similarly follows, that $\mathbb{E}\,[X_{T_n}|\mathcal{F}_{T_{n+1}}] \geq X_{T_{n+1}}$. Notice that the expectations $\mathbb{E}\,X_{T_n}$ form a decreasing sequence with lower bound $\mathbb{E}\,X_0$. We can thus apply Lemma 2.6 to conclude that the collection $\{X_{T_n} : n \geq 1\}$ is uniformly integrable. Of course the same holds for $\{X_{S_n} : n \geq 1\}$. By right-continuity of X we get $\lim_{n\to\infty} X_{T_n} = X_T$ a.s. and $\lim_{n\to\infty} X_{S_n} = X_S$ a.s. Uniform integrability implies that we also have L^1-convergence, hence we have from Equation (2.6) that $\mathbb{E}\,\mathbf{1}_A X_T \geq \mathbb{E}\,\mathbf{1}_A X_S$ for all $A \in \mathcal{F}_S$, which is what we had to prove. $\square$

Remark 2.8 The condition in Theorem 2.7 that X has a last element can be replaced by restriction to *bounded* stopping times $S \leq T$ and one also obtains $\mathbb{E}\,X_S \leq \mathbb{E}\,X_T$. In fact, if X is a right-continuous adapted process for which all $\mathbb{E}\,|X_t|$ are finite, and if $\mathbb{E}\,X_S \leq \mathbb{E}\,X_T$ for all bounded stopping times with $S \leq T$, then X is a submartingale. This can be seen by taking $t > s$, $S = s$ and $T = t\mathbf{1}_F + s\mathbf{1}_{F^c}$ for arbitrary $F \in \mathcal{F}_s$.

Corollary 2.9 *Let X be a right-continuous submartingale and let $S \leq T$ be stopping times. Then the stopped process X^T is a submartingale as well and $\mathbb{E}\,[X_{T\wedge t}|\mathcal{F}_S] \geq X_{S\wedge t}$ a.s. for every $t \geq 0$.*

Proof This is Exercise 2.3. $\square$

Here is another application of the optional sampling theorem.

Proposition 2.10 *Let X be a right-continuous nonnegative supermartingale. Let $T = \inf\{t > 0 : X_t = 0 \text{ or } X_{t-} = 0\}$. Then $X\mathbf{1}_{[T,\infty)}$ is indistinguishable from zero.*

Proof Take $X_\infty = 0$ as a last element of X. Let $T_n = \inf\{t : X_t \leq 1/n\}$ for $n \geq 1$. Then $T_n \leq T_{n+1} \leq T$ and $X_{T_n} \leq 1/n$ on $\{T_n < \infty\}$. From Theorem 2.7 we obtain for each stopping time $S \geq T_n$

$$\frac{1}{n} \geq \mathbb{E}\,X_{T_n} \geq \mathbb{E}\,X_S \geq 0.$$

Put $S = T + q$, where q is a positive rational number. It follows that $X_{T+q} = 0$ a.s., which yields the assertion by right-continuity. $\square$

2.3 Doob-Meyer decomposition

It is instructive to formulate the discrete time analogues of what will be described below. This will be left as exercises.

Definition 2.11

(i) An adapted process A is called *increasing* if a.s. we have $A_0 = 0$, $t \to A_t(\omega)$ is a nondecreasing right-continuous function and if $\mathbb{E}\, A_t < \infty$ for all t. An increasing process is called integrable if $\mathbb{E}\, A_\infty < \infty$, where $A_\infty = \lim_{t\to\infty} A_t$.

(ii) An increasing process A is called *natural* if for every right-continuous and bounded martingale ξ we have

$$\mathbb{E} \int_{(0,t]} \xi_s \, \mathrm{d}A_s = \mathbb{E} \int_{(0,t]} \xi_{s-} \, \mathrm{d}A_s, \forall t \geq 0. \tag{2.7}$$

Remark 2.12 The integrals $\int_{(0,t]} \xi_s \, \mathrm{d}A_s$ and $\int_{(0,t]} \xi_{s-} \, \mathrm{d}A_s$ in Equation (2.7) are defined pathwise as Lebesgue-Stieltjes integrals (see Section C) for all $t \geq 0$ and hence the corresponding processes are progressive (you check!). Furthermore, if the increasing process A is continuous, then it is natural. This follows from the fact the paths of ξ have only countably many discontinuities (Proposition 2.2(iv)).

Definition 2.13 A right-continuous process is said to belong to class D if the collection of X_T, where T runs through the set of all finite stopping times, is uniformly integrable. A right-continuous process X is said to belong to class DL if for all $a > 0$ the collection X_T, where T runs through the set of all stopping times bounded by a, is uniformly integrable.

With these definitions we can state and prove the celebrated Doob-Meyer decomposition (Theorem 2.16 below) of submartingales. In the proof of this theorem we use the following two lemmas. The first one (the Dunford-Pettis criterion) is a rather deep result in functional analysis, of which we give a partial proof in the appendix. The second one will also be used elsewhere.

Lemma 2.14 *A collection $\mathcal{Z}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a uniformly integrable family, iff for any sequence $\{Z_n, n \geq 1\} \subset \mathcal{Z}$ there exist an increasing subsequence (n_k) and a random variable Z such that for all bounded random variables ζ one has $\lim_{k\to\infty} \mathbb{E}\, Z_{n_k}\zeta = \mathbb{E}\, Z\zeta$. Moreover, if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then also $\lim_{k\to\infty} \mathbb{E}\,[\mathbb{E}\,[Z_{n_k}|\mathcal{G}]\zeta] = \mathbb{E}\,[\mathbb{E}\,[Z|\mathcal{G}]\zeta]$.*

Proof See Appendix D. □

We extend the concept of a natural increasing process as follows. First, we recall that a function is of bounded variation (over finite intervals), if it can be decomposed into a difference of two increasing functions (see Appendix C). A process A is called *natural and of bounded variation a.s.* if it can be written as the difference of two natural increasing processes.

Lemma 2.15 *Assume that the filtration satisfies the usual conditions. Let M be a right-continuous martingale which is natural and of bounded variation a.s. Then M is indistinguishable from a process that is constant over time.*

Proof Without loss of generality we suppose that $M_0 = 0$ a.s. Write $M = A - A'$, where the processes A and A' are natural and increasing. Then we have for any bounded right-continuous martingale ξ and for all $t \geq 0$ the equality (this is Exercise 2.6)

$$\mathbb{E}\, \xi_t A_t = \mathbb{E} \int_{(0,t]} \xi_{s-} \, \mathrm{d}A_s.$$

If we replace A with A' the above equality is then of course also true and then continues to hold if we replace A with M. So we have

$$\mathbb{E}\, \xi_t M_t = \mathbb{E} \int_{(0,t]} \xi_{s-} \, \mathrm{d}M_s. \tag{2.8}$$

The expectation on the right hand side of (2.8) is a limit of the expectations (use the Dominated Convergence Theorem separately to integrals w.r.t. A and A') $\mathbb{E} \sum_k \xi_{t_{k-1}}(M_{t_k} - M_{t_{k-1}})$, where the t_k come from a sequence of nested partitions whose mesh tend to zero. But $\mathbb{E} \sum_k \xi_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) = 0$, since M is a martingale. We conclude that $\mathbb{E}\, \xi_t M_t = 0$.

Let ζ be a bounded random variable and put $\xi_t \equiv \mathbb{E}\,[\zeta|\mathcal{F}_t]$, then ξ is a martingale. We know from Theorem 2.3 that ξ admits a right continuous modification and hence we get $\mathbb{E}\, M_t \zeta = 0$. Choosing ζ in an appropriate way (how?) we conclude that $M_t = 0$ a.s. We can do this for any $t \geq 0$ and then right-continuity yields that M and the zero process are indistinguishable. □

Theorem 2.16 *Let the filtration $\mathbb{F}$ satisfy the usual conditions and let X be a right-continuous submartingale of class* DL. *Then X can be decomposed according to*

$$X = A + M, \tag{2.9}$$

where M is a right-continuous martingale and A a natural increasing process. The decomposition (2.9) *is unique up to indistinguishability and is called Doob-Meyer decomposition. Under the stronger condition that X is of class* D, *M is uniformly integrable and A is integrable.*

Proof We first show the uniqueness assertion. Suppose that we have next to (2.9) another decomposition $X = A' + M'$ of the same type. Then $B := A - A' = M' - M$ is a right-continuous martingale that satisfies the assumptions of Lemma 2.15 with $B_0 = 0$. Hence it is indistinguishable from the zero process.

We now establish the existence of the decomposition for $t \in [0, a]$ for an arbitrary $a > 0$. Let Y be the process defined by $Y_t := X_t - \mathbb{E}\,[X_a|\mathcal{F}_t]$. Since X is a submartingale $Y_t \leq 0$ for $t \in [0, a]$. Moreover the process Y is a submartingale itself that has a right-continuous modification (Exercise 2.15), again denoted by Y. From now on we work with this modification.

Consider the nested sequence of dyadic partitions Π^n of $[0, a]$, $\Pi^n \ni t_j^n = j2^{-n}a$, $j = 0, \ldots, 2^n$. With time restricted to Π^n, the process Y becomes a discrete time submartingale to which we apply the Doob decomposition (Exercise 2.9) with an increasing process A^n, which results in $Y = A^n + M^n$, with $A_0^n = 0$. Since $Y_a = 0$, we get

$$A_{t_j^n}^n = Y_{t_j^n} + \mathbb{E}\,[A_a^n|\mathcal{F}_{t_j^n}]. \tag{2.10}$$

We now make the following claim.

$$\text{The family } \{A_a^n : n \geq 1\} \text{ is uniformly integrable.} \tag{2.11}$$

Supposing that the claim holds true, we finish the proof as follows. In view of Lemma 2.14 we can select from this family a subsequence, *again denoted by* A_a^n, and we can find a random variable A_a such that $\mathbb{E}\,\zeta A_a^n \to \mathbb{E}\,A_a\zeta$ for every bounded random variable ζ. We now define for $t \in [0, a]$ the random variable A_t as any version of

$$A_t = Y_t + \mathbb{E}\,[A_a|\mathcal{F}_t]. \tag{2.12}$$

The process A has a right-continuous modification on $[0, a]$, again denoted by A. We now show that A is increasing. Let $s, t \in \bigcup_n \Pi^n$, $s < t$. Then there is n_0 such that for all $n \geq n_0$ we have $s, t \in \Pi^n$. Using the second assertion of Lemma 2.14, equations (2.10) and (2.12), we obtain $\mathbb{E}\,\zeta(A_t^n - A_s^n) \to \mathbb{E}\,\zeta(A_t - A_s)$. Since for each n the A^n is increasing we get that $\mathbb{E}\,\zeta(A_t - A_s) \geq 0$ as soon as $\zeta \geq 0$. Take now $\zeta = \mathbf{1}_{\{A_s > A_t\}}$ to conclude that $A_t \geq A_s$ a.s. Since A is right-continuous, we have that A is increasing on the whole interval $[0, a]$. It follows from the construction that $A_0 = 0$ a.s. (Exercise 2.11).

Next we show that A is natural. Let ξ be a bounded and right-continuous martingale. The discrete time process A^n is predictable (and thus natural, Exercise 2.10). Restricting time to Π^n and using the fact that Y and A as well as Y and A^n differ by a martingale (equations (2.10) and (2.12)), we get

$$\begin{aligned}
\mathbb{E}\,\xi_a A_a^n &= \mathbb{E} \sum_j \xi_{t_{j-1}^n}(A_{t_j^n}^n - A_{t_{j-1}^n}^n) \\
&= \mathbb{E} \sum_j \xi_{t_{j-1}^n}(Y_{t_j^n} - Y_{t_{j-1}^n}) \\
&= \mathbb{E} \sum_j \xi_{t_{j-1}^n}(A_{t_j^n} - A_{t_{j-1}^n})
\end{aligned}$$

Let $n \to \infty$ to conclude that

$$\mathbb{E}\,\xi_a A_a = \mathbb{E}\int_{(0,a]} \xi_{s-}\,\mathrm{d}A_s, \tag{2.13}$$

by the definition of A_a and the dominated convergence theorem applied to the right hand side of the above string of equalities. Next we replace in (2.13) ξ with the, at the deterministic time $t \le a$, stopped process ξ^t. A computation shows that we can conclude

$$\mathbb{E}\,\xi_t A_t = \mathbb{E}\int_{(0,t]} \xi_{s-}\,\mathrm{d}A_s, \forall t \le a. \tag{2.14}$$

In view of Exercise 2.6 this shows that A is natural on $[0,a]$. The proof of the first assertion of the theorem is finished by setting $M_t = \mathbb{E}\,[X_a - A_a|\mathcal{F}_t]$. Clearly M is a martingale and $M_t = X_t - A_t$. Having shown that the decomposition is valid on each interval $[0,a]$, we invoke uniqueness to extend it to $[0,\infty)$.

If X belongs to class D, it has a last element (see Theorem 2.4) and we can repeat the above proof with a replaced by ∞. What is left however, is the proof of the claim (2.11). This will be given now.

Let $\lambda > 0$ be fixed. Define $T^n = \inf\{t^n_{j-1} : A^n_{t^n_j} > \lambda, j = 1,\dots,2^n\} \wedge a$. One checks that the T^n are stopping times, bounded by a, that $\{T^n < a\} = \{A^n_a > \lambda\}$ and that we have $A^n_{T^n} \le \lambda$ on this set. By the optional sampling theorem (in discrete time) applied to (2.10) one has $Y_{T^n} = A^n_{T^n} - \mathbb{E}\,[A^n_a|\mathcal{F}_{T^n}]$. Below we will often use that $Y \le 0$. To show uniform integrability, one considers $\mathbb{E}\,\mathbf{1}_{\{A^n_a > \lambda\}} A^n_a$ and with the just mentioned facts one gets

$$\begin{aligned}\mathbb{E}\,\mathbf{1}_{\{A^n_a>\lambda\}}A^n_a &= \mathbb{E}\,\mathbf{1}_{\{A^n_a>\lambda\}}A^n_{T^n} - \mathbb{E}\,\mathbf{1}_{\{A^n_a>\lambda\}}Y_{T^n} \\ &\le \lambda\mathbb{P}(T^n < a) - \mathbb{E}\,\mathbf{1}_{\{T^n<a\}}Y_{T^n}.\end{aligned} \tag{2.15}$$

Let $S^n = \inf\{t^n_{j-1} : A^n_{t^n_j} > \frac{1}{2}\lambda, j = 1,\dots,2^n\} \wedge a$. Then the S^n are bounded stopping times as well with $S^n \le T^n$ and we have, like above, $\{S^n < a\} = \{A^n_a > \frac{1}{2}\lambda\}$ and $A^n_{S^n} \le \frac{1}{2}\lambda$ on this set. Using $Y_{S^n} = A^n_{S^n} - \mathbb{E}\,[A^n_a|\mathcal{F}_{S^n}]$ we develop

$$\begin{aligned}-\mathbb{E}\,\mathbf{1}_{\{S^n<a\}}Y_{S^n} &= -\mathbb{E}\,\mathbf{1}_{\{S^n<a\}}A^n_{S^n} + \mathbb{E}\,\mathbf{1}_{\{S^n<a\}}A^n_a \\ &= \mathbb{E}\,\mathbf{1}_{\{S^n<a\}}(A^n_a - A^n_{S^n}) \\ &\ge \mathbb{E}\,\mathbf{1}_{\{T^n<a\}}(A^n_a - A^n_{S^n}) \\ &\ge \frac{1}{2}\lambda\mathbb{P}(T^n < a).\end{aligned}$$

It follows that $\lambda\mathbb{P}(T^n < a) \le -2\mathbb{E}\,\mathbf{1}_{\{S^n<a\}}Y_{S^n}$. Inserting this estimate into inequality (2.15), we obtain

$$\mathbb{E}\,\mathbf{1}_{\{A^n_a>\lambda\}}A^n_a \le -2\mathbb{E}\,\mathbf{1}_{\{S^n<a\}}Y_{S^n} - \mathbb{E}\,\mathbf{1}_{\{T^n<a\}}Y_{T^n}. \tag{2.16}$$

The assumption that X belongs to class DL implies that both $\{Y_{T^n}, n \geq 1\}$ and $\{Y_{S^n}, n \geq 1\}$ are uniformly integrable, since Y and X differ for $t \in [0,a]$ by a uniformly integrable martingale. Then we can find for all $\varepsilon > 0$ a $\delta > 0$ such that $-\mathbb{E}\,\mathbf{1}_F Y_{T^n}$ and $-\mathbb{E}\,\mathbf{1}_F Y_{S^n}$ are smaller than ε for any set F with $\mathbb{P}(F) < \delta$, uniformly in n. Since $Y_a = 0$, we have $\mathbb{E}\,A^n_a = -\mathbb{E}\,M^n_a = -\mathbb{E}\,M^n_0 = -\mathbb{E}\,Y_0$. Hence we get $\mathbb{P}(T_n < a) \leq \mathbb{P}(S^n < a) = \mathbb{P}(A^n_a > \lambda/2) \leq -2\mathbb{E}\,Y_0/\lambda$. Hence for λ bigger than $-2\mathbb{E}\,Y_0/\delta$, we can make both probabilities $\mathbb{P}(T_n < a)$ and $\mathbb{P}(S^n < a)$ of the sets in (2.16) less than δ, uniformly in n. This shows that the family $\{A^n_a : n \geq 1\}$ is uniformly integrable. □

The general statement of Theorem 2.16 only says that the processes A and M can be chosen to be right-continuous. If the process X is continuous, one would expect A and M to be continuous. This is true indeed, at least for certain nonnegative processes, see Exercise 2.13. However a stronger result can be proven.

Theorem 2.17 *Suppose that in addition to the conditions of Theorem* 2.16 *the submartingale is* weakly left-continuous, *i.e. it is such that* $\lim_{n\to\infty} \mathbb{E}\,X_{T^n} = \mathbb{E}\,X_T$ *for all bounded increasing sequences of stopping times* T^n *with limit* T. *Then the process* A *in equation* (2.9) *is continuous.*

Proof Let $a > 0$ be an upper bound for T and the T^n. According to Theorem 2.16 we have $X = A + M$. Hence, according to the optional sampling theorem we have $\mathbb{E}\,A_{T^n} = \mathbb{E}\,X_{T^n} - \mathbb{E}\,M_{T^n} = \mathbb{E}\,X_{T^n} - \mathbb{E}\,M_a$, which implies that

$$\mathbb{E}\,A_{T^n} \uparrow \mathbb{E}\,A_T. \tag{2.17}$$

Since A is increasing, we know that (A_{T^n}) has as a finite limit and by the monotone convergence theorem we can even conclude (check!) that $A_{T^n(\omega)}(\omega) \uparrow A_{T(\omega)}(\omega)$ for all ω outside a set of probability zero. This set however, in principle depends on the chosen sequence of stopping times, which is insufficient to establish a.s. continuity of A.

We proceed as follows. Assume for the time being that A is a bounded process. As in the proof of Theorem 2.16 we consider the dyadic partitions Π^n of the interval $[0,a]$. For every n and every $j = 0, \ldots, 2^n - 1$ we define $\xi^{n,j}$ to be the right-continuous modification of the martingale defined by $\mathbb{E}\,[A_{t^n_{j+1}}|\mathcal{F}_t]$. Then we define

$$\xi^n_t = \sum_{j=0}^{2^n-1} \xi^{n,j}_t \mathbf{1}_{(t^n_j, t^n_{j+1}]}(t).$$

The process ξ^n is on the whole interval $[0,a]$ right-continuous, except possibly at the points of the partition. Notice that we have $\xi^n_t \geq A_t$ a.s. for all $t \in [0,a]$ with equality at the points of the partition and $\xi^{n+1} \leq \xi^n$. Since A is a natural increasing process, we have for every n and j that

$$\mathbb{E}\int_{(t^n_j, t^n_{j+1}]} \xi^n_s \,\mathrm{d}A_s = \mathbb{E}\int_{(t^n_j, t^n_{j+1}]} \xi^n_{s-} \,\mathrm{d}A_s.$$

Summing this equality over all relevant j we get

$$\mathbb{E}\int_{(0,t]} \xi^n_s \,\mathrm{d}A_s = \mathbb{E}\int_{(0,t]} \xi^n_{s-}\,\mathrm{d}A_s, \text{ for all } t\in[0,a]. \tag{2.18}$$

We introduce the right-continuous nonnegative processes η^n defined by $\eta^n_t = (\xi^n_{t+} - A_t)\mathbf{1}_{[0,a)}(t)$. For fixed $\varepsilon > 0$ we define the bounded stopping times (see Proposition 1.10) $T^n = \inf\{t\in[0,a] : \eta^n_t > \varepsilon\}\wedge a$. Observe that also $T^n = \inf\{t\in[0,a] : \xi^n_t - A_t > \varepsilon\}\wedge a$ in view of the relation between η^n and ξ^n (Exercise 2.12). Since the ξ^n_t are decreasing in n, we have that the sequence T^n is increasing and thus has a limit, T say. Let $\phi^n(t) = \sum_{j=0}^{2^n-1} t^n_{j+1}\mathbf{1}_{(t^n_j,t^n_{j+1}]}(t)$ and notice that for all n and $t\in[0,a]$ one has $a \geq \phi^n(t) \geq \phi^{n+1}(t) \geq t$. It follows that $\lim_{n\to\infty}\phi^n(T^n) = T$ a.s. From the optional sampling theorem applied to the martingales $\xi^{n,j}$ it follows that

$$\begin{aligned}
\mathbb{E}\,\xi^n_{T^n} &= \mathbb{E}\sum_j \mathbb{E}\,[A_{t^n_{j+1}}|\mathcal{F}_{T^n}]\mathbf{1}_{(t^n_j,t^n_{j+1}]}(T^n)\\
&= \mathbb{E}\sum_j \mathbb{E}\,[A_{t^n_{j+1}}\mathbf{1}_{(t^n_j,t^n_{j+1}]}(T^n)|\mathcal{F}_{T^n}]\\
&= \mathbb{E}\sum_j \mathbb{E}\,[A_{\phi^n(T^n)}\mathbf{1}_{(t^n_j,t^n_{j+1}]}(T^n)|\mathcal{F}_{T^n}]\\
&= \mathbb{E}\,A_{\phi^n(T^n)}.
\end{aligned}$$

And then

$$\begin{aligned}
\mathbb{E}\,(A_{\phi^n(T^n)} - A_{T^n}) &= \mathbb{E}\,(\xi^n_{T^n} - A_{T^n})\\
&\geq \mathbb{E}\,\mathbf{1}_{\{T^n<a\}}(\xi^n_{T^n} - A_{T^n})\\
&\geq \varepsilon\mathbb{P}(T^n < a).
\end{aligned}$$

Thus

$$\mathbb{P}(T^n < a) \leq \frac{1}{\varepsilon}\mathbb{E}\,(A_{\phi^n(T^n)} - A_{T^n}) \to 0 \text{ for } n\to\infty, \tag{2.19}$$

by (2.17) and right-continuity and boundedness of A. But since $\{T^n < a\} = \{\sup_{t\in[0,a]}|\xi^n_t - A_t| > \varepsilon\}$ and since (2.19) holds for every $\varepsilon > 0$, we can find a subsequence (n_k) along which $\sup_{t\in[0,a]}|\xi^n_t - A_t| \overset{\text{a.s.}}{\to} 0$. Since both A and ξ are bounded, it follows from the dominated convergence theorem applied to Equation (2.18) and the chosen subsequence that

$$\mathbb{E}\int_{(0,t]} A_s\,\mathrm{d}A_s = \mathbb{E}\int_{(0,t]} A_{s-}\,\mathrm{d}A_s, \text{ for all } t\in[0,a],$$

and hence

$$\mathbb{E}\sum_{t\leq a}(\Delta A_t)^2 = \mathbb{E}\int_{(0,t]}(A_s - A_{s-})\,\mathrm{d}A_s = 0 \text{ for all } t\in[0,a]. \tag{2.20}$$

We conclude that $\Delta A_t = 0$ a.s. for all $t \geq 0$ and hence, by right-continuity that $A - A_-$ is indistinguishable from zero. Thus the paths of A are a.s. continuous.

The result has been proved under the assumption that A is bounded. If this assumption doesn't hold, we can introduce the stopping times $S_m = \inf\{t \geq 0 : A_t > m\}$ and replace everywhere above A with the stopped process A^{S_m} which is bounded. The conclusion from the analogue of (2.20) will be that the processes $(A - A_-)\mathbf{1}_{[0,S_m]}$ are indistinguishable from zero for all m and the final result follows by letting $m \to \infty$. □

We close this section with the following proposition.

Proposition 2.18 *Let X satisfy the conditions of Theorem 2.16 with Doob-Meyer decomposition $X = A + M$ and T a stopping time. Then also the stopped process X^T satisfies these conditions and its Doob-Meyer decomposition is given by*

$$X^T = A^T + M^T.$$

Proof That X^T also satisfies the conditions of Theorem 2.16 is straightforward. Of course, by stopping, we have $X^T = A^T + M^T$, so we only have to show that this indeed gives us the Doob-Meyer decomposition. By the optional sampling theorem and its Corollary 2.9, the stopped process M^T is a martingale. That the process A^T is natural, follows from the identity

$$\int_{(0,t]}\xi_s\,\mathrm{d}A^T_s = \int_{(0,t]}\xi^T_s\,\mathrm{d}A_s - \xi_T(A_t - A_{t\wedge T}),$$

a similar one for $\int_{(0,t]}\xi_{s-}\,\mathrm{d}A^T_s$, valid for any bounded measurable process ξ and the fact that A is natural. The uniqueness of the Doob-Meyer decomposition then yields the result. □

2.4 Exercises

2.1 Let X be a supermartingale with constant expectation. Show that X is in fact a martingale.

2.2 Why is $\sup_{s\leq u\leq t} X_u$ in formula (2.3) measurable?

2.3 Prove Corollary 2.9. To prove that X^T is a submartingale, you may want to use the commutation property of Exercise 1.8.

2.4 Suppose that $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ is a filtration and X is a right-continuous process and T a stopping time. Suppose that X is martingale w.r.t. the filtration $\{\mathcal{F}_{T\wedge t}, t \geq 0\}$. Show that X is also a martingale w.r.t. $\mathbb{F}$. (Exercise 1.8 may be helpful.)

2.5 Show that the random variables T_n in the proof of Theorem 2.7 are stopping times relative to the filtration $\{\mathcal{F}_{k2^{-n}}, k \geq 0\}$ and that they form a non-increasing sequence with pointwise limit T.

2.6 Show that an increasing process A is natural iff for all bounded right-continuous martingales ξ one has $\mathbb{E} \int_{(0,t]} \xi_{s-} \, \mathrm{d}A_s = \mathbb{E}\, \xi_t A_t$ for all $t \geq 0$. *Hint:* Consider first a partition $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$ and take ξ constant on the $(t_{k-1}, t_k]$. Then you approximate ξ with martingales of the above type.

2.7 Show that a (right-continuous) uniformly integrable martingale is of class D.

2.8 Let X be a right-continuous submartingale. Show that X is of class DL if X is nonnegative or if $X = M + A$, where A is an increasing process and M a martingale.

2.9 Let X be a discrete time process on some $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{F} = \{\mathcal{F}_n\}_{n \geq 0}$ be a filtration. Assume that X is adapted, $X_0 = 0$ and that X_n is integrable for all n. Define the process M by $M_0 = 0$ and $M_n = M_{n-1} + X_n - \mathbb{E}\,[X_n|\mathcal{F}_{n-1}$ for $n \geq 1$. Show that M is a martingale. Define then $A = X - M$. Show that A_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$ (one says that A is predictable) and that $A_0 = 0$. Show that A can be taken as an increasing process iff X is a submartingale.

2.10 A discrete time process A is called *increasing* if it is adapted, $A_0 = 0$ a.s., $A_n - A_{n-1} \geq 0$ a.s. and $\mathbb{E}\, A_n < \infty$ for all $n \geq 1$. An increasing process is natural if for all bounded martingales ξ one has $\mathbb{E}\, A_n \xi_n = \mathbb{E} \sum_{k=1}^n \xi_{k-1}(A_k - A_{k-1})$.

(a) Show that a process A is natural iff for all bounded martingales one has $\mathbb{E} \sum_{k=1}^n A_k(\xi_k - \xi_{k-1}) = 0$.

(b) Show that a predictable increasing process is natural.

(c) Show that a natural process is a.s. predictable. *Hint:* you have to show that $A_n = \tilde{A}_n$ a.s., where $\tilde{A}_n$ is a version of $\mathbb{E}\,[A_n|\mathcal{F}_{n-1}]$ for each n, which you do along the following steps. First you show that for all n one has $\mathbb{E}\,\xi_n A_n = \mathbb{E}\,\xi_{n-1} A_n = \mathbb{E}\,\xi_n \tilde{A}_n$. Fix n, take $\xi_k = \mathbb{E}\,[\mathrm{sgn}(A_n - \tilde{A}_n)|\mathcal{F}_k]$ and finish the proof.

2.11 Show that the A_0 in the proof of the Doob-Meyer decomposition (Theorem 2.16) is zero a.s.

2.12 Show that (in the proof of Theorem 2.17) $T^n = \inf\{t \in [0, a] : \xi^n_t - A_t > \varepsilon\} \wedge a$. *Hint:* use that ξ^n is right-continuous except possibly at the t^n_j.

2.13 A continuous nonnegative submartingale satisfies the conditions of Theorem 2.17. Show this. (Hence the predictable process in the Doob-Meyer decomposition of X can be taken to be continuous.)

2.14 Show (in the proof of Theorem 2.17) the convergence $\phi^n(T^n) \to T$ a.s. and the convergence in (2.19).

2.15 Suppose that X is a submartingale with right-continuous paths. Show that $t \mapsto \mathbb{E}\, X_t$ is right-continuous (use Lemma 2.6).

2.16 Let $N = \{N_t : t \geq 0\}$ be a Poisson process with parameter λ and let $\mathcal{F}_t = \sigma(N_s,\, s \leq t)$, $t \geq 0$. Show that N is a submartingale and that the process A in Theorem 2.16 is given by $A_t = \lambda t$. Note that A is continuous, which is a consequence of Theorem 2.17. (Optional exercise: show that N is weakly left-continuous.) Give also a decomposition of N as $N = B + M$, where M is martingale and B is increasing, but not natural. Let $X_t = (N_t - \lambda t)^2$. Show that X has the same natural increasing process A.

2.17 Here is a converse to the optional sampling theorem for martingales. Let X be a càdlàg process such that for all bounded stopping times T it holds that $\mathbb{E}\,|X_T|, \infty$ and $\mathbb{E}\, X_T = \mathbb{E}\, X_0$. Show that X is a martingale. *Hint: take* $s < t$, $A \in \mathcal{F}_s$ *and consider the stopping time* $T = t\mathbf{1}_{A^c} + s\mathbf{1}_A$.

3 Square integrable martingales

In later sections we will use properties of (continuous) *square integrable martingales*. Throughout this section we work with a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F}$ that satisfies the usual conditions.

3.1 Structural properties

Definition 3.1 A right-continuous martingale X is called *square integrable* if $\mathbb{E}\, X_t^2 < \infty$ for all $t \geq 0$. By $\mathcal{M}^2$ we denote the class of all square integrable martingales starting at zero and by $\mathcal{M}_c^2$ its subclass of *a.s continuous* square integrable martingales.

To study properties of $\mathcal{M}^2$ and $\mathcal{M}_c^2$ we endow these spaces with a metric and under additional assumptions with a norm.

Definition 3.2 Let $X \in \mathcal{M}^2$. We define for each $t \geq 0$ $||X||_t = (\mathbb{E}\, X_t^2)^{1/2}$, and $||X||_\infty = \sup ||X||_t$. The process X is called bounded in L^2 if $\sup ||X||_t < \infty$. For all $X \in \mathcal{M}^2$ we also define $||X|| = \sum_{n=1}^\infty 2^{-n}(||X||_n \wedge 1)$ and $d(X,Y) = ||X - Y||$ for all $X, Y \in \mathcal{M}^2$.

Remark 3.3 If we identify processes that are indistinguishable, then $(\mathcal{M}^2, d)$ becomes a metric space and on the subclass of martingales bounded in L^2 the operator $|| \cdot ||_\infty$ becomes a norm (see Exercise 3.1).

Proposition 3.4 *The metric space $(\mathcal{M}^2, d)$ is complete and $\mathcal{M}_c^2$ is a closed (w.r.t. the metric d) subspace of $\mathcal{M}^2$ and thus complete as well.*

Proof Let (X^m) be a Cauchy sequence in $(\mathcal{M}^2, d)$. Then for each fixed t the sequence (X_t^m) is Cauchy in the complete space $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ and thus has a limit, X_t say (which is the L^1-limit as well). We show that for $s \leq t$ one has $X_s = \mathbb{E}\,[X_t|\mathcal{F}_s]$. Consider for any $A \in \mathcal{F}_s$ the expectations $\mathbb{E}\,\mathbf{1}_A X_s$ and $\mathbb{E}\,\mathbf{1}_A X_t$. The former is the limit of $\mathbb{E}\,\mathbf{1}_A X_s^m$ and the latter the limit of $\mathbb{E}\,\mathbf{1}_A X_t^m$. But since each X^m is a martingale, one has $\mathbb{E}\,\mathbf{1}_A X_s^m = E\mathbf{1}_A X_t^m$, which yields the desired result. Choosing a right-continuous modification of the process X finishes the proof of the first assertion.

The proof of the second assertion is as follows. Let (X^m) be a sequence in $\mathcal{M}_c^2$, with limit $X \in \mathcal{M}^2$ say. We have to show that X is (almost surely) continuous. Using inequality (2.1), we have for every $\varepsilon > 0$ that

$$\mathbb{P}(\sup_{t \leq T} |X_t^m - X_t| > \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E}\,(X_T^m - X_T)^2 \to 0,\ m \to \infty.$$

Hence, for every $k \in \mathbb{N}$ there is m_k such that $\mathbb{P}(\sup_{t \leq T} |X_t^{m_k} - X_t| > \varepsilon) \leq 2^{-k}$. By the Borel-Cantelli lemma one has $\mathbb{P}(\liminf\{\sup_{t \leq T} |X_t^{m_k} - X_t| \leq \varepsilon\}) = 1$. Hence for all $T > 0$ and for almost all ω the functions $t \mapsto X^{m_k}(t, \omega) : [0, T] \to \mathbb{R}$ converge uniformly, which entails that limit functions $t \mapsto X(t, \omega)$ are continuous on each interval $[0, T]$. A.s. uniqueness of the limit on all these intervals with integer T yields a.s. continuity of X on $[0, \infty)$. □

3.2 Quadratic variation

Of a martingale $X \in \mathcal{M}^2$ we know that X^2 is a nonnegative submartingale. Therefore we can apply the Doob-Meyer decomposition, see Theorem 2.16 and Exercise 2.8, to obtain

$$X^2 = A + M, \tag{3.1}$$

where A is a natural increasing process and M a martingale that starts in zero. We also know from Theorem 2.17 and Exercise 2.13 that A and M are continuous if $X \in \mathcal{M}^2_c$.

Definition 3.5 For a process X in $\mathcal{M}^2$ the process A in the decomposition (3.1) is called the *quadratic variation process* and it is denoted by $\langle X \rangle$. So, $\langle X \rangle$ is the unique natural increasing process that makes $X^2 - \langle X \rangle$ a martingale.

Proposition 3.6 *Let X be a martingale in $\mathcal{M}^2$ and T a stopping time. Then the processes $\langle X^T \rangle$ and $\langle X \rangle^T$ are indistinguishable.*

Proof This is an immediate consequence of Proposition 2.18. □

The term quadratic variation process for the process $\langle X \rangle$ will be explained for $X \in \mathcal{M}^2_c$ in Proposition 3.8 below. We first introduce some notation. Let $\Pi = \{t_0, \ldots, t_n\}$ be a partition of the interval $[0,t]$ for some $t > 0$ with $0 = t_0 < \cdots < t_n = t$. For a process X we define

$$V_t(X, \Pi) = \sum_{j=1}^{n} (X_{t_j} - X_{t_{j-1}})^2.$$

Notice that $V_t(X,\Pi)$ is $\mathcal{F}_t$-measurable if X is adapted. The mesh $\mu(\Pi)$ of the partition is defined by $\mu(\Pi) = \max\{|t_j - t_{j-1}|, j = 1, \ldots, n\}$.

For any process X we denote by $m_T(X,\delta)$ the modulus of continuity:

$$m_T(X,\delta) = \max\{|X_t - X_s| : 0 \leq s,t \leq T, |s-t| < \delta\}.$$

If X is an a.s. continuous process, then for all $T > 0$ it holds that $m_T(X,\delta) \to 0$ a.s. for $\delta \to 0$.

The characterization of the quadratic variation process of Proposition 3.8 below will be proved using the following lemma.

Lemma 3.7 *Let $M \in \mathcal{M}^2$ and $t > 0$. Then for all $t \geq s$ one has*

$$\mathbb{E}\left[(M_t - M_s)^2 | \mathcal{F}_s\right] = \mathbb{E}\left[M_t^2 - M_s^2 | \mathcal{F}_s\right] \tag{3.2}$$

and

$$\mathbb{E}\left[(M_t - M_s)^2 - \langle M \rangle_t + \langle M \rangle_s | \mathcal{F}_s\right] = 0 \tag{3.3}$$

For all $t_1 \leq \cdots \leq t_n$ it holds that

$$\mathbb{E}\,[\sum_{j=i+1}^{n}(M_{t_j}-M_{t_{j-1}})^2|\mathcal{F}_{t_i}]=\mathbb{E}\,[M_{t_n}^2-M_{t_i}^2|\mathcal{F}_{t_i}]. \tag{3.4}$$

If M is bounded by a constant $K>0$, then

$$\mathbb{E}\,[\sum_{j=i+1}^{n}(M_{t_j}-M_{t_{j-1}})^2|\mathcal{F}_{t_i}]\leq K^2. \tag{3.5}$$

Proof Equation (3.2) follows by a simple computation and the martingale property of M. Then Equation (3.4) follows simply by iteration and reconditioning. Equation (3.3) follows from (3.2) and the definition of the quadratic variation process. Finally, Equation (3.5) is an immediate consequence of (3.4). □

Proposition 3.8 *Let X be in $\mathcal{M}_c^2$. Then $V_t(X,\Pi)$ converges in probability to $\langle X\rangle_t$ as $\mu(\Pi)\to 0$: for all $\varepsilon>0,\eta>0$ there exists a $\delta>0$ such that $\mathbb{P}(|V_t(X,\Pi)-\langle X\rangle_t|>\eta)<\varepsilon$ whenever $\mu(\Pi)<\delta$.*

Proof Supposing that we had already proven the assertion for bounded continuous martingales with bounded quadratic variation, we argue as follows. Let X be an arbitrary element of $\mathcal{M}_c^2$ and define for each $n\in\mathbb{N}$ the stopping times $T^n=\inf\{t\geq 0:|X_t|\geq n \text{ or } \langle X\rangle_t\geq n\}$. The T^n are stopping times in view of Proposition 1.10. Then

$$\{|V_t(X,\Pi)-\langle X\rangle_t|>\eta\}\subset\{T^n\leq t\}\cup\{|V_t(X,\Pi)-\langle X\rangle_t|>\eta,T^n>t\}.$$

The probability on the first event on the right hand side obviously tends to zero for $n\to\infty$. The second event is contained in $\{|V_t(X^{T^n},\Pi)-\langle X\rangle_t^{T^n}|>\eta\}$. In view of Proposition 3.6 we can rewrite it as $\{|V_t(X^{T^n},\Pi)-\langle X^{T^n}\rangle_t|>\eta\}$ and its probability can be made arbitrarily small, since the processes X^{T^n} and $\langle X\rangle^{T^n}=\langle X^{T^n}\rangle$ are both bounded, by what we supposed at the beginning of the proof.

We now show that the proposition holds for bounded continuous martingales with bounded quadratic variation. Actually we will show that we even have L^2-convergence and so we consider $\mathbb{E}\,(V_t(X,\Pi)-\langle X\rangle_t)^2$. Take a partition $\Pi=\{0=t_0<t_1<\cdots<t_n=t\}$ and put

$$M_k=\sum_{j=1}^{k}\big((X_{t_j}-X_{t_{j-1}})^2-(\langle X\rangle_{t_j}-\langle X\rangle_{t_{j-1}})\big).$$

This gives a martingale w.r.t. the filtration with σ-algebras $\mathcal{F}_{t_k}$. Indeed, using (3.2), we get

$$\begin{aligned}\mathbb{E}\,[M_k-M_{k-1}|\mathcal{F}_{t_{k-1}}]&=\mathbb{E}\,[(X_{t_k}-X_{t_{k-1}})^2-(\langle X\rangle_{t_k}-\langle X\rangle_{t_{k-1}})|\mathcal{F}_{t_{k-1}}]\\&=\mathbb{E}\,[X_{t_k}^2-X_{t_{k-1}}^2-(\langle X\rangle_{t_k}-\langle X\rangle_{t_{k-1}})|\mathcal{F}_{t_{k-1}}]\\&=0,\end{aligned}$$

because of (3.3). In view of Pythagoras's rule for square integrable martingales, we also have

$$\mathbb{E}\, M_n^2 = \sum_{k=1}^{n} \mathbb{E}\,(M_k - M_{k-1})^2.$$

Since $M_n = V_t(X,\Pi) - \langle X\rangle_t$, we get

$$\begin{aligned}
\mathbb{E}\,(V_t(X,\Pi) - \langle X\rangle_t)^2 &= \mathbb{E} \sum_k \big((X_{t_k} - X_{t_{k-1}})^2 - (\langle X\rangle_{t_k} - \langle X\rangle_{t_{k-1}})\big)^2 \\
&\leq 2 \sum_k \big(\mathbb{E}\,(X_{t_k} - X_{t_{k-1}})^4 + \mathbb{E}\,((\langle X\rangle_{t_k} - \langle X\rangle_{t_{k-1}})^2)\big) \\
&\leq 2 \sum_k \mathbb{E}\,(X_{t_k} - X_{t_{k-1}})^4 + \mathbb{E}\,\big(m_t(\langle X\rangle, \mu(\Pi))\langle X\rangle_t\big).
\end{aligned}$$

The second term in the last expression goes to zero in view of the bounded convergence theorem and continuity of $\langle X\rangle$ when $\mu(\Pi) \to 0$. We henceforth concentrate on the first term. First we bound the sum by

$$\mathbb{E} \sum_k (X_{t_k} - X_{t_{k-1}})^2 m_t(X, \mu(\Pi))^2 = \mathbb{E}\, V_t(X,\Pi) m_t(X,\mu(\Pi))^2,$$

which is by the Schwarz inequality less then $\big(\mathbb{E}\, V_t(X,\Pi)^2 \mathbb{E}\, m_t(X,\mu(\Pi))^4\big)^{1/2}$. By application of the dominated convergence theorem the last expectation tends to zero if $\mu(\Pi) \to 0$, so we have finished the proof as soon as we can show that $\mathbb{E}\, V_t(X,\Pi)^2$ stays bounded. Let $K > 0$ be an upper bound for X. Then, a two fold application of (3.5) leads to the inequalities below,

$$\begin{aligned}
&\mathbb{E}\,\big(\sum_k (X_{t_k} - X_{t_{k-1}})^2\big)^2 \\
&= \mathbb{E} \sum_k (X_{t_k} - X_{t_{k-1}})^4 + 2\mathbb{E} \sum_{i<j} \mathbb{E}\,[(X_{t_j} - X_{t_{j-1}})^2 (X_{t_i} - X_{t_{i-1}})^2 | \mathcal{F}_{t_i}] \\
&\leq 4K^2 \mathbb{E} \sum_k (X_{t_k} - X_{t_{k-1}})^2 + 2K^2 \mathbb{E} \sum_i (X_{t_i} - X_{t_{i-1}})^2 \\
&\leq 6K^4.
\end{aligned}$$

This finishes the proof. □

Having defined the quadratic variation (in Definition 3.5) of a square integrable martingale, we can also define the quadratic covariation (also called cross-variation) between two square integrable martingales X and Y. It is the process $\langle X, Y\rangle$ defined through the polarization formula

$$\langle X, Y\rangle = \frac{1}{4}(\langle X+Y\rangle - \langle X-Y\rangle).$$

Notice that the quadratic covariation process is a process of a.s. bounded variation. Moreover,

Proposition 3.9 *For all $X, Y \in \mathcal{M}^2$ the process $\langle X, Y \rangle$ is the unique process that can be written as the difference of two natural increasing processes such that the difference of XY and such a process is a martingale. Moreover, if $X, Y \in \mathcal{M}_c^2$, then the process $\langle X, Y \rangle$ is the unique continuous process of bounded variation such that the difference of XY and such a process is a martingale.*

Proof Exercise 3.4. □

3.3 Exercises

3.1 Let X be a martingale that is bounded in L^2. Then X_∞ exists as an a.s. limit of X_t, for $t \to \infty$. Show that $||X||_\infty = (\mathbb{E}\, X_\infty^2)^{1/2}$. Show also that $\{X \in \mathcal{M}^2 : ||X||_\infty < \infty\}$ is a vector space and that $||\cdot||_\infty$ is a norm on this space under the usual identification that two processes are "the same", if they are indistinguishable.

3.2 Give an example of a martingale (not continuous of course), for which the result of Proposition 3.8 doesn't hold. *Hint:* embed a very simple discrete time martingale in continuous time, by defining it constant on intervals of the type $[n, n+1)$.

3.3 Let $X, Y \in \mathcal{M}^2$. Show the following statements.

(a) $\langle X, Y \rangle = \frac{1}{2}(\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle)$.

(b) The quadratic covariation is a bilinear form.

(c) The Schwarz inequality $\langle X, Y \rangle_t^2 \leq \langle X \rangle_t \langle Y \rangle_t$ holds a.s. *Hint:* Show first that on a set with probability one has for all rational a and b $\langle aX + bY \rangle_t \geq 0$. Write this as a sum of three terms and show that the above property extends to real a and b. Use then that this defines a nonnegative quadratic form.

(d) If $V(s, t]$ denotes the total variation of $\langle X, Y \rangle$ over the interval $(s, t]$, then a.s. for all $t \geq s$

$$V(s, t] \leq \frac{1}{2}(\langle X \rangle_t + \langle Y \rangle_t - \langle X \rangle_s - \langle Y \rangle_s). \tag{3.6}$$

3.4 Prove Proposition 3.9.

3.5 Let $X \in \mathcal{M}^2$ and let T be a stopping time (not necessarily finite). If $\langle X \rangle_T = 0$, then $\mathbb{P}(X_{t \wedge T} = 0, \forall t \geq 0) = 1$.

3.6 If X is a martingale w.r.t. some filtration, it is also a martingale w.r.t. $\mathbb{F}^X$. Let X and Y be independent processes on some $(\Omega, \mathcal{F}, \mathbb{P})$ and assume that both are right-continuous martingales. Let $\mathcal{F}_t^0 = \sigma(X_s, Y_s,\, s \leq t)$ and let $\mathbb{F}$ be the filtration of σ-algebras $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^0 \vee \mathcal{N}$, where $\mathcal{N}$ is the collections of all $\mathbb{P}$-null sets of $\mathcal{F}$. Note that $\mathbb{P}$ can be uniquely extended to $\mathcal{F} \vee \mathcal{N}$. Show that X and Y are martingales w.r.t. $\mathbb{F}$ and that $\langle X, Y \rangle = 0$.

3.7 Let X be a nonnegative continuous process and A a continuous increasing process such $\mathbb{E}\, X_T \leq \mathbb{E}\, A_T$ for all bounded stopping times T. Define the process X^* by $X^*_t = \sup_{0\leq s\leq t} X_s$. Let now T be any stopping time. Show that

$$\mathbb{P}(X^*_T \geq \varepsilon, A_T < \delta) \leq \frac{1}{\varepsilon}\mathbb{E}\,(\delta \wedge A_T).$$

Deduce *Lenglart's inequality*:

$$\mathbb{P}(X^*_T \geq \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}\,(\delta \wedge A_T) + \mathbb{P}(A_T \geq \delta).$$

Finally, let $X = M^2$, where $M \in \mathcal{M}^2_c$. What is a good candidate for the process A in this case?

3.8 Formulate the analogue of Proposition 3.8 for the quadratic covariation process of two martingales in $\mathcal{M}^2_c$. Prove it by using the assertion of this proposition.

3.9 Consider an interval $[0, t]$ and let $\Pi_n = \{t^n_0, \ldots, t^n_{k_n}\}$ be a partition of it with $0 = t^n_0 < \cdots < t^n_{k_n} = t$ and denote by μ_n its mesh: $\mu_n = \max\{t^n_j - t^n_{j-1} : j = 1, \ldots, k_n\}$. Let W be a Brownian motion and $V^2_n = \sum_{j=1}^{k_n}(W(t^n_j) - W(t^n_{j-1}))^2$. Prove that $\mathbb{E}(V^2_n - t)^2 \to 0$ if $\mu_n \to 0$, and if moreover $\sum_{n=1}^{\infty} \mu_n < \infty$, then also $V^2_n \to t$ a.s.

3.10 Let W be a Brownian motion. Let X be the square integrable martingale defined by $X_t = W^2_t - t$. Show (use Proposition 3.8) that $\langle X\rangle_t = 4\int_0^t W^2_s\,\mathrm{d}s$.

3.11 Let N be a Poisson process, with intensity λ. Let $M_t = N_t - \lambda t$. Show that $\langle M\rangle_t = \lambda t$ for all $t \geq 0$. Show also that $V_t(M, \Pi) \to N_t$ a.s., when $\mu(\Pi) \to 0$.

3.12 Let $M \in \mathcal{M}^2$ have independent and stationary increments, the latter meaning that $M_{t+h} - M_t$ has the same distribution as $M_{s+h} - M_s$ for all $s, t, h > 0$. Show that $\langle M\rangle_t = t\,\mathbb{E}\,M^2_1$.

3.13 Show that the assertion of Proposition 3.8, under the same conditions, can be strengthened to $\mathbb{P}(\sup_{s\leq t} |V_s(X, \Pi) - \langle X\rangle_s| > \eta) < \varepsilon$ whenever $\mu(\Pi) < \delta$.

3.14 Let $M, N \in \mathcal{M}^2$, $F \in \mathcal{F}_{t_0}$ for some $t_0 \geq 0$. Show that X defined by $X_t = \mathbf{1}_F(M_t - M^{t_0}_t)$ also belongs to $\mathcal{M}^2$ and that $\langle X, N\rangle = \mathbf{1}_F\langle M - M^{t_0}, N\rangle$.

4 Local martingales

In proofs of results of previous sections we reduced a relatively difficult problem by stopping at judiciously chosen stopping times to an easier problem. This is a standard technique and it also opens the way to define wider classes of processes than the ones we have used thus far and that still share many properties of the more restricted classes. In this section we assume that the underlying filtration satisfies the usual conditions.

4.1 Localizing sequences and local martingales

Definition 4.1 A sequence of stopping times T^n is called a *fundamental* or a *localizing* sequence if $T^n \geq T^{n-1}$ for all $n \geq 1$ and if $\lim_{n\to\infty} T^n = \infty$ a.s. If $\mathcal{C}$ is a class of processes satisfying a certain property, then (usually) by $\mathcal{C}^{loc}$ we denote the class of processes X for which there exists a fundamental sequence of stopping times T^n such that all stopped processes X^{T^n} belong to $\mathcal{C}$.

In the sequel we will take for $\mathcal{C}$ various classes consisting of martingales with a certain property. The first one is defined in

Definition 4.2 An adapted right-continuous process X is called a *local martingale* if there exists a localizing sequence of stopping times T^n such that for every n the process $X^{T^n} = \{X_{T^n \wedge t}\ t \geq 0\}$ is a martingale. The class of martingales X with $X_0 = 0$ a.s. is denoted by $\mathcal{M}$ and the class of local martingales X with $X_0 = 0$ a.s. is denoted by $\mathcal{M}^{loc}$. The subclass of continuous local martingales X with $X_0 = 0$ a.s. is denoted by $\mathcal{M}_c^{loc}$.

Remark 4.3 In Definition 4.2 the stopped processes are required to be martingales. An alternative definition even requires uniform integrability, but one can show that this definition is equivalent to the former one, see Proposition 4.4.

In Definition 4.2, all $X_t^{T^n}$ have to be integrable. In particular for $t = 0$, one obtains that X_0 is integrable. There are other, not equivalent, definitions of local martingales. For instance one requires only that for every n the process $X^{T^n}\mathbf{1}_{\{T^n>0\}} = \{X_{T^n\wedge t}\mathbf{1}_{\{T^n>0\}} : t \geq 0\}$ is a martingale. Here we see the weaker requirement that $X_0\mathbf{1}_{\{T^n>0\}}$ has to be integrable for every n. Another choice is that one requires the processes $X^{T^n} - X_0$ to be martingales. As we will be mainly concerned with local martingales satisfying $X_0 = 0$ a.s., the difference between the definitions disappears. Also in these alternative definitions one can without loss of generality require uniform integrability.

By the optional stopping theorem, every martingale is a local martingale. This can be seen by choosing $T^n = n$ for all n. See also Exercise 4.7 for a characterization of a local martingale that makes use of the alternative definitions.

In almost all that follows we consider local martingales X with $X_0 = 0$. We have the following proposition.

Proposition 4.4 *Let X be an adapted right-continuous process such that $X_0 = 0$ a.s. Then the following statements are equivalent.*

(i) *X is a local martingale.*
(ii) *There exists a localizing sequence (T^n) such that the processes X^{T^n} are uniformly integrable martingales.*

Proof Exercise 4.1. □

4.2 Continuous local martingales

In the sequel we will deal mainly with continuous processes. The main results are as follows.

Proposition 4.5 *If X is a continuous local martingale with $X_0 = 0$ a.s., then there exist a localizing sequence of stopping times T^n such that the processes X^{T^n} are bounded martingales.*

Proof Exercise 4.2. □

Recall that called a (right-continuous) martingale is square integrable if $\mathbb{E}\, X_t^2 < \infty$ for all $t \geq 0$. Therefore we call a right-continuous adapted process X a *locally square integrable* martingale if $X_0 = 0$ and if there exists a localizing sequence of stopping times T^n such that the process X^{T^n} are all square integrable martingales. Obviously, one has that these processes are all local martingales. If we confine ourselves to continuous processes the difference disappears.

Proposition 4.6 *Let X be a continuous local martingale with $X_0 = 0$ a.s. Then it is also locally square integrable.*

Proof This follows immediately from Proposition 4.5. □

Some properties of local martingales are given in the next proposition.

Proposition 4.7
(i) *A local martingale of class* DL *is a martingale.*
(ii) *A local martingale of class* D *is a uniformly integrable martingale.*
(iii) *Any nonnegative local martingale is a supermartingale.*

Proof Exercise 4.3. □

Although local martingales X in general don't have a finite second moment, it is possible to define a quadratic variation process. We do this for continuous local martingales only and the whole procedure is based on the fact that for a localizing sequence (T^n) the processes X^{T^n} have a quadratic variation process in the sense of Section 3.2.

Proposition 4.8 *Let $X \in \mathcal{M}_c^{loc}$. Then there exists a unique (up to indistinguishability) continuous process $\langle X \rangle$ with a.s. increasing paths such that $X^2 - \langle X \rangle \in \mathcal{M}_c^{loc}$.*

Proof Choose stopping times T^n such that the stopped processes X^{T^n} are bounded martingales. This is possible in view of Proposition 4.5. Then, for each n there exists a unique natural (even continuous) increasing process A^n such that $(X^{T^n})^2 - A^n$ is a martingale. For $n > m$ we have that $(X^{T^n})^{T^m} = X^{T^m}$. Hence, by the uniqueness of the Doob-Meyer decomposition, one has $(A^n)^{T^m} = A^m$. So we can unambiguously define $\langle X\rangle$ by setting $\langle X\rangle_t = A^n_t$ if $t < T^n$. Moreover, we have

$$X^2_{t\wedge T^n} - \langle X\rangle_{t\wedge T^n} = (X^{T^n}_t)^2 - A^n_t,$$

which shows that for all n the process $(X^2)^{T^n} - \langle X\rangle^{T^n}$ is a martingale, and thus that $X^2 - \langle X\rangle$ is a (continuous) local martingale. □

Corollary 4.9 *If X and Y belong to $\mathcal{M}^{loc}_c$, then there exists a unique (up to indistinguishability) continuous process $\langle X, Y\rangle$ with paths of bounded variation a.s. such that $XY - \langle X, Y\rangle \in \mathcal{M}^{loc}_c$.*

Proof Exercise 4.4. □

Proposition 4.10 *Let X and Y belong to $\mathcal{M}^{loc}_c$ and T a stopping time. Then $\langle X, Y\rangle^T = \langle X^T, Y\rangle = \langle X^T, Y^T\rangle$.*

Proof Exercise 4.9. □

4.3 Exercises

4.1 Prove Proposition 4.4.

4.2 Prove Proposition 4.5.

4.3 Prove Proposition 4.7.

4.4 Prove Corollary 4.9.

4.5 Let $M \in \mathcal{M}^{loc}_c$ and let S be a stopping time. Put $X_t = M^2_t$ and define $X_\infty = \liminf_{t\to\infty} X_t$ (and $\langle M\rangle_\infty = \lim_{t\to\infty}\langle M\rangle_t$). Show that $\mathbb{E}\, X_S \leq \mathbb{E}\,\langle M\rangle_S$.

4.6 Let $M \in \mathcal{M}^2_c$ satisfy the property $\mathbb{E}\,\langle M\rangle_\infty < \infty$. Deduce from Exercise 4.5 that $\{M_S : S \text{ a finite stopping time}\}$ is uniformly integrable and hence that M_∞ is well defined as a.s. limit of M_t. Show that $\mathbb{E}\, M^2_\infty = \mathbb{E}\,\langle M\rangle_\infty$.

4.7 Let X be an adapted process and T a stopping time. Show that $X^T\mathbf{1}_{\{T>0\}}$ is a uniformly integrable martingale iff $X_0\mathbf{1}_{\{T>0\}}$ is integrable and $X^T - X_0$ is a uniformly integrable martingale. (*The latter property for a fundamental sequence of stopping times is also used as definition of local martingale.*)

4.8 Let $M \in \mathcal{M}^{loc}_c$ satisfy the property $\langle M\rangle_T = 0$ for some stopping time T. Show that $\mathbb{P}(M^T_t = 0, \forall t \geq 0) = 1$. See also Exercise 3.5.

4.9 Prove Proposition 4.10.

4.10 Let T^n be increasing stopping times such that $T^n \to \infty$ and let M be a process such that $M^{T^n} \in \mathcal{M}^{loc}_c$ for all n. Show that $M \in \mathcal{M}^{loc}_c$.

5 Spaces of progressive processes

Throughout this section we work with a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that is endowed with a filtration $\mathbb{F}$ that satisfies the usual conditions. All properties below that are defined relative to a filtration (such as adaptedness) are assumed to be defined in terms of this filtration.

5.1 Doléans measure

In this section we often work with (Lebesgue-Stieltjes) integrals w.r.t. a process of finite variation A that satisfies $A_0 = 0$. So, for an arbitrary process X with the right measurability properties we will look at integrals of the form

$$\int_{[0,T]} X_t(\omega)\,\mathrm{d}A_t(\omega), \tag{5.1}$$

with t the integration variable and where this integral has to be evaluated ω-wise. We follow the usual convention for random variables by omitting the variable ω. But in many cases we also omit the integration variable t and hence an expression like (5.1) will often be denoted by

$$\int_{[0,T]} X\,\mathrm{d}A.$$

We also use the notation $\int_0^T$ instead of $\int_{(0,T]}$ if the process A has a.s. continuous paths. Let M be a square integrable continuous martingale ($M \in \mathcal{M}^2_c$). Recall that $\langle M\rangle$ is the unique continuous increasing process such that $M^2 - \langle M\rangle$ is a martingale.

Definition 5.1 The *Doléans measure* μ_M on $([0,\infty) \times \Omega, \mathcal{B}([0,\infty)) \times \mathcal{F})$ is defined by

$$\mu_M(A) = \int_\Omega \int_0^\infty \mathbf{1}_A(t,\omega)\,\mathrm{d}\langle M\rangle_t(\omega)\mathbb{P}(d\omega) = \mathbb{E}\int_0^\infty \mathbf{1}_A\,\mathrm{d}\langle M\rangle. \tag{5.2}$$

For a measurable, adapted process X we define for every $T \in [0,\infty)$

$$||X||_{M,T} = (\mathbb{E}\int_0^T X_t^2\,\mathrm{d}\langle M\rangle_t)^{1/2} = (\int X^2 \mathbf{1}_{[0,T]\times\Omega}\,\mathrm{d}\mu_M)^{1/2} \tag{5.3}$$

and

$$||X||_M = \sum_{n=1}^\infty 2^{-n}(1 \wedge ||X||_{M,n}). \tag{5.4}$$

We will also use

$$||X||_{M,\infty} = (\mathbb{E}\int_0^\infty X_t^2\,\mathrm{d}\langle M\rangle_t)^{1/2} = (\int X^2\,\mathrm{d}\mu_M)^{1/2}. \tag{5.5}$$

Note that $||\cdot||_{M,\infty}$ is nothing else but the L^2-norm for the Doléans measure. We call two measurable, adapted processes X and Y (M-)equivalent if $||X-Y||_M = 0$. By $\mathcal{P}$ we denote the class of progressive processes X for which $||X||_{M,T} < \infty$ for all $T \in [0,\infty)$.

Remark 5.2 It is not obvious that that μ_M as in (5.2) indeed defines a measure on $([0,\infty) \times \Omega, \mathcal{B}([0,\infty)) \times \mathcal{F})$. The inner integral in (5.2) is well defined. But is the resulting expression $\mathcal{F}$-measurable? The latter is needed to have the expectation well defined. It is easy to check that the answer to this question is affirmative if A is a disjoint union of sets of the form $(a,b] \times F$ with $F \in \mathcal{F}$. The extension of μ_M as a measure to the full product σ-algebra $\mathcal{B}([0,\infty)) \times \mathcal{F}$ can be guaranteed to exist by Carathéodory's extension theorem and requires some work.

Remark 5.3 Notice that, with time restricted to $[0,T]$, the function $||\cdot||_{M,T}$ defines an L^2-norm on the space of measurable, adapted processes if we identify equivalent processes. Similarly, $d_M(X,Y) := ||X-Y||_M$ defines a metric. Notice also that X and Y are equivalent iff $\int_0^T (X-Y)^2 \,\mathrm{d}\langle M\rangle = 0$ a.s. for all $T \in [0,\infty)$.

In addition to the class $\mathcal{P}$ introduced above we also need the classes $\mathcal{P}_T$ for $T \in [0,\infty]$. These are defined in

Definition 5.4 For $T < \infty$ the class $\mathcal{P}_T$ is the set of processes X in $\mathcal{P}$ for which $X_t = 0$ if $t > T$. The class $\mathcal{P}_\infty$ is the subclass of processes $X \in \mathcal{P}$ for which $\mathbb{E} \int_0^\infty X^2 \,\mathrm{d}\langle M\rangle < \infty$.

Remark 5.5 A process X belongs to $\mathcal{P}_T$ iff $X_t = 0$ for $t > T$ and $||X||_{M,T} < \infty$.

Remark 5.6 All the classes, norms and metrics above depend on the martingale M. When other martingales than M play a role in a certain context, we emphasize this dependence by e.g. writing $\mathcal{P}_T(M)$, $\mathcal{P}(M)$, etc.

Proposition 5.7 *For $T \leq \infty$ the class $\mathcal{P}_T$ is a Hilbert space with inner product given by*

$$(X,Y)_T = \mathbb{E} \int_0^T XY \,\mathrm{d}\langle M\rangle,$$

if we identify two processes X and Y that satisfy $||X-Y||_{M,T} = 0$.

Proof Let $T < \infty$ and let (X^n) be a Cauchy sequence in $\mathcal{P}_T$. Since $\mathcal{P}_T$ is a subset of the Hilbert space $L^2([0,T] \times \Omega, \mathcal{B}([0,T]) \times \mathcal{F}_T, \mu_M)$, the sequence (X^n) has a limit X in this space. We have to show that $X \in \mathcal{P}_T$, but it is not clear that the limit process X is progressive (a priori we can only be sure that X is $\mathcal{B}([0,T]) \times \mathcal{F}_T$-measurable). We will replace X with an equivalent process as follows. First we select a subsequence X^{n_k} that converges to X almost everywhere w.r.t. μ_M. We set $Y_t(\omega) = \limsup_{k\to\infty} X_t^{n_k}(\omega)$. Then Y is a progressive process, since the X^n are progressive and $||X-Y||_{M,T} = 0$, so Y is equivalent to X and $||X^n - Y||_{M,T} \to 0$. For $T = \infty$ the proof is similar. $\square$

Finally we enlarge the classes $\mathcal{P}_T$ by dropping the requirement that the expectations are finite and by relaxing the condition that $M \in \mathcal{M}_c^2$. We have

Definition 5.8 Let $M \in \mathcal{M}_c^{loc}$. The class $\mathcal{P}^*$ is the equivalence class of progressive processes X such that $\int_0^T X^2 \, \mathrm{d}\langle M \rangle < \infty$ a.s. for every $T \in [0, \infty)$.

Remark 5.9 If $M \in \mathcal{M}_c^{loc}$, there exist a fundamental sequence of stopping times R^n such that $M^{R^n} \in \mathcal{M}_c^2$. If we take $X \in \mathcal{P}^*$, then the bounded stopping times $S^n = n \wedge \inf\{t \geq 0 : \int_0^t X^2 \, \mathrm{d}\langle M \rangle \geq n\}$ also form a fundamental sequence. Consider then the stopping times $T^n = R^n \wedge S^n$. These form a fundamental sequence as well. Moreover, $M^{T^n} \in \mathcal{M}_c^2$ and the processes X^n defined by $X_t^n = X_t \mathbf{1}_{\{t \leq T^n\}}$ belong to $\mathcal{P}(M^{T^n})$.

5.2 Simple processes

We start with a definition.

Definition 5.10 A process X is called *simple* if there exists a strictly increasing sequence of real numbers t_n with $t_0 = 0$ and $t_n \to \infty$, a uniformly bounded sequence of random variables ξ_n with the property that ξ_n is $\mathcal{F}_{t_n}$-measurable for each n, such that

$$X_t = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{n=0}^{\infty} \xi_n \mathbf{1}_{(t_n, t_{n+1}]}(t), \; t \geq 0.$$

The class of simple processes is denoted by $\mathcal{S}$.

Remark 5.11 Notice that simple processes are progressive and bounded. If $M \in \mathcal{M}_c^2$, then a simple process X belongs to $\mathcal{P} = \mathcal{P}(M)$.

The following lemma is crucial for the construction of the stochastic integral.

Lemma 5.12 *Let X be a bounded progressive process. Then there exists a sequence of simple processes X^n such that for all $T \in [0, \infty)$ one has*

$$\lim_{n \to \infty} \mathbb{E} \int_0^T (X_t^n - X_t)^2 \, \mathrm{d}t = 0. \tag{5.6}$$

Proof Suppose that we had found for each $T \in [0, \infty)$ a sequence of simple processes $X^{n,T}$ (depending on T) such that (5.6) holds. Then for all integers n there exist integers m_n such that

$$\mathbb{E} \int_0^n (X_t^{m_n, n} - X_t)^2 \, \mathrm{d}t \leq \frac{1}{n}.$$

One verifies that the sequence with elements $X^n = X^{m_n, n}$ then has the asserted property: fix $T > 0$, choose $n > T$ and verify that

$$\mathbb{E} \int_0^T (X_t^n - X_t)^2 \, \mathrm{d}t \leq \frac{1}{n}.$$

Therefore, we will keep T fixed in the remainder of the proof and construct a sequence of simple processes X^n for which (5.6) holds. This is relatively easy if X is continuous. Consider the sequence of approximating processes (slightly different from the ones we used in the proof of Proposition 1.7)

$$X^n = X_0 \mathbf{1}_{\{0\}}(\cdot) + \sum_{k=1}^{2^n} \mathbf{1}_{((k-1)2^{-n}T, k2^{-n}T]}(\cdot) X_{(k-1)2^{-n}T}.$$

This sequence has the desired property in view of the bounded convergence theorem.

If X is merely progressive (but bounded), we proceed by first approximating it with continuous processes to which we apply the preceding result. For $t \leq T$ we define the bounded continuous ('primitive') process F by $F_t(\omega) = \int_0^t X_s(\omega)\,\mathrm{d}s$ for $t \geq 0$ and $F_t(\omega) = 0$ for $t < 0$ and for each integer m

$$Y_t^m(\omega) = m(F_t(\omega) - F_{t-1/m}(\omega)).$$

By the fundamental theorem of calculus, for each ω, the set of t for which $Y_t^m(\omega)$ does not converge to $X_t(\omega)$ for $m \to \infty$ has Lebesgue measure zero. Hence, by dominated convergence, we also have $\mathbb{E} \int_0^T (Y_t^m - X_t)^2\,\mathrm{d}t \to 0$. By the preceding result we can approximate each of the continuous Y^m by simple processes $Y^{m,n}$ in the sense that $\mathbb{E} \int_0^T (Y_t^m - Y_t^{m,n})^2\,\mathrm{d}t \to 0$. But then we can also find a sequence of simple processes $X^m = Y^{m,n_m}$ for which $\mathbb{E} \int_0^T (X_t^m - X_t)^2\,\mathrm{d}t \to 0$. $\square$

The preceding lemma enables us to prove

Proposition 5.13 *Let M be in $\mathcal{M}_c^2$ and assume that there exists a progressive nonnegative process f such that $\langle M \rangle$ is indistinguishable from $\int_0^\cdot f_s\,\mathrm{d}s$ (the process $\langle M \rangle$ is said to be a.s. absolutely continuous). Then the set $\mathcal{S}$ of simple processes is dense in $\mathcal{P}$ with respect to the metric d_M defined in Remark 5.3.*

Proof Let $X \in \mathcal{P}$ and assume that X is bounded. By Lemma 5.12 we can find simple processes X^n such that for all $T > 0$ it holds that $\mathbb{E} \int_0^T (X^n - X)^2\,\mathrm{d}t \to 0$. But then we can select a subsequence (X^{n_k}) such that $X^{n_k} \to X$ for $dt \times \mathbb{P}$-almost all (t, ω). By the dominated convergence theorem we then also have for all $T > 0$ that $\mathbb{E} \int_0^T (X^{n_k} - X)^2 f_t\,\mathrm{d}t \to 0$.

If X is not bounded we truncate it and introduce the processes $X^n = X\mathbf{1}_{\{|X| \leq n\}}$. Each of these bounded processes can be approximated by simple processes in view of the previous case. The result then follows upon noticing that $\mathbb{E} \int_0^T (X^n - X)^2\,\mathrm{d}\langle M \rangle = \mathbb{E} \int_0^T X^2 \mathbf{1}_{\{|X| > n\}}\,\mathrm{d}\langle M \rangle \to 0$. $\square$

Remark 5.14 The approximation results can be strengthened. For instance, in the previous lemma we didn't use progressive measurability. The space $\mathcal{S}$ is also dense in the set of measurable processes. Furthermore, if we drop the requirement that the process $\langle M \rangle$ is a.s. absolutely continuous, the assertion of Proposition 5.13 is still true, but the proof is much more complicated. For most, if not all, of our purposes the present version is sufficient.

5.3 Exercises

5.1 Let X be a simple process given by $X_t = \sum_{k=1}^{\infty} \xi_{k-1} \mathbf{1}_{(t_{k-1},t_k]}(t)$ and let M be an element of $\mathcal{M}_c^2$. Consider the discrete time process V defined by $V_n = \sum_{k=1}^{n} \xi_{k-1}(M_{t_k} - M_{t_{k-1}})$.

(a) Show that V is a martingale w.r.t. an appropriate filtration $\mathbb{G} = (\mathcal{G}_n)$ in discrete time.

(b) Compute $A_n := \sum_{k=1}^{n} \mathbb{E}\,[(V_k - V_{k-1})^2 | \mathcal{G}_{k-1}]$.

(c) Compute also $\sum_{k=1}^{n} \mathbb{E}\,[\int_{t_{k-1}}^{t_k} X_t^2 \,\mathrm{d}\langle M \rangle_t | \mathcal{G}_{k-1}]$ and show that this is equal to A_n.

(d) Finally, show that $V_n^2 - A_n$ is a $\mathbb{G}$-martingale.

5.2 Let W be a Brownian motion and let for each n a partition $\Pi^n = \{0 = t_0^n, \ldots, t_n^n = T\}$ of $[0,T]$ be given with $\mu(\Pi^n) \to 0$ for $n \to \infty$. Let $h_n : \mathbb{R} \to \mathbb{R}$ be defined by $h_n(x) = x\mathbf{1}_{[-n,n]}(x)$ and put

$$\bar{W}_t^n = \sum_{j=1}^{n} W_{t_{j-1}^n} \mathbf{1}_{(t_{j-1}^n, t_j^n]}(t).$$

and

$$W_t^n = \sum_{j=1}^{n} h_n(W_{t_{j-1}^n}) \mathbf{1}_{(t_{j-1}^n, t_j^n]}(t).$$

Then $W^n \in \mathcal{S}$ for all n. Show that $\bar{W}^n \to W$ and $W^n \to W$ in $\mathcal{P}_T(W)$.

6 Stochastic Integral

In previous sections we have already encountered integrals where both the integrand and the integrator were stochastic processes, e.g. in the definition of a natural process. In all these cases the integrator was an increasing process or, more general, a process with paths of bounded variation over finite intervals. In the present section we will consider integrals where the integrator is a continuous martingale. Except for trivial exceptions, these have paths of unbounded variation so that a pathwise definition of these integrals in the Lebesgue-Stieltjes sense cannot be given (e.g. Proposition C.5 cannot be applied). As a matter of fact, if one aims at a sensible pathwise definition of such an integral, one finds himself in a (seemingly) hopeless position in view of Proposition 6.1 below. For integrals, over the interval $[0,1]$ say, defined in the Stieltjes sense, the sums

$$S_\Pi h = \sum h(t_k)(g(t_{k+1}) - g(t_k))$$

converge for continuous h and g of bounded variation, if we sum over the elements of partitions $\Pi = \{0 = t_0 < \cdots < t_n = 1\}$ whose mesh $\mu(\Pi)$ tends to zero (Proposition C.5).

Proposition 6.1 *Suppose that the fixed function g is such that for all continuous functions h one has that $S_\Pi h$ converges, if $\mu(\Pi) \to 0$. Then g is of bounded variation.*

Proof We view the S_Π as linear operators on the Banach space of continuous functions on $[0,1]$ endowed with the sup-norm $||\cdot||$. Notice that $|S_\Pi h| \leq ||h|| \sum |g(t_{k+1}) - g(t_k)| = ||h|| V^1(g;\Pi)$, where $V^1(g;\Pi)$ denotes the variation of g over the partition Π. Hence the operator norm $||S_\Pi||$ is less than $V^1(g;\Pi)$. For any partition $\Pi = \{0 = t_0 < \cdots < t_n = 1\}$ we can find (by linear interpolation) a continuous function h_Π (bounded by 1) such that $h_\Pi(t_k) = \text{sgn}(g(t_{k+1}) - g(t_k))$. Then we have $S_\Pi h_\Pi = V^1(g;\Pi)$. It follows that $||S_\Pi|| = V^1(g;\Pi)$. By assumption, for any h we have that the sums $S_\Pi h$ converge if $\mu(\Pi) \to 0$, so that for any h the set with elements $|S_\Pi h|$ (for such Π) is bounded. By the Banach-Steinhaus theorem (Theorem E.5), also the $||S_\Pi||$ form a bounded set. The result follows since we had already observed that $||S_\Pi|| = V^1(g;\Pi)$. □

The function h in the above proof evaluated at points t_k uses the value of g at a 'future' point t_{k+1}. Excluding functions that use 'future' information, one also says that such functions are anticipating, is one of the ingredients that allow us to nevertheless finding a coherent notion of the (stochastic) integral with martingales as integrator.

6.1 Construction

The basic formula for the construction of the stochastic integral is formula (6.1) below. We consider a process $X \in \mathcal{S}$ as in Definition 5.10. The stochastic integral, also called Itô integral of X w.r.t. $M \in \mathcal{M}_c^2$ is a stochastic process, denoted by $X \cdot M$, that at each time t is defined to be the random variable

$$(X \cdot M)_t = \sum_{i=0}^{\infty} \xi_i (M_{t \wedge t_{i+1}} - M_{t \wedge t_i}). \tag{6.1}$$

For each $t \in [0, \infty)$ there is a unique $n = n(t)$ such that $t_n \le t < t_{n+1}$. Hence Equation (6.1) then takes the form

$$(X \cdot M)_t = \sum_{i=0}^{n-1} \xi_i (M_{t_{i+1}} - M_{t_i}) + \xi_n (M_t - M_{t_n}). \tag{6.2}$$

As in classical Lebesgue integration theory one can show that the given expression for $(X \cdot M)_t$ is independent of the chosen representation of X. Notice that (6.2) expresses $(X \cdot M)_t$ as a *martingale transform*. With this observation we now present the first properties of the stochastic integral.

Proposition 6.2 *For $X \in \mathcal{S}$ and $M \in \mathcal{M}_c^2$ the process $X \cdot M$ is a continuous square integrable martingale with initial value zero, quadratic variation process*

$$\langle X \cdot M \rangle = \int_0^{\cdot} X_u^2 \, \mathrm{d}\langle M \rangle_u, \tag{6.3}$$

and

$$\mathbb{E} \, (X \cdot M)_t^2 = \mathbb{E} \int_0^t X_u^2 \, \mathrm{d}\langle M \rangle_u. \tag{6.4}$$

Thus, for each fixed t, we can view $I_t : X \to (X \cdot M)_t$ as a linear operator on the space of simple processes that are annihilated after t with values in $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ It follows that $||X \cdot M||_t = ||X||_{M,t}$, where the first $||\cdot||_t$ is as in Definition 3.2 and the second $||\cdot||_{M,t}$ as in Definition 5.1. Hence I_t is an isometry. With the pair of 'norms' $||\cdot||$ and $||\cdot||_M$ in the same definitions, we also have $||X \cdot M|| = ||X||_M$.

Proof From (6.1) it follows that $X \cdot M$ is a continuous process. Consider the identities

$$\mathbb{E} \, [(X \cdot M)_t | \mathcal{F}_s] = (X \cdot M)_s \text{ a.s.}, \tag{6.5}$$

$$\mathbb{E} \, [((X \cdot M)_t - (X \cdot M)_s)^2] | \mathcal{F}_s] = \mathbb{E} \, [\int_s^t X_u^2 \, \mathrm{d}\langle M \rangle_u | \mathcal{F}_s] \text{ a.s.} \tag{6.6}$$

To prove (6.5) and (6.6), we assume without loss of generality that $t = t_n$ and $s = t_m$. Then $(X \cdot M)_t - (X \cdot M)_s = \sum_{i=m}^{n-1} \xi_i (M_{t_{i+1}} - M_{t_i})$. In fact, the sequence $(X \cdot M)_{t_k}$ can be considered as a discrete time martingale transform, and so (6.5) follows from the one step ahead version $\mathbb{E}\,[\xi_i(M_{t_{i+1}} - M_{t_i})|\mathcal{F}_{t_i}] = 0$.

Similarly, we have

$$\begin{aligned}\mathbb{E}\,[(\xi_i(M_{t_{i+1}} - M_{t_i}))^2|\mathcal{F}_{t_i}] &= \xi_i^2\mathbb{E}\,[(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{t_i}] \\ &= \xi_i^2\mathbb{E}\,[\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i}|\mathcal{F}_{t_i}] \\ &= \mathbb{E}\,[\int_{t_i}^{t_{i+1}} X_u^2 \,\mathrm{d}\langle M\rangle_u|\mathcal{F}_{t_i}],\end{aligned}$$

from which (6.6) follows. We conclude that $X \cdot M \in \mathcal{M}_c^2$ and obtain the expression (6.3) for its quadratic variation process. Then we also have (6.4) and the equality $||X \cdot M||_t = ||X||_{M,t}$ immediately follows, as well as $||X \cdot M|| = ||X||_M$. Linearity of I_t can be proved as in Lebesgue integration theory. □

Theorem 6.3 *Let $M \in \mathcal{M}_c^2$. For all $X \in \mathcal{P}$, there exists a unique (up to indistinguishability) process $X \cdot M \in \mathcal{M}_c^2$ with the property $||X^n \cdot M - X \cdot M|| \to 0$ for every sequence (X^n) in $\mathcal{S}$ such that $||X^n - X||_M \to 0$. Moreover, its quadratic variation is given by (6.3). This process is called the stochastic integral or Itô integral of X w.r.t. M.*

Proof First we show existence. Let $X \in \mathcal{P}$. From Proposition 5.13 and Remark 5.14 we know that there is a sequence of $X^n \in \mathcal{S}$ such that $||X - X^n|| \to 0$. By Proposition 6.2 we have for each t that $\mathbb{E}\,((X^m \cdot M)_t - (X^n \cdot M)_t)^2 = \mathbb{E}\int_0^t (X_s^m - X_s^n)^2 \,\mathrm{d}\langle M\rangle_s$ and hence $||X^m \cdot M - X^n \cdot M|| = ||X^m - X^n||_M$. This shows that the sequence $(X^n \cdot M)$ is Cauchy in the complete space $\mathcal{M}_c^2$ (Proposition 3.4) and thus has a limit in this space. We call it $X \cdot M$. The limit can be seen to be independent of the particular sequence (X^n) by the following familiar trick. Let (Y^n) be another sequence in $\mathcal{S}$ converging to X. Mix the two sequences as follows: $X^1, Y^1, X^2, Y^2, \ldots$. Also this sequence converges to X. Consider the sequence of corresponding stochastic integrals $X^1 \cdot M, Y^1 \cdot M, X^2 \cdot M, Y^2 \cdot M, \ldots$. This sequence has a unique limit in $\mathcal{M}_c^2$ and hence its subsequences $(X^n \cdot M)$ and $(Y^n \cdot M)$ must converge to the same limit, which then must be $X \cdot M$. The proof of (6.3) is left as Exercise 6.6. □

We will frequently need the following extension of Proposition 6.2.

Lemma 6.4 *The mapping $I : \mathcal{P} \to \mathcal{M}_c^2$, $I(X) = X \cdot M$, is linear and $I_T : (\mathcal{P}_T, ||\cdot||_M) \to L^2(\Omega, \mathcal{F}_T, \mathbb{P})$, $I_T(X) = (X \cdot M)_T$, an isometry from $(\mathcal{P}_T, ||\cdot||_M)$ onto its image.*

Proof Exercise 6.7. □

The proposition below gives two alternative ways to express a stopped stochastic integral.

Proposition 6.5 *Let $X \in \mathcal{P}$, $M \in \mathcal{M}_c^2$ and T a stopping time. Then*

$$\begin{aligned}(X \cdot M)^T &= X \cdot M^T && (6.7)\\ &= \mathbf{1}_{[0,T]} X \cdot M. && (6.8)\end{aligned}$$

Proof First we prove (6.7). If $X \in \mathcal{S}$, the result is trivial. For $X \in \mathcal{P}$, we take $X^n \in \mathcal{S}$ such that $||X^n - X||_M \to 0$. Using the assertion for processes in $\mathcal{S}$ (check this) and the linearity property of Lemma 6.4 we can write

$$(X \cdot M)^T - X \cdot M^T = ((X - X^n) \cdot M)^T - (X - X^n) \cdot M^T.$$

For any $t > 0$ we have

$$\begin{aligned}\mathbb{E}\left(((X - X^n) \cdot M)_t^T\right)^2 &= \mathbb{E}\int_0^{t\wedge T} (X_u - X_u^n)^2 \,\mathrm{d}\langle M\rangle_u \\ &\le \mathbb{E}\int_0^t (X_u - X_u^n)^2 \,\mathrm{d}\langle M\rangle_u \to 0.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}\left((X - X^n) \cdot M^T\right)_t^2 &= \mathbb{E}\int_0^t (X_u - X_u^n)^2 \,\mathrm{d}\langle M^T\rangle_u \\ &= \mathbb{E}\int_0^t (X_u - X_u^n)^2 \,\mathrm{d}\langle M\rangle_u^T \\ &= \mathbb{E}\int_0^{t\wedge T} (X_u - X_u^n)^2 \,\mathrm{d}\langle M\rangle_u \\ &\le \mathbb{E}\int_0^t (X_u - X_u^n)^2 \,\mathrm{d}\langle M\rangle_u \to 0,\end{aligned}$$

This shows (6.7).

To prove (6.8), we first show that $\mathbf{1}_{[0,T]}X \cdot M = (\mathbf{1}_{[0,T]}X \cdot M)^T$. From (6.7) we have $(\mathbf{1}_{[0,T]}X \cdot M)^T = \mathbf{1}_{[0,T]}X \cdot M^T$. Hence, exploiting the linearity of the stochastic integral in the integrator, Exercise 6.12, we obtain

$$\begin{aligned}\mathbf{1}_{[0,T]}X \cdot M - (\mathbf{1}_{[0,T]}X \cdot M)^T &= \mathbf{1}_{[0,T]}X \cdot M - \mathbf{1}_{[0,T]}X \cdot M^T \\ &= \mathbf{1}_{[0,T]}X \cdot (M - M^T),\end{aligned}$$

which has quadratic variation $\int \mathbf{1}_{[0,T]}X^2 \,\mathrm{d}\langle M - M^T\rangle = \int \mathbf{1}_{[0,T]}X^2 \,\mathrm{d}(\langle M\rangle - \langle M\rangle^T) = 0$. We can thus write, using (6.7) again,

$$\begin{aligned}(X \cdot M)^T - (\mathbf{1}_{[0,T]}X \cdot M) &= (X \cdot M)^T - (\mathbf{1}_{[0,T]}X \cdot M)^T \\ &= X \cdot M^T - \mathbf{1}_{[0,T]}X \cdot M^T \\ &= (1 - \mathbf{1}_{[0,T]})X \cdot M^T \\ &= \mathbf{1}_{(T,\infty)}X \cdot M^T,\end{aligned}$$

whose quadratic variation $\int \mathbf{1}_{(T,\infty)}X^2 \,\mathrm{d}\langle M\rangle^T$ vanishes, which proves (6.8). □

Proposition 6.6 *Let $X, Y \in \mathcal{P}$, $M \in \mathcal{M}_c^2$ and $S \le T$ be stopping times. Then*

$$\mathbb{E}\left[(X \cdot M)_{T\wedge t}|\mathcal{F}_S\right] = (X \cdot M)_{S\wedge t}$$

and

$$\mathbb{E}\,[((X\cdot M)_{T\wedge t}-(X\cdot M)_{S\wedge t})((Y\cdot M)_{T\wedge t}-(Y\cdot M)_{S\wedge t})|\mathcal{F}_S] =$$
$$\mathbb{E}\,[\int_{S\wedge t}^{T\wedge t} X_uY_u\,\mathrm{d}\langle M\rangle_u|\mathcal{F}_S].$$

Proof The first property follows from Corollary 2.9. The second property is first proved for $X=Y$ by applying Corollary 2.9 to the martingale $(X\cdot M)^2-\int_0^{\cdot} X^2\,\mathrm{d}\langle M\rangle$ and then by polarization. □

Remark 6.7 Integrals of the type $(X\cdot M)_\infty=\int_0^\infty X\,\mathrm{d}M$ can be defined similarly for $X\in\mathcal{P}_\infty$. The Itô isometry property then reads $\mathbb{E}\,(X\cdot M)_\infty^2=\mathbb{E}\int_0^\infty X^2\,\mathrm{d}\langle M\rangle$, which is finite by $X\in\mathcal{P}_\infty$. Furthermore, one can show the relation $(X\cdot M)_t=(X\mathbf{1}_{[0,t]}\cdot M)_\infty$. Also properties in the next section carry over from finite t to $t=\infty$ under the appropriate assumptions.

6.2 Characterizations and further properties

One of the aims of this section is the computation of the quadratic covariation between the martingales $X\cdot M$ and $Y\cdot N$, where $X\in\mathcal{P}(M)$, $Y\in\mathcal{P}(N)$ and $M,N\in\mathcal{M}_c^2$. For $X,Y\in\mathcal{S}$ this is (relatively) straightforward (Exercise 6.5) since the integrals become sums and the result is

$$\langle X\cdot M, Y\cdot N\rangle=\int_0^{\cdot} XY\,\mathrm{d}\langle M,N\rangle. \tag{6.9}$$

The extension to more general X and Y will be established in a number of steps. The first step is a result known as the *Kunita-Watanabe inequality.*

Proposition 6.8 *Let $M,N\in\mathcal{M}_c^2$, $X\in\mathcal{P}(M)$ and $Y\in\mathcal{P}(N)$. Let V be the total variation process of the process $\langle M,N\rangle$. Then for $t\geq 0$,*

$$|\int_0^t XY\,\mathrm{d}\langle M,N\rangle|\leq\big(\int_0^t |X|^2\,\mathrm{d}\langle M\rangle\big)^{1/2}\big(\int_0^t |Y|^2\,\mathrm{d}\langle N\rangle\big)^{1/2}\text{ a.s.,} \tag{6.10}$$
$$\int_0^t |XY|\,\mathrm{d}V\leq\big(\int_0^t |X|^2\,\mathrm{d}\langle M\rangle\big)^{1/2}\big(\int_0^t |Y|^2\,\mathrm{d}\langle N\rangle\big)^{1/2}\text{ a.s.} \tag{6.11}$$

Proof We first focus on (6.10). Observe that $(\langle M,N\rangle_t-\langle M,N\rangle_s)^2\leq(\langle M\rangle_t-\langle M\rangle_s)(\langle N\rangle_t-\langle N\rangle_s)$ a.s. (similar to Exercise 3.3). Assume that X and Y are simple. The final result then follows by taking limits. The process X we can represent on $(0,t]$ as $\sum_k x_k\mathbf{1}_{(t_k,t_{k+1}]}$ with the last $t_n=t$. For Y we similarly have $\sum_k y_k \mathbf{1}_{(t_k,t_{k+1}]}$. It follows that

$$
\begin{aligned}
&|\int_0^t XY \,\mathrm{d}\langle M,N\rangle| \\
&\leq \sum |x_k||y_k||\langle M,N\rangle_{t_{k+1}} - \langle M,N\rangle_{t_k}| \\
&\leq \sum_k |x_k||y_k|\big((\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k})(\langle N\rangle_{t_{k+1}} - \langle N\rangle_{t_k})\big)^{1/2} \\
&\leq \big(\sum_k x_k^2(\langle M\rangle_{t_{k+1}} - \langle M\rangle_{t_k})\big)^{1/2}\big(\sum_k y_k^2(\langle N\rangle_{t_{k+1}} - \langle N\rangle_{t_k})\big)^{1/2} \\
&= \big(\int_0^t |X|^2 \,\mathrm{d}\langle M\rangle\big)^{1/2}\big(\int_0^t |Y|^2 \,\mathrm{d}\langle N\rangle\big)^{1/2}.
\end{aligned}
$$

The remainder of the proof of (6.10) is left as Exercise 6.13.

Consider then (6.11). Observe that $\langle M,N\rangle$ is absolutely continuous w.r.t. V, with density process $\mathrm{d}\langle M,N\rangle/\mathrm{d}V$ taking values in $\{-1,1\}$. Write $\int_0^t |XY|\,\mathrm{d}V$ as $\int_0^t XY\mathrm{sgn}(XY)\frac{\mathrm{d}\langle M,N\rangle}{\mathrm{d}V}\,\mathrm{d}\langle M,N\rangle$. Relabeling $X\mathrm{sgn}(XY)$ as X and $Y\frac{\mathrm{d}\langle M,N\rangle}{\mathrm{d}V}$ as Y, we simply apply (6.10). □

Lemma 6.9 *Let $M,N \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M)$. Then*

$$\langle X\cdot M, N\rangle = \int_0^\cdot X \,\mathrm{d}\langle M,N\rangle$$

or, in shorthand notation,

$$\mathrm{d}\,\langle X\cdot M, N\rangle = X\,\mathrm{d}\langle M,N\rangle.$$

Proof Choose $X^n \in \mathcal{S}$ such that $||X^n - X||_M \to 0$. Then we can find for every $T > 0$ a subsequence, again denoted by X^n, such that $\int_0^T (X^n - X)^2 \,\mathrm{d}\langle M\rangle \to 0$ a.s. But then $\langle X^n\cdot M - X\cdot M, N\rangle_T^2 \leq \langle X^n\cdot M - X\cdot M\rangle_T\langle N\rangle_T \to 0$ a.s. But for the simple X^n one relatively easily (verify this!) verifies $\langle X^n\cdot M, N\rangle_T = \int_0^T X^n\,\mathrm{d}\langle M,N\rangle$ (similar to Exercise 6.5). Application of the Kunita-Watanabe inequality to $|\int_0^T (X^n - X)\,\mathrm{d}\langle M,N\rangle|$ yields the result. □

Proposition 6.10 *Let $M,N \in \mathcal{M}_c^2$, $X \in \mathcal{P}(M)$ and $Y \in \mathcal{P}(N)$. Then Equation (6.9) holds.*

Proof We apply Lemma 6.9 twice and get

$$\langle X\cdot M, Y\cdot N\rangle = \int_0^\cdot Y\,\mathrm{d}\langle X\cdot M, N\rangle = \int_0^\cdot XY\,\mathrm{d}\langle M,N\rangle,$$

which is the desired equality. □

We are now in the position to state an important characterization of the stochastic integral. In certain books it is taken as a definition.

Theorem 6.11 *Let $M \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M)$. Let $I \in \mathcal{M}_c^2$ be such that $\langle I, N\rangle = \int_0^\cdot X \,\mathrm{d}\langle M, N\rangle$ for all $N \in \mathcal{M}_c^2$. Then I is indistinguishable from $X \cdot M$.*

Proof Since $\langle X \cdot M, N\rangle = \int_0^\cdot X \,\mathrm{d}\langle M, N\rangle$ we get $\langle I - X \cdot M, N\rangle = 0$ for all $N \in \mathcal{M}_c^2$ by subtraction. In particular for $N = I - X \cdot M$. The result follows, since $\langle N\rangle = 0$ implies $N = 0$. □

The characterization is a useful tool in the proof of the following 'chain rule'.

Proposition 6.12 *Let $M \in \mathcal{M}_c^2$, $X \in \mathcal{P}(M)$ and $Y \in \mathcal{P}(X \cdot M)$. Then $XY \in \mathcal{P}(M)$ and $Y \cdot (X \cdot M) = (XY) \cdot M$ up to indistinguishability.*

Proof Since $\langle X \cdot M\rangle = \int_0^\cdot X^2 \,\mathrm{d}\langle M\rangle$, it immediately follows that $XY \in \mathcal{P}(M)$. Furthermore, for any martingale $N \in \mathcal{M}_c^2$ we have

$$\begin{aligned}
\langle (XY) \cdot M, N\rangle &= \int_0^\cdot XY \,\mathrm{d}\langle M, N\rangle \\
&= \int_0^\cdot Y \,\mathrm{d}\langle X \cdot M, N\rangle \\
&= \langle Y \cdot (X \cdot M), N\rangle.
\end{aligned}$$

It follows from Theorem 6.11 that $Y \cdot (X \cdot M) = (XY) \cdot M$. □

The construction of the stochastic integral that we have developed here is founded on 'L^2-theory'. We have not defined the stochastic integral w.r.t. a continuous martingale in a pathwise way. Nevertheless, there exists a 'pathwise-uniqueness result'.

Proposition 6.13 *Let $M_1, M_2 \in \mathcal{M}_c^2$, $X_1 \in \mathcal{P}(M_1)$ and $X_2 \in \mathcal{P}(M_2)$. Let T be a stopping time and suppose that M_1^T and M_2^T as well as X_1^T and X_2^T are indistinguishable. Then the same holds for $(X_1 \cdot M_1)^T$ and $(X_2 \cdot M_2)^T$.*

Proof For any $N \in \mathcal{M}_c^2$ we have that $\langle M_1 - M_2, N\rangle^T = 0$. Hence,

$$\begin{aligned}
\langle (X_1 \cdot M_1)^T - (X_2 \cdot M_2)^T, N\rangle &= \int_0^\cdot X_1 \,\mathrm{d}\langle M_1^T, N\rangle - \int_0^\cdot X_2 \,\mathrm{d}\langle M_2^T, N\rangle \\
&= \int_0^\cdot (X_1 - X_2) \,\mathrm{d}\langle M_1^T, N\rangle \\
&= \int_0^\cdot (X_1 - X_2) \,\mathrm{d}\langle M_1, N\rangle^T \\
&= \int_0^\cdot (X_1 - X_2)\mathbf{1}_{[0,T]} \,\mathrm{d}\langle M_1, N\rangle \\
&= 0.
\end{aligned}$$

The assertion follows by application of Theorem 6.11. □

6.3 Integration w.r.t. local martingales

In this section we extend the definition of stochastic integral into two directions. In the first place we relax the condition that the integrator is a martingale (in $\mathcal{M}_c^2$) and in the second place, we put less severe restrictions on the integrand.

In this section M will be a continuous local martingale. We have

Definition 6.14 For $M \in \mathcal{M}_c^{loc}$ the class $\mathcal{P}^* = \mathcal{P}^*(M)$ is defined as the collection of progressive processes X with the property that $\int_0^T X^2 \,\mathrm{d}\langle M\rangle < \infty$ a.s. for all $T \geq 0$.

Recall that for local martingales the quadratic (co-)variation processes exist.

Theorem 6.15 *Let $M \in \mathcal{M}_c^{loc}$ and $X \in \mathcal{P}^*(M)$. Then there exists a unique local martingale, denoted by $X \cdot M$, such that for all $N \in \mathcal{M}_c^{loc}$ it holds that*

$$\langle X \cdot M, N\rangle = \int_0^{\cdot} X \,\mathrm{d}\langle M, N\rangle. \tag{6.12}$$

This local martingale is called the stochastic integral of X w.r.t. M. If furthermore $Y \in \mathcal{P}^*(N)$, then equality (6.9) is still valid.

Proof Define the stopping times S^n as a localizing sequence for M and $T^n = \inf\{t \geq 0 : M_t^2 + \int_0^t X^2 \,\mathrm{d}\langle M\rangle \geq n\} \wedge S^n$. Then the T^n also form a localizing sequence, $|M^{T^n}| \leq n$ and $X^{T^n} \in \mathcal{P}(M^{T^n})$. Therefore the stochastic integrals $I^n := X^{T^n} \cdot M^{T^n}$ can be defined as before. It follows from e.g. Proposition 6.13 that I^{n+1} and I^n coincide on $[0, T^n]$. Hence we can unambiguously define $(X \cdot M)_t$ as I_t^n for any n such that $T^n \geq t$. Since $(X \cdot M)^{T^n} = I^n$ is a martingale, $X \cdot M$ is a local martingale. Furthermore, for any $N \in \mathcal{M}^2$ we have $\langle X \cdot M, N\rangle^{T^n} = \langle I^n, N\rangle = \int_0^{\cdot} X \,\mathrm{d}\langle M, N\rangle^{T^n}$. By letting $n \to \infty$ we obtain (6.12). The uniqueness follows as in the proof of Theorem 6.11 for $N \in \mathcal{M}^2$, and then also for $N \in \mathcal{M}_c^{loc}$. $\square$

Remark 6.16 Let $M \in \mathcal{M}_c^{loc}$ and $\mathcal{P}_\infty^*$ be the class of progressive processes X satisfying $\int_0^\infty X^2 \,\mathrm{d}\langle M\rangle < \infty$ a.s. Then one can also define parallel to Remark 6.7 integrals $(X \cdot M)_\infty$ by adapting the proof of Theorem 6.15.

6.4 Exercises

6.1 Let (X^n) be a sequence in $\mathcal{P}$ that converges to X w.r.t. the metric of Definition 5.1. Show that the stochastic integrals $X^n \cdot M$ converge to $X \cdot M$ w.r.t. the metric of Definition 3.2 and also that $\sup_{s \leq t} |(X^n \cdot M)_s - (X \cdot M)_s| \xrightarrow{\mathbb{P}} 0$. (The latter convergence is called uniform convergence on compacts in probability, abbreviated by ucp convergence, often denoted $\xrightarrow{\text{ucp}}$).

6.2 Show, not referring to Proposition 6.5, that $(\mathbf{1}_{[0,T]} X) \cdot M = (X \cdot M)^T$ for any finite stopping time T, $M \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M)$.

6.3 Let S and T be stopping times such that $S \leq T$ and let ζ be a bounded $\mathcal{F}_S$-measurable random variable. Show that the process $X = \zeta \mathbf{1}_{(S,T]}$ is progressive. Let $M \in \mathcal{M}_c^2$. Show that $((\zeta \mathbf{1}_{(S,T]}) \cdot M)_t = \zeta(M_{T\wedge t} - M_{S\wedge t})$.

6.4 Prove the second assertion of Proposition 6.6.

6.5 Show the equality (6.9) for X and Y in $\mathcal{S}$.

6.6 Finish the proof of Theorem 6.3. I.e. show that the quadratic variation of $X \cdot M$ is given by (6.3) if $M \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M)$.

6.7 Prove the linearity in Lemma 6.4, so show that $(X+Y)\cdot M$ and $X\cdot M+Y\cdot M$ are indistinguishable if $M \in \mathcal{M}_c^2$ and $X, Y \in \mathcal{P}(M)$.

6.8 Let W be standard Brownian motion.

(a) Find a sequence of piecewise constant processes W^n such that $\mathbb{E}\int_0^T |W_t^n - W_t|^2 \,\mathrm{d}t \to 0$.

(b) Compute $\int_0^T W_t^n \,\mathrm{d}W_t$ and show that it 'converges' (in what sense?) to $\frac{1}{2}(W_T^2 - T)$, if we consider smaller and smaller intervals of constancy.

(c) Deduce that $\int_0^T W_t \,\mathrm{d}W_t = \frac{1}{2}(W_T^2 - T)$.

6.9 Let $M \in \mathcal{M}_c^{loc}$, $X, Y \in \mathcal{P}^*(M)$ and $a, b \in \mathbb{R}$. Show that $(aX + bY)\cdot M$ and $a\,X \cdot M + b\,Y \cdot M$ are indistinguishable.

6.10 Let W be Brownian motion and T a stopping time with $\mathbb{E}\,T < \infty$. Show (use stochastic integrals) that $\mathbb{E}\,W_T = 0$ and $\mathbb{E}\,W_T^2 = \mathbb{E}\,T$.

6.11 Define for $M \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M)$ for $0 \leq s < t \leq T$ the random variable $\int_s^t X \,\mathrm{d}M$ as $(X \cdot M)_t - (X \cdot M)_s$. Show that $\int_s^t X \,\mathrm{d}M = ((\mathbf{1}_{(s,t]}X)\cdot M)_T$.

6.12 Let $M, N \in \mathcal{M}_c^2$ and $X \in \mathcal{P}(M) \cap \mathcal{P}(N)$. Show that $X \cdot M + X \cdot N$ and $X \cdot (M + N)$ are indistinguishable.

6.13 Finish the proof of Proposition 6.8 as follows. Show that we can deduce from the given proof that inequality (6.10) holds for all *bounded* processes $X \in \mathcal{P}(M)$ and $Y \in \mathcal{P}(N)$ and then for all $X \in \mathcal{P}(M)$ and $Y \in \mathcal{P}(N)$.

6.14 Let C be a finite $\mathcal{F}_0$-measurable random variable and suppose that the process X belongs to $\mathcal{P}^*$. Let $M \in \mathcal{M}_c^{loc}$. Show that the stochastic integral process $(CX) \cdot M$ is well defined and that $(CX) \cdot M = C(X \cdot M)$.

7 Semimartingales and the Itô formula

As almost always we assume also in this section that the filtration $\mathbb{F}$ on the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the usual conditions.

7.1 Semimartingales

Definition 7.1 A process X is called a (continuous) *semimartingale* if it admits a decomposition

$$X = X_0 + A + M, \tag{7.1}$$

where $M \in \mathcal{M}_c^{loc}$ and A is a continuous process with paths of bounded variation (over finite intervals) and $A_0 = 0$ a.s.

Remark 7.2 The decomposition (7.1) is unique up to indistinguishability. This follows from the fact that continuous local martingales that are of bounded variation are a.s. constant in time (Lemma 2.15). We call (7.1) the semimartingale decomposition of X.

Every continuous submartingale is a semimartingale and its semimartingale decomposition of Definition 7.1 coincides with the Doob-Meyer decomposition.

Definition 7.3 If X and Y are semimartingales with semimartingale decompositions $X = X_0 + A + M$ and $Y = Y_0 + B + N$ where M and N are local martingales and A and B processes of bounded variation, then we define their quadratic covariation process $\langle X, Y \rangle$ as $\langle M, N \rangle$. If $X = Y$, we write $\langle X \rangle$ for $\langle X, X \rangle$, the quadratic variation process of X.

The bounded variation parts of semimartingales is thus ignored in the definition of quadratic variation and covariation. Proposition 7.4 below gives an intuitively appealing justification for this.

Let X be a semimartingale and B a continuous bounded variation process. Then $X + B$ is again a semimartingale, which has the same (unique) local martingale part as X. It follows that with every other semimartingale Y one has $\langle X + B, Y \rangle = \langle X, Y \rangle$. We will often use this property.

Proposition 7.4 *The given definition of $\langle X, Y \rangle$ for semimartingales coincides with our intuitive understanding of quadratic covariation. If (Π^n) is a sequence of partitions of $[0, \infty)$ whose meshes $\mu(\Pi^n)$ tend to zero, then for every $T > 0$ we have*

$$\sup_{t \leq T} |V_t(X, Y; \Pi^n) - \langle X, Y \rangle_t| \xrightarrow{\mathbb{P}} 0,$$

where for any partition Π we write $V_t(X, Y; \Pi) = \sum (X_{t_{k+1}} - X_{t_k})(Y_{t_{k+1}} - Y_{t_k})$ with the summation over the $t_k, t_{k+1} \in \Pi \cap [0, t]$.

Proof It is sufficient to prove this for $X = Y$ with $X_0 = 0$. Let $X = X_0 + A + M$ according to Equation (7.1). Write $V_t(X, X; \Pi^n) = V_t(M, M; \Pi^n) + 2V_t(A, M; \Pi^n) + V_t(A, A; \Pi^n)$. Since A has paths of bounded variation and A and M have continuous paths, the last two terms tend to zero a.s. We concentrate henceforth on $V_t(M, M; \Pi^n)$. Let (T^m) be a localizing sequence for M such that M^{T^m} as well as $\langle M^{T^m} \rangle$ are bounded by m. For given $\varepsilon, \delta, T > 0$, we can choose T^m such that $\mathbb{P}(T^m \leq T) < \delta$. We thus have

$$\begin{aligned}
&\mathbb{P}(\sup_{t \leq T} |V_t(M, M; \Pi^n) - \langle M \rangle_t| > \varepsilon) \leq \\
&\quad \delta + \mathbb{P}(\sup_{t \leq T} |V_t(M^{T^m}, M^{T^m}; \Pi^n) - \langle M^{T^m} \rangle_t| > \varepsilon, T < T^m). \qquad (7.2)
\end{aligned}$$

Realizing that $V(M^{T^m}, M^{T^m}; \Pi^n) - \langle M^{T^m} \rangle$ is a bounded martingale, we apply Doob's inequality to show that the second term in (7.2) tends to zero when the mesh $\mu(\Pi^n)$ tends to zero. Actually, similar to the proof of Proposition 3.8, one shows L^2-convergence (Exercise 3.13). Then let $m \to \infty$. (We thus obtain by almost the same techniques an improvement of Proposition 3.8). □

Definition 7.5 A progressive process Y is called *locally bounded* if there exists a localizing sequence (T^n) such that the stopped processes Y^{T^n} are bounded.

Clearly, all continuous processes are locally bounded and locally bounded processes belong to $\mathcal{P}^*(M)$ for any continuous local martingale M. Moreover, for locally bounded processes Y and continuous processes A that are of bounded variation, the pathwise Lebesgue-Stieltjes integrals $\int_0^t Y_s(\omega)\, \mathrm{d}A_s(\omega)$ are (a.s.) defined for every $t > 0$ and finite a.s.

Definition 7.6 Let Y be a locally bounded process and X a semimartingale with decomposition (7.1). Then the stochastic integral of Y w.r.t. X is defined as $Y \cdot M + \int_0^\cdot Y\, \mathrm{d}A$. Here the first integral is the stochastic integral of Theorem 6.15 and the second one is the Stieltjes integral that we just mentioned. From now on we will use the notation $\int_0^\cdot Y\, \mathrm{d}X$ or $Y \cdot X$ to denote the stochastic integral of Y w.r.t. X.

Note that, due to the uniqueness of the decomposition in Definition 7.1, the definition of $Y \cdot X$ is unambiguous. Moreover, $Y \cdot X$ is again a semimartingale and

$$Y \cdot X = Y \cdot A + Y \cdot M,$$

is its unique semimartingale decomposition. Indeed, $Y \cdot M$ is a local martingale and $Y \cdot A$ has paths of bounded variation. As a matter of fact, its total variation process $V(Y \cdot M)$ is given by $V(Y \cdot M)_t = (|Y| \cdot V(A))_t$, which is a.s. finite, where $V(A)$ stands for the total variation of A.

As a consequence, for two semimartingales X_1 and X_2 and two locally bounded processes Y_1 en Y_2 we have $\langle Y_1 \cdot X_1, Y_2 \cdot X_2 \rangle = (Y_1 Y_2) \cdot \langle M_1, M_2 \rangle$, where the M_i are the local martingale parts of the X_i, and then also $\langle Y_1 \cdot X_1, Y_2 \cdot X_2 \rangle = (Y_1 Y_2) \cdot \langle X_1, X_2 \rangle$. Hence we have the analogue of (6.9) for stochastic integrals w.r.t. semimartingales.

Proposition 7.7 *Let Y, U be locally bounded processes and X a semimartingale.*

(i) If T is a stopping time, then $(Y \cdot X)^T = Y \cdot X^T = (\mathbf{1}_{[0,T]}Y) \cdot X$.

(ii) The product YU is locally bounded, $Y \cdot X$ is a semimartingale and $U \cdot (Y \cdot X)$ and $(YU) \cdot X$ are indistinguishable.

Proof As the assertions are obvious, if X itself is of bounded variation, we only have to take care of the (local) martingale parts. But for these we have Propositions 6.5 and 6.12, that we combine with an appropriate stopping argument. $\square$

Proposition 7.8 *Let X be a continuous semimartingale and (Y^n) a sequence of locally bounded progressive processes that converge to zero pointwise. If there exists a locally bounded progressive process Y that dominates all the Y^n, then $Y^n \cdot X$ converges to zero uniformly in probability on compact sets (also called ucp convergence, notation: $Y^n \cdot X \stackrel{\text{ucp}}{\to} 0$), meaning that $\sup_{s \leq t} |\int_0^s Y^n \,\mathrm{d}X|$ tends to zero in probability for all $t \geq 0$.*

Proof Let $X = A + M$. Then $\sup_{s \leq t} |\int_0^s Y^n \,\mathrm{d}A|$ tends to zero pointwise in view of Lebesgue's dominated convergence theorem. Therefore we prove the proposition for $X = M$, a continuous local martingale. Choose stopping times T^m such that each M^{T^m} is a square integrable martingale and Y^{T^m} bounded. Then for all $T > 0$, $\mathbb{E} \int_0^T (Y^n)^2 \,\mathrm{d}\langle M^{T^m} \rangle \to 0$ for $n \to \infty$ and by the isometry property of the stochastic integral, valid since $Y^n \in \mathcal{P}(M^{T^m})$, we also have for all $T > 0$ that $\mathbb{E}\,(Y^n \cdot M^{T^m})_T^2 \to 0$ and consequently, see the proof of Proposition 7.4, we have $Y^n \cdot M^{T^m} \stackrel{\text{ucp}}{\to} 0$ for $n \to \infty$. The dependence on T^m can be removed in the same way as in the proof of Proposition 7.4. $\square$

We close this section with a result that says that $Y \cdot X$ is in a good sense a limit of sums of the type (6.2), which was the point of departure of developing the theory of stochastic integrals.

Corollary 7.9 *Let Y be a continuous adapted process and X a semimartingale. Let Π^n be a sequence of partitions of $[0, \infty)$ whose meshes tend to zero. Let $Y^n = \sum_k Y_{t_k^n} \mathbf{1}_{(t_k^n, t_{k+1}^n]}$, with the t_k^n in Π^n. Then $Y^n \cdot X \stackrel{\text{ucp}}{\to} Y \cdot X$.*

Proof Since Y is locally bounded, we can find stopping times T^m such that $|Y^{T^m}| \leq m$ and hence $\sup_t |(Y^n)_t^{T^m}| \leq m$. We can therefore apply the preceding proposition to the sequence $((Y^n)^{T^m})_{n \geq 1}$ that converges pointwise to Y^{T^m} and with X replaced with X^{T^m}. We thus obtain $Y^n \cdot (X^{T^m}) \stackrel{\text{ucp}}{\to} Y \cdot (X^{T^m})$, which is nothing else but $\sup_{t \leq T} |\int_0^{t \wedge T^m} Y^n \,\mathrm{d}X - \int_0^{t \wedge T^m} Y \,\mathrm{d}X| \stackrel{\mathbb{P}}{\to} 0$ for all $T > 0$. Finally, since for each t the probability $\mathbb{P}(t \leq T^m) \to 1$, we can remove the stopping time in the last ucp convergence. $\square$

7.2 Integration by parts

The following (first) stochastic calculus rule is the foundation for the Itô formula of the next section.

Proposition 7.10 *Let X and Y be (continuous) semimartingales. Then*

$$X_tY_t = X_0Y_0 + \int_0^t X\,\mathrm{d}Y + \int_0^t Y\,\mathrm{d}X + \langle X, Y\rangle_t \quad \text{a.s.}\,(t \geq 0). \tag{7.3}$$

A special case occurs when $Y = X$ in which case (7.3) becomes

$$X_t^2 = X_0^2 + 2\int_0^t X\,\mathrm{d}X + \langle X\rangle_t \quad \text{a.s.}\,(t \geq 0). \tag{7.4}$$

Proof It is sufficient to prove (7.4), because then (7.3) follows by polarization. Let then Π be a subdivision of $[0, t]$. Then, summing over the elements of the subdivision, we have

$$X_t^2 - X_0^2 = 2\sum X_{t_k}(X_{t_{k+1}} - X_{t_k}) + \sum (X_{t_{k+1}} - X_{t_k})^2 \,\text{a.s.}\,(t \geq 0).$$

To the first term on the right we apply Corollary 7.9 and for the second term we use Proposition 7.4. This yields the assertion. □

If the semimartingales X and Y don't possess a local martingale part, then (7.3) reduces to the familiar integration by parts formula

$$X_tY_t = X_0Y_0 + \int_0^t X\,\mathrm{d}Y + \int_0^t Y\,\mathrm{d}X \quad \text{a.s.}\,(t \geq 0).$$

Another consequence of Proposition 7.10 is that we can use Equation (7.3) to *define* the quadratic covariation between two semimartingales. Indeed, some authors take this as their point of view.

7.3 Itô's formula

Theorem 7.11 below contains the celebrated Itô formula (7.5), perhaps the most famous and a certainly not to be underestimated result in stochastic analysis.

Theorem 7.11 *Let X be a continuous semimartingale and f a twice continuously differentiable function on $\mathbb{R}$. Then $f(X)$ is a continuous semimartingale as well and it holds that*

$$f(X_t) = f(X_0) + \int_0^t f'(X)\,\mathrm{d}X + \frac{1}{2}\int_0^t f''(X)\,\mathrm{d}\langle X\rangle, \text{a.s.}\,(t \geq 0). \tag{7.5}$$

Before giving the proof of Theorem 7.11, we comment on the integrals in Equation (7.5). The first integral we have to understand as a sum as in Definition 7.6. With M the local martingale part of X we therefore have to consider the integral $f'(X) \cdot M$, which is well defined since $f'(X)$ is continuous and thus locally bounded. Moreover it is a continuous local martingale. With A the finite variation part of X, the integral $\int_0^{\cdot} f'(X)\,\mathrm{d}A$ has to be understood in the (pathwise) Lebesgue-Stieltjes sense and thus it becomes a process of finite variation, as is the case for the integral $\int_0^{\cdot} f''(X)\,\mathrm{d}\langle X\rangle$. Hence, $f(X)$ is a continuous semimartingale and its local martingale part is $f'(X) \cdot M$.

Proof The theorem is obviously true for affine functions and for f given by $f(x) = x^2$, Equation (7.5) reduces to (7.4). We show by induction that (7.5) is true for any monomial and hence, by linearity, for every polynomial. The general case follows at the end.
Let $f(x) = x^n = x^{n-1}x$. We apply the integration by parts formula (7.3) with $Y_t = X_t^{n-1}$ and assume that (7.5) is true for $f(x) = x^{n-1}$. We obtain

$$X_t^n = X_0^n + \int_0^t X^{n-1}\,\mathrm{d}X + \int_0^t X\,\mathrm{d}X^{n-1} + \langle X, X^{n-1}\rangle_t. \tag{7.6}$$

By assumption we have

$$X_t^{n-1} = X_0^{n-1} + \int_0^t (n-1)X^{n-2}\,\mathrm{d}X + \frac{1}{2}\int_0^t (n-1)(n-2)X^{n-3}\,\mathrm{d}\langle X\rangle. \tag{7.7}$$

We obtain from this equation that $\langle X^{n-1}, X\rangle$ is given by $(n-1)\int_0^{\cdot} X^{n-2}\,\mathrm{d}\langle X\rangle$ (remember that the quadratic covariation between two semimartingales is determined by their local martingale parts). Inserting this result as well as (7.7) into (7.6) and using Proposition 7.7 we get the result for $f(x) = x^n$ and hence for f equal to an arbitrary polynomial.
Suppose now that f is twice continuously differentiable and that X is bounded, with values in $[-K, K]$, say. Then, since f'' is continuous, we can (by the Weierstraß approximation theorem) view it on $[-K, K]$ as the uniform limit of a sequence of polynomials, p_n'' say: for all $\varepsilon > 0$ there is n_0 such that $\sup_{[-K,K]} |f''(x) - p_n''(x)| < \varepsilon$, if $n > n_0$. But then f' is the uniform limit of p_n' defined by $p_n'(x) = f'(-K) + \int_{-K}^x p_n''(u)\,\mathrm{d}u$ and f as the uniform limit of the polynomials p_n defined by $p_n(x) = f(-K) + \int_{-K}^x p_n'(u)\,\mathrm{d}u$. For the polynomials p_n we already know that (7.5) holds true. Write R for the difference of the left hand side of (7.5) minus its right hand side. Then

$$\begin{aligned} R &= f(X_t) - p_n(X_t) - (f(X_0) - p_n(X_0)) \\ &\quad - \int_0^t (f'(X) - p_n'(X))\,\mathrm{d}X - \frac{1}{2}\int_0^t (f''(X) - p_n''(X))\,\mathrm{d}\langle X\rangle. \end{aligned}$$

The first two differences in this equation can be made arbitrarily small by the definition of the p_n. To the first (stochastic) integral we apply Proposition 7.8 and the last integral has absolute value less than $\varepsilon\langle X\rangle_t$ if $n > n_0$.

Let now X be arbitrary and let $T^K = \inf\{t \geq 0 : |X_t| > K\}$. Then, certainly X^{T^K} is bounded by K and we can apply the result of the previous step. We have

$$f(X_t^{T^K}) = f(X_0) + \int_0^t f'(X^{T^K})\,\mathrm{d}X^{T^K} + \frac{1}{2}\int_0^t f''(X^{T^K})\,\mathrm{d}\langle X\rangle^{T^K},$$

which can be rewritten as

$$f(X_{t\wedge T^K}) = f(X_0) + \int_0^{t\wedge T^K} f'(X)\,\mathrm{d}X + \frac{1}{2}\int_0^{t\wedge T^K} f''(X)\,\mathrm{d}\langle X\rangle.$$

We trivially have $f(X_{t\wedge T^K}) \to f(X_t)$ a.s. and the right hand side of the previous equation is on $\{t < T^K\}$ (whose probability tends to 1) equal to the right hand side of (7.5). The theorem has been proved. □

Formula (7.5) is often represented in *differential notation*, a short hand way of writing the formula down without integrals. We write

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle X\rangle_t,$$

or merely

$$\mathrm{d}f(X) = f'(X)\,\mathrm{d}X + \frac{1}{2}f''(X)\,\mathrm{d}\langle X\rangle.$$

If T is a finite stopping time, then one can replace t in Formula (7.5) with T. This substitution is then also valid for $T \wedge t$ for an arbitrary stopping time T.

Remark 7.12 With minor changes in the proof one can show that also the following multivariate extension of the Itô formula (7.5) holds true. If $X = (X^1, \ldots, X^d)$ is a d-dimensional vector of semimartingales and $f : \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable in all its arguments, then

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X)\,\mathrm{d}X^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X)\,\mathrm{d}\langle X^i, X^j\rangle. \tag{7.8}$$

Notice that in this expression we only need second order derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$, if the corresponding components X^i and X^j of X both have a non-vanishing martingale part. The integration by parts formula (7.3) is a special case of (7.8).

7.4 Applications of Itô's formula

Let X be a (continuous) semimartingale with $X_0 = 0$ and define the process Z by $Z_t = Z_0 \exp(X_t - \frac{1}{2}\langle X\rangle_t)$, where Z_0 is an $\mathcal{F}_0$-measurable random variable. Application of Itô's formula gives

$$Z_t = Z_0 + \int_0^t Z\,\mathrm{d}X. \tag{7.9}$$

This is a linear (stochastic) integral equation, that doesn't have the usual exponential solution as in ordinary calculus (unless $\langle X\rangle = 0$). The process Z with $Z_0 = 1$ is called the *Doléans exponential* of X and we have in this case the special notation

$$Z = \mathcal{E}(X). \tag{7.10}$$

Notice that if X is a local martingale, Z is a local martingale as well. Later on we will give conditions on X that ensure that Z becomes a martingale.

An application of Itô's formula we present in the proof of *Lévy's characterization* of Brownian motion, Proposition 7.13.

Proposition 7.13 *Let M be a continuous local martingale (w.r.t. the filtration $\mathbb{F}$) with $M_0 = 0$ and $\langle M\rangle_t \equiv t$. Then M is a Brownian motion (w.r.t. $\mathbb{F}$).*

Proof By splitting into real and imaginary part one can show that Itô's formula also holds for complex valued semimartingales (here, by definition both the real and the imaginary part are semimartingales). Let $u \in \mathbb{R}$ be arbitrary and define the process Y by $Y_t = \exp(\mathrm{i}uM_t + \frac{1}{2}u^2 t)$. Applying Itô's formula, we obtain

$$Y_t = 1 + \mathrm{i}u \int_0^t Y_s \,\mathrm{d}M_s.$$

It follows that Y is a complex valued local martingale. We stop Y at the fixed time point t_0. Then, the stopped process Y^{t_0} is bounded and thus a martingale and since $Y \neq 0$ we get for all $s < t < t_0$

$$\mathbb{E}\,[\frac{Y_t}{Y_s}|\mathcal{F}_s] = 1.$$

This identity in explicit form is equal to

$$\mathbb{E}\,[\exp(\mathrm{i}u(M_t - M_s) + \frac{1}{2}u^2(t-s))|\mathcal{F}_s] = 1,$$

which is valid for all $t > s$, since t_0 is arbitrary. Rewriting this as

$$\mathbb{E}\,[\exp(\mathrm{i}u(M_t - M_s)|\mathcal{F}_s] = \exp(-\frac{1}{2}u^2(t-s)),$$

we conclude that $M_t - M_s$ is independent of $\mathcal{F}_s$ and has a normal distribution with zero mean and variance $t - s$. Since this is true for all $t > s$ we conclude that M is a Brownian motion w.r.t. $\mathbb{F}$. □

An important generalization of Doob's inequality (2.4) is one of the Burkholder-Davis-Gundy inequalities, Proposition 7.14 below. It is stated for $p \geq 2$ and continuous local martingales, although there are related results for $0 < p < 2$ and arbitrary local martingales.

Proposition 7.14 *Let $p \geq 2$. There exists a constant C_p such that for all continuous local martingales M with $M_0 = 0$ and all finite stopping times T one has*

$$\mathbb{E} \sup_{t \leq T} |M_t|^p \leq C_p \, \mathbb{E} \, \langle M \rangle_T^{p/2}. \tag{7.11}$$

Proof Assume first that M is bounded. The function $x \mapsto |x|^p$ is in $C^2(\mathbb{R})$ so we can apply Itô's formula to get

$$|M_T|^p = p \int_0^T \operatorname{sgn}(M_t)|M_t|^{p-1} \, \mathrm{d}M_t + \frac{1}{2}p(p-1) \int_0^T |M_t|^{p-2} \, \mathrm{d}\langle M \rangle_t.$$

Since M is bounded, it is an honest martingale and so is the first term in the above equation. Taking expectations, we thus get $M_T^* := \sup_{t \leq T} |M_t|$ and Hölder's inequality at the last step below

$$\begin{aligned}
\mathbb{E} \, |M_T|^p &= \frac{1}{2}p(p-1)\mathbb{E} \int_0^T |M_t|^{p-2} \, \mathrm{d}\langle M \rangle_t \\
&\leq \frac{1}{2}p(p-1)\mathbb{E} \left(\sup_{t \leq T} |M_t|^{p-2} \langle M \rangle_T \right) \\
&\leq \frac{1}{2}p(p-1)(\mathbb{E} \, (M_T^*)^p)^{1-2/p} (\mathbb{E} \, \langle M \rangle_T^{p/2})^{2/p}.
\end{aligned}$$

Doob's inequality $\mathbb{E} \, (M_T^*)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E} \, |M_T|^p$ (see (2.3)) then gives

$$\mathbb{E} \, (M_T^*)^p \leq (\frac{p}{p-1})^p \frac{1}{2}p(p-1)(\mathbb{E} \, (M_T^*)^p)^{1-2/p} (\mathbb{E} \, \langle M \rangle_T^{p/2})^{2/p},$$

from which we obtain

$$\mathbb{E} \, (M_T^*)^p \leq (\frac{p}{p-1})^{p^2/2} (\frac{1}{2}p(p-1))^{p/2} \mathbb{E} \, \langle M \rangle_T^{p/2},$$

which proves the assertion for bounded M. If M is not bounded we apply the above result to M^{T^n}, where the stopping times T^n are such that M^{T^n} are martingales bounded by n,

$$\mathbb{E} \, (M_{T \wedge T^n}^*)^p = \mathbb{E} \, ((M^{T^n})_T^*)^p \leq C_p \mathbb{E} \, \langle M^{T^n} \rangle_T^{p/2} = C_p \mathbb{E} \, \langle M \rangle_{T \wedge T^n}^{p/2} \leq C_p \mathbb{E} \, \langle M \rangle_T^{p/2}.$$

The result is obtained by applying Fatou's lemma to the left hand side of the above display. □

Remark 7.15 If in Proposition 7.14 a stopping time T is not finite, one may for instance replace the left hand side in (7.11) with $\mathbb{E} \sup_{t<T} |M_t|^p$ and leave the right hand side unaltered to still have a valid inequality.

7.5 Fubini's theorem for stochastic integrals

The main result of this section is a version of Fubini's theorem for stochastic integrals w.r.t. semimartingales, Theorem 7.18. In the special case where the semimartingale is of finite variation, there is not much more to be done than invoking the ordinary Fubini theorem. But if the semimartingale has a non vanishing martingale part, e.g. if it is itself a (local) martingale, additional arguments are needed.

Here are the preliminaries of this section. Apart from the usual filtered probability space, we have an additional measure space $(X, \mathcal{X}, \mu)$. It is assumed throughout this section that μ is a finite measure. We will often consider functions $H : X \times [0, \infty) \times \Omega \to \mathbb{R}$, with values $H(x, t, \omega)$, also written as $H_t(x, \omega)$. By $H(x)$ we then denote the mapping $(t, \omega) \mapsto H(x, t, \omega)$, and the mappings $\omega \mapsto H(x, t, \omega)$ are denoted $H_t(x)$. Under appropriate measurability conditions on H, see below, $H(x)$ will be a stochastic process and $H_t(x)$ a random variable. To prepare for Fubini's theorem in the present context, we need a number of technical results. We simple write $\mathcal{B}$ for the Borel sets of $[0, \infty)$.

Lemma 7.16 *Let $Y^n : X \times [0, \infty) \times \Omega \to \mathbb{R}$ be $\mathcal{X} \times \mathcal{B} \times \mathcal{F}$-measurable for each n. Suppose that the functions $t \mapsto Y^n(x, t, \omega)$ are right-continuous and that $Y^n(x)$ is ucp-convergent to a limit process $Y(x)$ for all $x \in X$. Then the limit process can be chosen to be $\mathcal{X} \times \mathcal{B} \times \mathcal{F}$-measurable and a.s. right-continuous for each x.*

Proof Let $S^{ij}_u(x) = \sup_{s \le u} |Y^i_s(x) - Y^j_s(x)|$. By right-continuity, $(x, \omega) \mapsto S^{ij}_u(x, \omega)$ is $\mathcal{X} \times \mathcal{F}$-measurable and the hypothesis implies $S^{ij}_u(x) \xrightarrow{\mathbb{P}} 0$. We will modify the $S^{ij}_u(x)$ to obtain a.s. convergence. Thereto we define $n_0(x) \equiv 1$ and inductively

$$n_k(x) = \inf\{m > \max\{k, n_{k-1}(x)\} : \sup_{i,j \ge m} \mathbb{P}(S^{ij}_k(x) > 2^{-k}) \le 2^{-k}\}.$$

Note that $n_k(x) \to \infty$ for all x. Put $Z^k(x, t, \omega) = Y^{n_k(x)}(x, t, \omega)$ and verify that the Z^k are measurable functions of x by the fact that the n_k depend measurably on x, which follows from the usual Fubini setup. Similar to the $S^{ij}_u(x)$ above, we define $T^{ij}_u(x) = \sup_{s \le u} |Z^i_s(x) - Z^j_s(x)|$ and we have $(x, \omega) \mapsto T^{ij}_u(x, \omega)$ is $\mathcal{X} \times \mathcal{F}$-measurable, by the fact that the $Z^i(x)$ have right-continuous paths. By construction, we have $\mathbb{P}(T^{k,k+m}_k(x) > 2^{-k}) \le 2^{-k}$ for $m \ge 1$ and then by Borel-Cantelli $\mathbb{P}(\liminf_{k \to \infty} E_k) = 1$, where $E_k = \{T^{k,k+m}_k(x) \le 2^{-k}\}$. Hence eventually, for all $u > 0$ one has $\sup_{i,j \ge k} T^{ij}_u(x) \le 2^{-k}$. In other words we have, as desired, $\lim_{i,j \to \infty} T^{ij}_u(x) = 0$ a.s. This entails that the existence of random variables $Z_t(x)$ such that $\sup_{t \le u} |Z^i_t(x) - Z_t(x)| \to 0$ a.s. Hence for all x there exist a set $L(x) \in \mathcal{X} \times \mathcal{F}$ of probability one such that for all $\omega \in L(x)$ we have the uniform convergence $\sup_{t \le u} |Z^i_t(x, \omega) - Z_t(x, \omega)| \to 0$ and hence the functions $t \to Z_t(x, \omega)$ are right-continuous. By defining $Y_t(x, \omega) = \mathbf{1}_{L(x)}(\omega) Z_t(x, \omega)$, we obtain the desired limit. □

The assertion of the lemma is also true, if one replaces right-continuity with left-continuity, continuity or càdlàg. In the next proposition $\mathcal{L}$ denotes the σ-algebra on $[0,\infty) \times \Omega$ generated by the adapted locally bounded processes.

Proposition 7.17 *Let X be a semimartingale, $X_0 = 0$ and $H : X \times [0,\infty) \times \Omega \to \mathbb{R}$ be $\mathcal{X} \times \mathcal{L}$-measurable. Then there exists $Z : X \times [0,\infty) \times \Omega \to \mathbb{R}$ that is $\mathcal{X} \times \mathcal{B} \times \mathcal{F}$-measurable, such that for all $x \in X$, $t \mapsto Z_t(x)$ is a continuous, adapted version of $H(x) \cdot X$.*

Proof We first give the proof for bounded H, then certainly $H \in \mathcal{L}$. Let $\mathcal{H}$ be the set of bounded $\mathcal{X}\times\mathcal{B}\times\mathcal{F}$-measurable processes for which the assertion holds true. Clearly, $\mathcal{H}$ is a vector space and the H given by $H(x,t,\omega) = f(x)K(t,w)$ for a bounded and $\mathcal{X}$-measurable f and K bounded and progressive belong to $\mathcal{H}$, as is straightforward to verify. Moreover, the H of the above product form generate the σ-algebra $\mathcal{X} \times \mathcal{L}$.

Let then $(H^n) \subset \mathcal{H}$ and suppose that one has $H^n \to H$ pointwise in all the arguments for some bounded H. By Proposition 7.8 we conclude that $H^n(x)\cdot X \overset{\text{ucp}}{\to} H(x)\cdot X$. Invoking Lemma 7.16 allows us to conclude the existence of a continuous adapted version of $H(x) \cdot X$ that inherits the $\mathcal{X} \times \mathcal{B} \times \mathcal{F}$-measurability. So $H \in \mathcal{H}$. Application of the Monotone class theorem concludes the proof for bounded H.

In the last step we drop the requirement that H is bounded. For such H we have a continuous $\mathcal{X}\times\mathcal{B}\times\mathcal{F}$-measurable version $Z^k(x)$ of $H(x)\mathbf{1}_{\{|H(x)|\leq k\}}\cdot X$ by the first part of the proof. By Proposition 7.8 again we have $Z^k(x) \overset{\text{ucp}}{\to} H(x)\cdot X$. The result now follows in view of Lemma 7.16. □

Here is Fubini's theorem for stochastic integrals.

Theorem 7.18 *Let X be a semimartingale, H as in Proposition 7.17 and bounded. Let $Z(x)$ be a continuous version of $H(x) \cdot X$ for each x. Then the process Z given by $Z_t(\omega) := \int_X Z(x,t,\omega)\,\mu(\mathrm{d}x)$ is a continuous version of $\bar{H}\cdot X$, where $\bar{H}$ is the process defined by $\bar{H}_t(\omega) = \int_X H(x,t,\omega)\,\mu(\mathrm{d}x)$. In appealing notation, we thus have*

$$\int_X \left(\int_0^t H_s(x)\mathrm{d}X_s\right)\,\mu(\mathrm{d}x) = \int_0^t \left(\int_X H_s(x)\,\mu(\mathrm{d}x)\right)\,\mathrm{d}X_s.$$

Proof If we decompose a stochastic integral as in Definition 7.6, we can apply the ordinary version of Fubini's theorem to the integral w.r.t. finite variation part. We therefore proceed under the assumption that $X \in \mathcal{M}_c^{loc}$. For the time being we even assume that $\mathbb{E}\,\langle X\rangle_\infty < \infty$. Since μ is a finite measure, we henceforth assume, by normalization, without loss of generality that μ is a probability measure. It follows that also $\mu \times \mathbb{P}$ is a probability measure on $(X \times \Omega, \mathcal{X} \times \mathcal{F})$, which will be exploited below.

Suppose that $H(x,t,\omega) = f(x)K(t,\omega)$ as in the proof of Proposition 7.17. Then $K \in \mathcal{L}$, $\mu(|f|) < \infty$ and $Z_t(x) = f(x)(K \cdot X)_t$. This immediately yields $\int_X Z_t(x)\,\mu(\mathrm{d}x) = (\bar{H}\cdot X)_t$. By linearity this identity continues to hold for linear combinations of functions of the type $H(x,t,\omega) = f(x)K(t,\omega)$. Note that these linear combinations generate a vector space, V say.

Let now $(H^n) \subset V$ and assume pointwise convergence of the H^n to H. Based on the Monotone class theorem, we will show that the result is also valid for H. Let $Z^n(x)$ be a continuous adapted version of $H^n(x) \cdot X$, which exists in view of Proposition 7.17, and let $Z(x)$ be a continuous adapted version of $H(x) \cdot X$. Then by the Cauchy-Schwarz and Doob inequalities we have

$$\begin{aligned}\left(\mathbb{E}\int_X \sup_t |Z^n_t(x) - Z_t(x)|\,\mu(\mathrm{d}x)\right)^2 &\leq \mathbb{E}\int_X \sup_t |Z^n_t(x) - Z_t(x)|^2\,\mu(\mathrm{d}x)\\ &= \int_X \mathbb{E}\sup_t |Z^n_t(x) - Z_t(x)|^2\,\mu(\mathrm{d}x)\\ &\leq 4\int_X \mathbb{E}\,|Z^n_\infty(x) - Z_\infty(x)|^2\,\mu(\mathrm{d}x).\end{aligned}$$

Of course, the right hand side is equal to

$$4\int_X \mathbb{E}\,\langle Z^n(x) - Z(x)\rangle_\infty\,\mu(\mathrm{d}x) = 4\int_X \mathbb{E}\int_0^\infty (H^n_s(x) - H_s(x))^2\mathrm{d}\langle X\rangle_s\,\mu(\mathrm{d}x).$$

A threefold application of the Dominated convergence theorem (recall that H is bounded) to the triple integral gives $\mathbb{E}\left(\int_X \sup_t |Z^n_t(x) - Z_t(x)|\,\mu(\mathrm{d}x)\right)^2 \to 0$. It then also follows that $\int_X \sup_t |Z^n_t(x) - Z_t(x)|\,\mu(\mathrm{d}x)$ is a.s. finite and the same holds for $\int_X |Z_t(x)|\,\mu(\mathrm{d}x)$. Consider now

$$\mathbb{E}\int_X \sup_t |Z^n_t(x) - Z_t(x)|\,\mu(\mathrm{d}x) \leq \mathbb{E}\int_X \sup_t |Z^n_t(x) - Z_t(x)|\,\mu(\mathrm{d}x), \quad (7.12)$$

which tends to zero by the above. With $\bar{H}^n_t = \int_X H^n_t(x)\,\mu(\mathrm{d}x)$ we have $\bar{H}^n \cdot X = \int_X Z^n_t(x)\,\mu(\mathrm{d}x)$ (recall $H^n \in V$) and the $\int_X Z^n_t(x)\,\mu(\mathrm{d}x)$ converge ucp-wise to $\int_X Z_t(x)\,\mu(\mathrm{d}x)$ in view of (7.12). But Proposition 7.8 yields $\bar{H}^n \cdot X \overset{\text{ucp}}{\to} \bar{H} \cdot X$ and so we must have $(\bar{H} \cdot X)_t = \int_X Z_t(x)\,\mu(\mathrm{d}x)$. This concludes the proof under the assumption that $X \in \mathcal{M}^{loc}_c$ with $\mathbb{E}\,\langle X\rangle_\infty < \infty$. The general case follows by a usual stopping time argument, Exercise 7.9. □

7.6 Exercises

7.1 The Hermite polynomials h_n are defined as

$$h_n(x) = (-1)^n \exp(\frac{1}{2}x^2)\frac{\mathrm{d}^n}{\mathrm{d}x^n}\exp(-\frac{1}{2}x^2).$$

Let $H_n(x,y) = y^{n/2}h_n(x/\sqrt{y})$.

(a) Show that $\frac{\partial}{\partial x}H_n(x,y) = nH_{n-1}(x,y)$, for which you could first prove that

$$\sum_{n\geq 0}\frac{u^n}{n!}h_n(x) = \exp(ux - \frac{1}{2}u^2).$$

(b) Show also that $\frac{\partial}{\partial y}H_n(x,y) + \frac{1}{2}\frac{\partial^2}{\partial x^2}H_n(x,y) = 0$.

(c) Show finally that for a Brownian motion W it holds that $H_n(W_t, t) = n\int_0^t H_{n-1}(W_s, s)\,\mathrm{d}W_s$.

7.2 Let W be Brownian motion and $X \in \mathcal{S}$. Let $M = X \cdot W$ and $Z = \mathcal{E}(M)$. Show that M and Z are martingales.

7.3 Let X and Y be (continuous) semimartingales. Show that $\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + \langle X, Y\rangle)$. Deduce from this the formula $\mathcal{E}(X)/\mathcal{E}(Y) = \mathcal{E}(X - Y - \langle X - Y, Y\rangle)$. Suppose further that $X = H \cdot Z$ and $Y = K \cdot Z$ (H and K both satisfying appropriate conditions) for some semimartingale Z. Show that $\mathcal{E}(H \cdot Z)/\mathcal{E}(K \cdot Z) = \mathcal{E}((H - K) \cdot \tilde{Z})$, where $\tilde{Z} = Z - K \cdot \langle Z\rangle$.

7.4 Let X be a strictly positive continuous semimartingale with $X_0 = 1$ and define the process Y by

$$Y_t = \int_0^t \frac{1}{X}\,\mathrm{d}X - \frac{1}{2}\int_0^t \frac{1}{X^2}\,\mathrm{d}\langle X\rangle.$$

Let the process Z be given by $Z_t = e^{Y_t}$. Compute $\mathrm{d}Z_t$ and show that $Z = X$.

7.5 Let W_1, W_2, W_3 be three independent Brownian motions. Let

$$M_t = \left(W_{1,t}^2 + W_{2,t}^2 + W_{3,t}^2\right)^{-1/2},\ t \geq 1.$$

Show that $M = \{M_t : t \geq 1\}$ is a local martingale with $\sup_{t\geq 1} \mathbb{E}\, M_t^p < \infty$ if $0 \leq p < 3$ (the sum of squares has a Gamma distribution!) and that M is *not* a martingale. Is M uniformly integrable?

7.6 Let $f : [0,\infty) \to \mathbb{R}$ be measurable, $\int_0^t f(s)^2\,\mathrm{d}s < \infty$ for all $t \geq 0$ and W a Brownian motion. Then $X_t = \int_0^t f(s)\,\mathrm{d}W_s$ is well defined for all $t \geq 0$. Show that X_t has a normal distribution and determine its mean and variance.

7.7 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be twice continuously differentiable in the first variable and continuously differentiable in the second variable. Let X be a continuous semimartingale and B a continuous process of finite variation over finite intervals. Show, depart from the formula in Remark 7.12, that for $t \geq 0$

$$\begin{aligned} f(X_t, B_t) = f(X_0, B_0) &+ \int_0^t f_x(X, B)\,\mathrm{d}X \\ &+ \int_0^t f_y(X, B)\,\mathrm{d}B + \frac{1}{2}\int_0^t f_{xx}(X, B)\,\mathrm{d}\langle X\rangle, \text{ a.s.} \end{aligned}$$

7.8 Let X be a strictly positive (continuous) semimartingale with standard decomposition $X = X_0 + A + M$.

(a) Show that X also admits a unique *multiplicative* decomposition $X = X_0\tilde{A}\tilde{M}$, where $\tilde{A}$ a positive process of finite variation over bounded intervals and $\tilde{M}$ a positive local martingale.

(b) Express $\tilde{A}$ and $\tilde{M}$ in X, A and M.

(c) What extra can you say, if X is a supermartingale, of the paths of $\tilde{A}$? Are $\tilde{A}$ and $\tilde{M}$ convergent in this case?

7.9 Finish the proof of Theorem 7.18 by dropping the assumption $\mathbb{E}\,\langle X\rangle_\infty < \infty$ for $X \in \mathcal{M}_c^{loc}$.

8 Integral representations

We assume that the underlying filtration satisfies the usual conditions. However, in Section 8.2 we will encounter a complication and see how to repair this.

8.1 First representation

The representation result below explains a way how to view any continuous local martingale as a stochastic integral w.r.t. a suitably defined process W that is a Brownian motion.

Proposition 8.1 *Let M be a continuous local martingale with $M_0 = 0$ whose quadratic variation process $\langle M \rangle$ is almost surely absolutely continuous, $\langle M \rangle$ can be written as $\int_0^\cdot X_s \,\mathrm{d}s$ for some process X. Then there exists, possibly defined on an enlarged probability space, an $\mathbb{F}$-Brownian motion W and process Y in $\mathcal{P}^*(W)$ such that $M = Y \cdot W$ (up to indistinguishability).*

Proof Let X^n for $n \in \mathbb{N}$ be defined by $X^n_t = n(\langle M \rangle_t - \langle M \rangle_{t-1/n})$ (with $\langle M \rangle_s = 0$ for $s < 0$). Then all the X^n are progressive processes and thanks to the fundamental theorem of calculus their limit X' for $n \to \infty$ exists for Lebesgue almost all $t > 0$ a.s. by assumption, and is progressive as well. Hence we can assume that X is progressive.

Suppose that $X_t > 0$ a.s. for all $t > 0$. Then we define $f_t = X_t^{-1/2}$ and $W = f \cdot M$ and we notice that $f \in \mathcal{P}^*(M)$. From the calculus rules for computing the quadratic variation of a stochastic integral we see that $\langle W \rangle_t \equiv t$. Hence, by Lévy's characterization (Proposition 7.13), W (which is clearly adapted) is a Brownian motion. By the chain rule for stochastic integrals (Proposition 6.12) we obtain that $X^{1/2} \cdot W$ is indistinguishable from M, so we can take $Y = X^{1/2}$.

If X_t assumes for some t the value zero with positive probability, we cannot define the process f as we did above. Let in this case $(\Omega', \mathbb{F}', \mathbb{P}')$ be another probability space that is rich enough to support a Brownian motion B. We consider now the product space of $(\Omega, \mathcal{F}, \mathbb{P})$ with $(\Omega', \mathbb{F}', \mathbb{P}')$ and redefine in the obvious way M and B (as well as the other processes that we need below) on this product space. For instance, one has $M_t(\omega, \omega') = M_t(\omega)$. Notice that everything defined on the original space now becomes independent of B. Next, one needs a filtration on the enlarged probability space. The proper choice is to take the family of σ-algebras $\mathcal{F}_t \times \mathcal{F}^B_t$. Having done so, one finds for instance that M is still a local martingale and B remains a Brownian motion. Hence the stochastic integrals $\mathbf{1}_{\{X>0\}} X^{-1/2} \cdot M$ and $\mathbf{1}_{\{X=0\}} \cdot B$ are well defined on the product space and $\mathbf{1}_{\{X>0\}} X^{-1/2} \cdot M$ is a local martingale w.r.t. the product filtration and $\mathbf{1}_{\{X=0\}} \cdot B$ a martingale. These two (local) martingales have zero quadratic covariation, since $\langle M, B \rangle = 0$ in view of the independence (Exercise 3.6).

Now we are able to define in the present situation the process W by

$$W = \mathbf{1}_{\{X>0\}} X^{-1/2} \cdot M + \mathbf{1}_{\{X=0\}} \cdot B.$$

Observe that $\langle W\rangle_t = \int_0^t \mathbf{1}_{\{X>0\}} X^{-1}\,\mathrm{d}\langle M\rangle + \int_0^t \mathbf{1}_{\{X=0\}}\mathrm{d}\langle B\rangle = t$. Hence W is also a Brownian motion in this case. Finally, again by the chain rule for stochastic integrals, $X^{1/2}\cdot W = \mathbf{1}_{\{X>0\}}\cdot M + \mathbf{1}_{\{X=0\}}X^{1/2}\cdot B = \mathbf{1}_{\{X>0\}}\cdot M$. Hence $M - X^{1/2}\cdot W = \mathbf{1}_{\{X=0\}}\cdot M$ has quadratic variation identically zero and is thus indistinguishable from the zero martingale. Again we can take $Y = X^{1/2}$.□

8.2 Representation of Brownian local martingales

Suppose that on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we can define a Brownian motion W. Let $\mathbb{F}^W$ be the filtration generated by this process. Application of the preceding theory to situations that involve this filtration is not always justified since this filtration doesn't satisfy the usual conditions (one can show that it is not right-continuous, see Exercise 8.1). However, we have

Proposition 8.2 *Let $\mathbb{F}$ be the filtration with σ-algebras $\mathcal{F}_t = \mathcal{F}_t^W \vee \mathcal{N}$, where $\mathcal{N}$ is the collection of sets that are contained in sets of $\mathcal{F}_\infty^W$ with zero probability. Then $\mathbb{F}$ satisfies the usual conditions and W is also a Brownian motion w.r.t. this filtration.*

Proof The proof of right-continuity involves arguments that use the Markovian character of Brownian motion (the result can be extended to hold for more general Markov processes) and will not be given here. Clearly, by adding null sets to the original filtration the distributional properties of W don't change. □

Remark 8.3 It is remarkable that the addition of the null sets to the σ-algebras of the given filtration renders the filtration right-continuous.

Any process that is a martingale w.r.t. the filtration of Proposition 8.2 (it will be referred to as the augmented Brownian filtration) is called a *Brownian martingale*. Below, in Theorem 8.7, we sharpen the result of Proposition 8.1 in the sense that the Brownian motion is now given and not constructed and that moreover the integrand process X is progressive w.r.t. the *augmented* Brownian filtration $\mathbb{F}$. In the proof of that theorem we shall use the following result.

Lemma 8.4 *Let $T > 0$ and $\mathcal{R}_T$ be the subset of $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ consisting of the random variables $(X\cdot W)_T$ for $X \in \mathcal{P}_T(W)$ and let $\mathcal{R}$ be the class of stochastic integrals $X \cdot W$, where X ranges through $\mathcal{P}(W)$. Then $\mathcal{R}_T$ is a closed subspace of $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$. Moreover, every martingale M in $\mathcal{M}^2$ can be uniquely (up to indistinguishability) written as the sum $M = N + Z$, where $N \in \mathcal{R}$ and $Z \in \mathcal{M}^2$ that is such that $\langle Z, N'\rangle = 0$ for every $N' \in \mathcal{R}$.*

Proof Let $((X^n\cdot W)_T)$ be a converging sequence in $\mathcal{R}_T$, then it is also Cauchy. By the construction of the stochastic integrals (the isometry property in particular), we have that (X^n) is Cauchy in $\mathcal{P}_T(W)$. But this is a complete space (Proposition 5.7) and thus (X^n) has a limit X in this space. By the isometry property again, we have that $(X^n\cdot W)_T$ converges in L^2 to $(X\cdot W)_T$, which is an element of $\mathcal{R}_T$.

The uniqueness of the decomposition in the second assertion is established as follows. Suppose that a given $M \in \mathcal{M}^2$ can be decomposed as $N^1 + Z^1 = N^2 + Z^2$. Then $0 = \langle Z^1 - Z^2, N^2 - N^1 \rangle = \langle Z^1 - Z^2 \rangle$, hence the uniqueness follows.

We now prove the existence of the decomposition on an arbitrary (but fixed) interval $[0, T]$. Invoking the already established uniqueness, one can extend the existence to $[0, \infty)$. Since $M_T \in L^2(\Omega, \mathcal{F}_T, \mathbb{P})$ and $\mathcal{R}_T$ is closed, we have the unique decomposition $M_T = N_T + Z_T$, with $N_T \in \mathcal{R}_T$ and $\mathbb{E}\, Z_T N'_T = 0$ for any $N'_T \in \mathcal{R}_T$. Let Z be the right-continuous modification of the martingale defined by $\mathbb{E}\,[Z_T|\mathcal{F}_t]$ and note that $\mathbb{E}\,[N_T|\mathcal{F}_t] \in \mathcal{R}_t$. Then $M = N + Z$ on $[0, T]$. We now show that $\langle Z, N' \rangle_t = 0$ for $t \leq T$, where $N' \in \mathcal{R}$. Equivalently, we show that ZN' is a martingale on $[0, T]$. Write $N'_T = (X \cdot W)_T$ for some $X \in \mathcal{P}_T(W)$. Let $F \in \mathcal{F}_t$. Then the process Y given by $Y_u = \mathbf{1}_F \mathbf{1}_{(t,T]}(u) X_u$ is progressive and in $\mathcal{P}_T(W)$, and we have $(Y \cdot W)_T = \mathbf{1}_F (N'_T - N'_t)$. (This is not immediately obvious, since $\mathbf{1}_F$ is random, Exercise 8.7.) It follows that

$$
\begin{aligned}
\mathbb{E}\,[\mathbf{1}_F Z_T N'_T] &= \mathbb{E}\,[\mathbf{1}_F Z_T (N'_T - N'_t)] + \mathbb{E}\,[\mathbf{1}_F Z_t N'_t] \\
&= \mathbb{E}\,[Z_T (Y \cdot W)_T] + \mathbb{E}\,[\mathbf{1}_F Z_t N'_t].
\end{aligned}
$$

Since Z_T is orthogonal in the L^2-sense to $\mathcal{R}_T$, the expectation $\mathbb{E}\,[Z_T (Y \cdot W)_T] = 0$. We obtain

$$
\mathbb{E}\,[\mathbf{1}_F Z_T N'_T] = \mathbb{E}\,[\mathbf{1}_F Z_t N'_t],
$$

equivalently, $\mathbb{E}\,[Z_T N'_T|\mathcal{F}_t] = Z_t N'_t$, since $F \in \mathcal{F}_t$ is arbitrary. This shows that ZN' is a martingale on $[0, T]$. □

Remark 8.5 In the proof of the above lemma we have not exploited the fact that we deal with Brownian motion, nor did we use the special structure of the filtration (other than it satisfies the usual conditions). The lemma can therefore be extended to other martingales than Brownian motion.

However, Theorem 8.7 shows that the process Z of Lemma 8.4 in the Brownian context is actually zero. It is known as the (Brownian) *martingale representation theorem*. In its proof we use the following lemma.

Lemma 8.6 *Let Z be a Brownian martingale, that is orthogonal to the space $\mathcal{R}$ in Lemma 8.4. Let $t > 0$. For any $n \in \mathbb{N}$ and bounded Borel measurable functions $f_k : \mathbb{R} \to \mathbb{R}$ ($k = 1, \ldots, n$) one has for $0 = s_0 \leq \cdots \leq s_n \leq t$ that*

$$
\mathbb{E}\,\big(Z_t \prod_{k=1}^{n} f_k(W_{s_k})\big) = 0. \tag{8.1}
$$

Proof For $n = 0$ this is trivial since $Z_0 = 0$. We use induction to show (8.1). Suppose that (8.1) holds true for a given n and suppose w.l.o.g. that $s_n < t$ and let $s \in [s_n, t]$. By P_n we abbreviate the product $\prod_{k=1}^{n} f_k(W_{s_k})$. Put for all $\theta \in \mathbb{R}$

$$
\phi(s, \theta) = \mathbb{E}\,\big(Z_t P_n e^{\mathrm{i}\theta W_s}\big).
$$

We keep θ fixed for the time being. Observe that by the martingale property of Z we have $\phi(s,\theta) = \mathbb{E}\, Z_s P_n e^{\mathrm{i}\theta W_s}$ and by the induction assumption we have $\phi(s_n,\theta) = 0$. Below we will need that

$$\mathbb{E}\,[Z_s P_n \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u] = 0, \tag{8.2}$$

which is by reconditioning a consequence of

$$\mathbb{E}\,[Z_s \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] = 0. \tag{8.3}$$

Indeed,

$$\begin{aligned}\mathbb{E}\,[Z_s P_n \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u] &= \mathbb{E}\,\mathbb{E}\,[Z_s P_n \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] \\ &= \mathbb{E}\,\big(P_n \mathbb{E}\,[Z_s \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}]\big) = 0,\end{aligned}$$

by Equation (8.3). To prove this equation, we compute

$$\begin{aligned}&\mathbb{E}\,[Z_s \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] \\ &= \mathbb{E}\,[(Z_s - Z_{s_n}) \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] + \mathbb{E}\,[Z_{s_n} \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] \\ &= \mathbb{E}\,[(Z_s - Z_{s_n}) \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u|\mathcal{F}_{s_n}] \\ &= \mathbb{E}\,[\langle Z, \int_0^{\cdot} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u\rangle_s - \langle Z, \int_0^{\cdot} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u\rangle_{s_n}|\mathcal{F}_{s_n}] \\ &= 0,\end{aligned}$$

since Z is orthogonal to stochastic integrals that are in $\mathcal{R}$ (Lemma 8.4). As in the proof of Proposition 7.13 we use Itô's formula to get

$$e^{\mathrm{i}\theta W_s} = e^{\mathrm{i}\theta W_{s_n}} + \mathrm{i}\theta \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}W_u - \frac{1}{2}\theta^2 \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}u. \tag{8.4}$$

Multiplication of Equation (8.4) by $Z_s P_n$, taking (conditional) expectations and using the just shown fact (8.2), yields

$$\mathbb{E}\,[Z_s P_n e^{\mathrm{i}\theta W_s}] = \mathbb{E}\,[Z_{s_n} P_n e^{\mathrm{i}\theta W_{s_n}}] - \frac{1}{2}\theta^2 \mathbb{E}\,[Z_s P_n \int_{s_n}^{s} e^{\mathrm{i}\theta W_u}\,\mathrm{d}u].$$

Use the fact that we also have $\phi(s,\theta) = \mathbb{E}\left(Z_s P_n e^{\mathrm{i}\theta W_s}\right)$, Fubini and reconditioning to obtain

$$\phi(s,\theta) = \phi(s_n,\theta) - \frac{1}{2}\theta^2 \int_{s_n}^{s} \mathbb{E}\left(P_n e^{\mathrm{i}\theta W_u}\mathbb{E}\left[Z_s|\mathcal{F}_u\right]\right) \mathrm{d}u,$$

which then becomes

$$\phi(s,\theta) = \phi(s_n,\theta) - \frac{1}{2}\theta^2 \int_{s_n}^{s} \phi(u,\theta)\,\mathrm{d}u.$$

Since $\phi(s_n,\theta) = 0$, the unique solution to this integral equation is the zero solution. Hence $\phi(s_{n+1},\theta) = 0$ for all θ. Stated otherwise, the Fourier transform of the signed measure μ on $\mathcal{B}(\mathbb{R})$ given by $\mu(B) = \mathbb{E}\left(\mathbf{1}_B(W_{s_{n+1}})Z_t P_n\right)$ is identically zero. But then μ must be the zero measure and thus for any bounded measurable function f_{n+1} we have $\int f_{n+1}\,\mathrm{d}\mu = 0$, nothing else but $\mathbb{E}\left(f_{n+1}(W_{s_{n+1}})Z_t P_n\right)$. Hence we have (8.1) also for n replaced with $n+1$ and thus for all n, which completes the proof. □

Theorem 8.7 *Let M be a square integrable Brownian martingale with $M_0 = 0$. Then there exists a process $X \in \mathcal{P}(W)$ such that $M = X \cdot W$. The process X is unique in the sense that for any other process X' with the same property one has $||X - X'||_W = 0$.*

Proof As a starting point we take the decomposition $M = N + Z$ of Lemma 8.4. We will show that Z is indistinguishable from zero. We use Lemma 8.6. By the monotone class theorem (Exercise 8.2) we conclude from (8.1) that $\mathbb{E}\, Z_t\xi = 0$ for all bounded $\mathcal{F}_t^W$-measurable functions ξ. But then also for all bounded ξ that are $\mathcal{F}_t$-measurable, because the two σ-algebras differ only by null sets. We conclude that $Z_t = 0$ a.s. Since this holds true for each t separately, we have by right-continuity of Z that Z is indistinguishable from the zero process. The uniqueness of the assertion is trivial. □

Corollary 8.8 *Let M be a right-continuous local martingale adapted to the augmented Brownian filtration $\mathbb{F}$ with $M_0 = 0$. Then there exists a process $X \in \mathcal{P}^*(W)$ such that $M = X \cdot W$. In particular all such local martingales are continuous.*

Proof Let (T^n) be a localizing sequence for M such that the stopped processes M^{T^n} are bounded martingales. Then we have according to Theorem 8.7 the existence of processes X^n such that $M^{T^n} = X^n \cdot W$. By the uniqueness of the representation we have that $X^n \mathbf{1}_{[0,T^{n-1}]} = X^{n-1}\mathbf{1}_{[0,T^{n-1}]}$. Hence we can unambiguously define $X_t = \lim_n X_t^n$ and it follows from $X^n\mathbf{1}_{[0,T^n]} = X\mathbf{1}_{[0,T^n]}$ that $M_t = \lim_n M_{t\wedge T^n} = (X \cdot W)_t$. □

8.3 Exercises

8.1 The filtration $\mathbb{F}^W$ is not right-continuous. Let $\Omega = C[0,\infty)$ and $W_t(\omega) = \omega(t)$ for $t \geq 0$. Fix $t > 0$ and let F be the set of functions $\omega \in \Omega$ that have a local maximum at t. Show that $F \in \mathcal{F}_{t+}$. Suppose that $F \in \mathcal{F}_t$. Since any set G in $\mathcal{F}_t$ is determined by the paths of functions ω up to time t, such a set is unchanged if we have continuous continuations of such functions after time t. In particular, if $\omega \in G$, then also $\omega' \in G$, where $\omega'(s) = \omega(s \wedge t) + (s-t)^+$. Notice that for any ω the function ω' doesn't have a local maximum at t. Conclude that $F \notin \mathcal{F}_t$.

8.2 Complete the proof of Theorem 8.7 by writing down the Monotone Class argument.

8.3 The result of Theorem 8.7 is not constructive, it is not told how to construct the process X from the given Brownian martingale M. In the following cases we can give an explicit expression for X. (If $M_0 \neq 0$, you have to adjust this theorem slightly.)

(a) $M_t = W_t^3 - c\int_0^t W_s \, \mathrm{d}s$ for a suitable constant c (which one?).
(b) For some fixed time T we have $M_t = \mathbb{E}[e^{W_T} | \mathcal{F}_t]$.
(c) For some fixed time T we take $M_t = \mathbb{E}[\int_0^T W_s \, \mathrm{d}s | \mathcal{F}_t]$.
(d) If v is a solution to the *backward heat equation*

$$\frac{\partial v}{\partial t}(t,x) + \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(t,x) = 0,$$

then $M_t = v(t, W_t)$ is a martingale. Show this to be true under a to be specified integrability condition. Give also two examples of a martingale M that can be written in this form.
(e) Suppose a square integrable martingale M is of the form $M_t = v(t, W_t)$, where W is a standard Brownian motion and $v : \mathbb{R}^2 \to \mathbb{R}$ is continuously differentiable in the first variable and twice continuously differentiable in the second variable. Show that v satisfies the backward heat equation.

8.4 Let M be as in Theorem 8.7. Show that $\langle M, W \rangle$ is a.s. absolutely continuous. Express X in terms of $\langle M, W \rangle$.

8.5 Let ξ be a square integrable random variable that is $\mathcal{F}_\infty^W$-measurable, where $\mathcal{F}_\infty^W$ is the σ-algebra generated by a Brownian motion on $[0,\infty)$. Show that there exists a unique process $X \in \mathcal{P}(W)_\infty$ such that

$$\xi = \mathbb{E}\,\xi + \int_0^\infty X \, dW.$$

8.6 The uniqueness result of Exercise 8.5 relies on $X \in \mathcal{P}(W)_\infty$. For any $T > 0$ we define $S_T = \inf\{t > T : W_t = 0\}$. It is known that the S_T are a.s. finite stopping times, but with infinite expectation. Take $\xi = 0$ in Exercise 8.5. Show that $\xi = \int_0^\infty \mathbf{1}_{[0,S_T]} \, \mathrm{d}W$ for any T. Why is this not a contradiction?

8.7 Show the equality $(Y \cdot W)_T = \mathbf{1}_F(N'_T - N'_t)$ in the proof of Lemma 8.4. (Although not necessary, you may use Exercise 6.14.)

9 Absolutely continuous change of measure

In Section 7.3 we have seen that the class of semimartingales is closed under smooth transformations; if X is a semimartingale, so is $f(X)$ if f is twice continuously differentiable. In the present section we will see that the semimartingale property is preserved, when we change the underlying probability measure in an absolutely continuous way. This result is absolutely not obvious. Indeed, consider a semimartingale X with decomposition $X = X_0 + A + M$ with all paths of A of bounded variation. Then this property of A evidently still holds, if we change the probability measure $\mathbb{P}$ into any other one. But it is less clear what happens to M. As a matter of fact, M (suppose it is a martingale under $\mathbb{P}$) will in general loose the martingale property. We will see later on that it becomes a semimartingale and, moreover, we will be able to give its semimartingale decomposition under the new probability measure.

9.1 Absolute continuity

Let $(\Omega, \mathcal{F})$ be a measurable space. We consider two measures on this space, μ and ν. One says that ν is absolutely continuous with respect to μ (the notation is $\nu \ll \mu$) if $\nu(F) = 0$ for every $F \in \mathcal{F}$ for which $\mu(F) = 0$. If we have both $\nu \ll \mu$ and $\mu \ll \nu$ we say that μ and ν are equivalent and we write $\mu \sim \nu$. If there exists a set $\Omega_0 \in \mathcal{F}$ for which $\nu(\Omega_0) = 0$ and $\mu(\Omega_0^c) = 0$, then μ and ν are called mutually singular.

If Z is a nonnegative measurable function on Ω, then $\nu(F) = \int_F Z \, \mathrm{d}\mu$ defines a measure on $\mathcal{F}$ that is absolutely continuous w.r.t. μ. The consequence of the Radon-Nikodym theorem (Theorem 9.1 below, stated in a more general version than we need later on) is that this is, loosely speaking, the only case of absolute continuity.

Theorem 9.1 *Let μ be σ-finite measure on $(\Omega, \mathcal{F})$ and let ν be a finite measure on $(\Omega, \mathcal{F})$. Then we can uniquely decompose ν as the sum of two measures ν_s and ν_0, where ν_s and μ are mutually singular and $\nu_0 \ll \mu$. Moreover, there exists a μ-a.e. unique nonnegative $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ such that $\nu_0(F) = \int_F Z \, \mathrm{d}\mu$. This function is called the Radon-Nikodym derivative of ν_0 w.r.t. μ and is often written as*

$$Z = \frac{\mathrm{d}\nu_0}{\mathrm{d}\mu}.$$

We will be interested in the case where μ and ν are probability measures, as usual called $\mathbb{P}$ and $\mathbb{Q}$, and will try to describe the function Z in certain cases. Moreover, we will always assume (unless otherwise stated) that $\mathbb{Q} \ll \mathbb{P}$. (See section F for a proof of the Radon-Nikodym theorem for the case $\mathbb{Q} \ll \mathbb{P}$.) Then one has $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$, with $\mathbb{E}_{\mathbb{P}} Z = 1$, $\mathbb{Q}(Z = 0) = 0$ and for $\mathbb{Q} \sim \mathbb{P}$ also $\mathbb{P}(Z = 0) = 0$. To distinguisg between expectations w.r.t. $\mathbb{P}$ and $\mathbb{Q}$, these are often denoted by $\mathbb{E}_{\mathbb{P}}$ and $\mathbb{E}_{\mathbb{Q}}$ respectively. If $X \geq 0$ or $XZ \in \mathcal{L}(\Omega, \mathcal{F}, \mathbb{P})$, it follows from

Theorem 9.1 by application of the *standard machine* that $\mathbb{E}_{\mathbb{Q}} X = \mathbb{E}_{\mathbb{P}} XZ$. For *conditional* expectations, the corresponding result is different.

Lemma 9.2 *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathcal{F})$ and assume that $\mathbb{Q} \ll \mathbb{P}$ with $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let X be a random variable. Then $\mathbb{E}_{\mathbb{Q}} |X| < \infty$ iff $\mathbb{E}_{\mathbb{P}} |X|Z < \infty$ and in either case we have*

$$\mathbb{E}_{\mathbb{Q}} [X|\mathcal{G}] = \frac{\mathbb{E}_{\mathbb{P}} [XZ|\mathcal{G}]}{\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}]} \text{ a.s. w.r.t. } \mathbb{Q}. \tag{9.1}$$

Proof Let $G \in \mathcal{G}$. We have, using the defining property of conditional expectation both under $\mathbb{Q}$ and $\mathbb{P}$

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{Q}} [X|\mathcal{G}]\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}]\mathbf{1}_G\right] &= \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{Q}} [X|\mathcal{G}]\mathbf{1}_G Z\right] \\
&= \mathbb{E}_{\mathbb{Q}} \left[\mathbb{E}_{\mathbb{Q}} [X|\mathcal{G}]\mathbf{1}_G\right] \\
&= \mathbb{E}_{\mathbb{Q}} [X\mathbf{1}_G] \\
&= \mathbb{E}_{\mathbb{P}} [X\mathbf{1}_G Z] \\
&= \mathbb{E}_{\mathbb{P}} \left[\mathbb{E}_{\mathbb{P}} [XZ|\mathcal{G}]\mathbf{1}_G\right].
\end{aligned}$$

Since this holds for any $G \in \mathcal{G}$, we conclude that $\mathbb{E}_{\mathbb{Q}} [X|\mathcal{G}]\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}] = \mathbb{E}_{\mathbb{P}} [XZ|\mathcal{G}]$ $\mathbb{P}$- (and thus $\mathbb{Q}$-)a.s. Because

$$\mathbb{Q}(\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}] = 0) = \mathbb{E}_{\mathbb{P}} Z\mathbf{1}_{\{\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}]=0\}} = \mathbb{E}_{\mathbb{P}} \mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}]\mathbf{1}_{\{\mathbb{E}_{\mathbb{P}} [Z|\mathcal{G}]=0\}} = 0,$$

the division in (9.1) is $\mathbb{Q}$-a.s. justified. □

9.2 Change of measure on filtered spaces

We consider a measurable space $(\Omega, \mathcal{F})$ together with a right-continuous filtration $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ and two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined on it. The restrictions of $\mathbb{P}$ and $\mathbb{Q}$ to the $\mathcal{F}_t$ will be denoted by $\mathbb{P}_t$ and $\mathbb{Q}_t$ respectively. Similarly, for a stopping time T, we will denote by $\mathbb{P}_T$ the restriction of $\mathbb{P}$ to $\mathcal{F}_T$. Below we will always assume that $\mathbb{P}_0 = \mathbb{Q}_0$. If $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}$, then necessarily every restriction $\mathbb{Q}_t$ of $\mathbb{Q}$ to $\mathcal{F}_t$ is absolutely continuous w.r.t. the restriction $\mathbb{P}_t$ of $\mathbb{P}$ to $\mathcal{F}_t$ and thus we have a family of *densities* (Radon-Nikodym derivatives) Z_t, defined by

$$Z_t = \frac{\mathrm{d}\mathbb{Q}_t}{\mathrm{d}\mathbb{P}_t}.$$

The process $Z = \{Z_t, t \geq 0\}$ is called the density process (of $\mathbb{Q}$ w.r.t. $\mathbb{P}$). Here is the first property.

Proposition 9.3 *If $\mathbb{Q} \ll \mathbb{P}$ (on $\mathcal{F}$), then the density process Z is a nonnegative uniformly integrable martingale w.r.t. the probability measure $\mathbb{P}$ and $Z_0 = 1$.*

Proof Exercise 9.1. □

However, we will encounter many interesting situations, where we only have $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$ and where Z is not uniformly integrable. One may also envisage a reverse situation. One is given a nonnegative martingale Z, is it then possible to find probability measures $\mathbb{Q}_t$ (or $\mathbb{Q}$) such that $\mathbb{Q}_t \ll \mathbb{P}_t$ for all t (or $\mathbb{Q} \ll \mathbb{P}$)? The answers are given in the next two propositions.

Proposition 9.4 *Let Z be a nonnegative uniformly integrable martingale under the probability measure $\mathbb{P}$ with $\mathbb{E}_\mathbb{P} Z_t = 1$ (for all t). Then there exists a probability measure $\mathbb{Q}_\infty$ on $\mathcal{F}$ that is absolutely continuous w.r.t. $\mathbb{P}$. If we denote by $\mathbb{P}_t$, respectively $\mathbb{Q}_t$, the restrictions of $\mathbb{P}$, respectively $\mathbb{Q}$, to $\mathcal{F}_t$, then $\frac{\mathrm{d}\mathbb{Q}_t}{\mathrm{d}\mathbb{P}_t} = Z_t$ for all t.*

Proof Notice that there exists a $\mathcal{F}_\infty$-measurable random variable Z_∞ such that $\mathbb{E}_\mathbb{P}[Z_\infty|\mathcal{F}_t] = Z_t$ a.s., for all $t \geq 0$. Simply define $\mathbb{Q}_\infty$ on $\mathcal{F}$ by $\mathbb{Q}_\infty(F) = \mathbb{E}_\mathbb{P} \mathbf{1}_F Z_\infty$. $\mathbb{Q}_\infty$ is a probability measure since $\mathbb{E}_\mathbb{P} Z_\infty = 1$. If F is also in $\mathcal{F}_t$, then we have by the martingale property that $\mathbb{Q}_t(F) = \mathbb{Q}(F) = \mathbb{E}_\mathbb{P} \mathbf{1}_F Z_t$. □

Suppose $\mathbb{Q}$ is a probability measure on $\mathcal{F}$ with $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \zeta$. Then the Z_t in Proposition 9.4 are given by $Z_t = \mathbb{E}[\zeta|\mathcal{F}_t]$ and the Z_∞ in the proof is $\mathbb{E}[\zeta|\mathcal{F}_\infty]$. It follows that $\mathbb{Q}$ and $\mathbb{Q}_\infty$ coincide on $\mathcal{F}_\infty$, but may be different on $\mathcal{F}$.

If we drop the uniform integrability requirement of Z in Proposition 9.4, then the conclusion can not be drawn. There will be a family of probability measures $\mathbb{Q}_t$ on the $\mathcal{F}_t$ that is consistent in the sense that the restriction of $\mathbb{Q}_t$ to $\mathcal{F}_s$ coincides with $\mathbb{Q}_s$ for all $s < t$, but there will in general not exist a probability measure $\mathbb{Q}$ on $\mathcal{F}_\infty$ that is absolutely continuous w.r.t. the restriction of $\mathbb{P}$ to the σ-algebra $\mathcal{F}_\infty$. The best possible general result in that direction is the following.

Proposition 9.5 *Let Ω be the space of real continuous functions on $[0,\infty)$ and let X be the coordinate process ($X_t(\omega) = \omega(t)$) on this space. Let $\mathbb{F} = \mathbb{F}^X$ and let $\mathbb{P}$ be a given probability measure on $\mathcal{F}^X_\infty$. Let Z be a nonnegative martingale with $\mathbb{E}_\mathbb{P} Z_t = 1$. Then there exists a unique probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}^X_\infty)$ such that the restrictions $\mathbb{Q}_t$ of $\mathbb{Q}$ to $\mathcal{F}_t$ are absolutely continuous w.r.t. the restrictions $\mathbb{P}_t$ of $\mathbb{P}$ to $\mathcal{F}_t$ and $\frac{\mathrm{d}\mathbb{Q}_t}{\mathrm{d}\mathbb{P}_t} = Z_t$.*

Proof Let $\mathcal{A}$ be the algebra $\bigcup_{t\geq 0} \mathcal{F}_t$. Observe that one can define a set function $\mathbb{Q}$ on $\mathcal{A}$ unambiguously by $\mathbb{Q}(F) = \mathbb{E}_\mathbb{P} \mathbf{1}_F Z_t$ if $F \in \mathcal{F}_t$. The assertion of the proposition follows from Caratheodory's extension theorem as soon as one has shown that $\mathbb{Q}$ is countably additive on $\mathcal{A}$. We omit the proof. □

Remark 9.6 Notice that in Proposition 9.5 it is not claimed that $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}^X_\infty$. In general this will not happen, see Exercise 9.3.

The next propositions will be crucial in what follows.

Proposition 9.7 *Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$ and $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$. Let Z be their density process and let X be any adapted càdlàg* *process. Then XZ is a martingale under $\mathbb{P}$ iff X is a martingale under $\mathbb{Q}$. If XZ is a local martingale under $\mathbb{P}$ then X is a local martingale under $\mathbb{Q}$. In the latter case equivalence holds under the extra condition that $\mathbb{P}_t \ll \mathbb{Q}_t$ for all $t \geq 0$.*

Proof We prove the 'martingale version' only. Using Lemma 9.2, we have for $t > s$ that

$$\mathbb{E}_{\mathbb{Q}}\left[X_t | \mathcal{F}_s\right] = \frac{\mathbb{E}_{\mathbb{P}}\left[X_t Z_t | \mathcal{F}_s\right]}{Z_s},$$

from which the assertion immediately follows. The 'local martingale' case is left as Exercise 9.4. □

Proposition 9.8 *Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $(\Omega, \mathcal{F})$ and $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$. Let Z be their density process and assume it is right-continuous. Then Z is strictly positive $\mathbb{Q}$-a.s.*

Proof Let $T = \inf\{t \geq 0 : Z_t = 0 \text{ or } Z_{t-} = 0\}$, then $\{\inf_{t \geq 0} Z_t = 0\} = \{T < \infty\}$. Since Z is a $\mathbb{P}$-martingale, we have in view of Proposition 2.10 that $Z_t \mathbf{1}_{\{T \leq t\}} = 0$ ($\mathbb{P}$-a.s.). Then, since $\{T \leq t\} \in \mathcal{F}_t$, we have $\mathbb{Q}(T \leq t) = \mathbb{E}_{\mathbb{P}} Z_t \mathbf{1}_{\{T \leq t\}} = 0$ and hence $\mathbb{Q}(T < \infty) = 0$. □

9.3 The Girsanov theorem

In this section we will explicitly describe how the decomposition of a semimartingale changes under an absolutely continuous change of measure. This is the content of what is known as *Girsanov's theorem*, Theorem 9.9. The standing assumption in this and the next section is that *the density process Z is continuous.*

Theorem 9.9 *Let X be a continuous semimartingale on $(\Omega, \mathcal{F}, \mathbb{P})$ w.r.t. a filtration $\mathbb{F}$ with semimartingale decomposition $X = X_0 + A + M$. Let $\mathbb{Q}$ be another probability measure on $(\Omega, \mathcal{F})$ such that $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$ and with density process Z. Then the stochastic integral $Z^{-1} \cdot \langle Z, M\rangle$ is well defined under $\mathbb{Q}$ and X is also a continuous semimartingale on $(\Omega, \mathcal{F}, \mathbb{Q})$ whose local martingale part $M^{\mathbb{Q}}$ under $\mathbb{Q}$ is given by*

$$M^{\mathbb{Q}} = M - Z^{-1} \cdot \langle Z, M\rangle.$$

Moreover, if X and Y are semimartingales (under $\mathbb{P}$), then their quadratic covariation process $\langle X, Y\rangle$ is the same under $\mathbb{P}$ and $\mathbb{Q}$.

Proof Let $T^n = \inf\{t > 0 : Z_t < 1/n\}$. On each $[0, T^n]$, the stochastic integral $Z^{-1} \cdot \langle Z, M\rangle$ is $\mathbb{P}$-a.s. finite and therefore also $\mathbb{Q}$-a.s. finite. Since $T^n \to \infty$ $\mathbb{Q}$-a.s. (Exercise 9.6), the process $Z^{-1} \cdot \langle Z, M\rangle$ is well defined under $\mathbb{Q}$.

To show that $M^{\mathbb{Q}}$ is a local martingale under $\mathbb{Q}$ we use Proposition 9.7 and Itô's formula for products. We obtain

$$\begin{aligned}(M^{\mathbb{Q}}Z)_t^{T^n} &= \int_0^{T^n\wedge t} M^{\mathbb{Q}}\,\mathrm{d}Z + \int_0^{T^n\wedge t} Z\,\mathrm{d}M^{\mathbb{Q}} + \langle M^{\mathbb{Q}}, Z\rangle_{T^n\wedge t} \\ &= \int_0^{T^n\wedge t} M^{\mathbb{Q}}\,\mathrm{d}Z + \int_0^{T^n\wedge t} Z\,\mathrm{d}M,\end{aligned}$$

where in the last step we used that $Z \times \frac{1}{Z} = 1$ on $[0, T_n]$. This shows that the stopped processes $(M^{\mathbb{Q}}Z)^{T^n}$ are local martingales under $\mathbb{P}$. But then also each $(M^{Q})^{T^n}Z$ is a local martingale under $\mathbb{P}$ (Exercise 9.7) and therefore $(M^{\mathbb{Q}})^{T^n}$ is a local martingale under $\mathbb{Q}$ (use Proposition 9.7). Since $T^n \to \infty$ $\mathbb{Q}$-a.s., $M^{\mathbb{Q}}$ is a local martingale under $\mathbb{Q}$ (Exercise 4.10), which is what we had to prove.

The statement concerning the quadratic covariation process one can prove along the same lines (you show that $(M^{\mathbb{Q}}N^{\mathbb{Q}} - \langle M, N\rangle)Z$ is a local martingale under $\mathbb{P}$, where N and $N^{\mathbb{Q}}$ are the local martingale parts of Y under $\mathbb{P}$ and $\mathbb{Q}$ respectively), or by invoking Proposition 7.4 and by noticing that addition or subtraction of a finite variation process has no influence on the quadratic variation. □

Girsanov's theorem becomes simpler to prove if for all t the measures $\mathbb{P}_t$ and $\mathbb{Q}_t$ are equivalent, in which case the density process is also strictly positive $\mathbb{P}$-a.s. and can be written as a Doléans exponential.

Proposition 9.10 *Let Z be a strictly positive continuous local martingale with $Z_0 = 1$. Then there exists a unique continuous local martingale μ such that $Z = \mathcal{E}(\mu)$.*

Proof Since $Z > 0$ we can define $Y = \log Z$, a semimartingale whose semimartingale decomposition $Y = \mu + A$ satisfies $\mu_0 = A_0 = 0$. We apply the Itô formula to Y and obtain

$$\mathrm{d}Y_t = \frac{1}{Z_t}\,\mathrm{d}Z_t - \frac{1}{2Z_t^2}\,\mathrm{d}\langle Z\rangle_t.$$

Hence

$$\mu_t = \int_0^t \frac{1}{Z_s}\,\mathrm{d}Z_s,$$

and we observe that $A = -\frac{1}{2}\langle\mu\rangle$. Hence $Z = \exp(\mu - \frac{1}{2}\langle\mu\rangle) = \mathcal{E}(\mu)$. Showing the uniqueness is the content of Exercise 9.10. □

Proposition 9.11 *Let μ be a continuous local martingale and let $Z = \mathcal{E}(\mu)$. Let $T > 0$ and assume that $\mathbb{E}_{\mathbb{P}}\, Z_T = 1$. Then Z is a martingale under $\mathbb{P}$ on $[0, T]$. If M is a continuous local martingale under $\mathbb{P}$, then $M^{\mathbb{Q}} := M - \langle M, \mu\rangle$ is a continuous local martingale under the measure $\mathbb{Q}_T$ defined on $(\Omega, \mathcal{F}_T)$ by $\frac{\mathrm{d}\mathbb{Q}_T}{\mathrm{d}\mathbb{P}_T} = Z_T$ with time restricted to $[0, T]$.*

Proof That Z is a martingale follows from Proposition 4.7 (ii) and Exercise 2.1. Then we apply Theorem 9.9 and use that $\langle Z, M\rangle = \int_0^\cdot Z \,\mathrm{d}\langle \mu, M\rangle$ to find the representation for $M^{\mathbb{Q}}$. $\square$

Remark 9.12 In the situation of Proposition 9.11 we have actually $\mathbb{P}_T \sim \mathbb{Q}_T$ and the density $\frac{\mathrm{d}\mathbb{P}_T}{\mathrm{d}\mathbb{Q}_T} = Z_T^{-1}$, alternatively given by $\mathcal{E}(-\mu^{\mathbb{Q}})_T$, with $\mu^{\mathbb{Q}} = \mu - \langle\mu\rangle$, in agreement with the notation of Proposition 9.11. Moreover, if $M^{\mathbb{Q}}$ is a local martingale under $\mathbb{Q}_T$ over $[0,T]$, then $M^{\mathbb{Q}} + \langle M^{\mathbb{Q}}, \mu^{\mathbb{Q}}\rangle = M^{\mathbb{Q}} + \langle M^{\mathbb{Q}}, \mu\rangle$ is a local martingale under $\mathbb{P}$ on $[0,T]$.

Corollary 9.13 *Let W be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ relative to a filtration $\mathbb{F}$. Assume that the conditions of Proposition 9.11 are in force with $\mu = X \cdot W$, where $X \in \mathcal{P}_T^*(W)$. Then the process $W^{\mathbb{Q}} = W - \int_0^\cdot X_s \,\mathrm{d}s$ is a Brownian motion on $[0,T]$ under $\mathbb{Q}_T$ w.r.t. the filtration $\{\mathcal{F}_t, t \in [0,T]\}$.*

Proof We know from Proposition 9.11 that $W^{\mathbb{Q}}$ is a continuous local martingale on $[0,T]$. Since the quadratic variation of this process is the same under $\mathbb{Q}$ as under $\mathbb{P}$ (Theorem 9.9), the process $W^{\mathbb{Q}}$ must be a Brownian motion in view of Lévy's characterization (Proposition 7.13). $\square$

What happens in Corollary 9.13 (and in a wider context a weaker version of this property in Theorem 9.9) is that the change of measure and the shift from W to $W^{\mathbb{Q}}$ compensate each other, the process $W^{\mathbb{Q}}$ has under $\mathbb{Q}$ the same distribution as W under $\mathbb{P}$. This can be seen as the counterpart in continuous time of a similar phenomenon for Gaussian random variables. Suppose that a random variable X has a $N(0,1)$ distribution (a measure on the Borel sets of $\mathbb{R}$), which we call $\mathbb{P}$. Let $\mathbb{Q}$ be the $N(\theta,1)$ distribution as an alternative distribution of X. Under $\mathbb{Q}$ one has that $X^{\mathbb{Q}} := X - \theta$ has a $N(0,1)$ distribution. We see again that the change of distribution from $\mathbb{P}$ to $\mathbb{Q}$ is such that $X^{\mathbb{Q}}$ has the same distribution under $\mathbb{Q}$ as X has under $\mathbb{P}$.

Let p be the density of $\mathbb{P}$ and q the density of $\mathbb{Q}$. In this case one computes the Radon-Nikodym derivative (likelihood ratio)

$$Z_1 := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \frac{q(X)}{p(X)} = \exp(\theta X - \frac{1}{2}\theta^2).$$

Supposing that W is a Brownian motion, we know that W_1 has the $N(0,1)$ distribution. Taking the above X as W_1, we can write

$$Z_1 = \exp(\int_0^1 \theta \,\mathrm{d}W_s - \frac{1}{2}\int_0^1 \theta^2 \,\mathrm{d}s) = \mathcal{E}(\theta \cdot W)_1,$$

which is a special case (at $t = 1$) of the local martingale Z in Proposition 9.11.

Corollary 9.13 only gives us a Brownian motion $W^{\mathbb{Q}}$ under $\mathbb{Q}_T$ on $[0,T]$. Suppose that this would be the case for every T, can we then say that $W^{\mathbb{Q}}$ is a Brownian motion on $[0,\infty)$? For an affirmative answer we would have to extend the family of probability measures $\mathbb{Q}_T$ defined on the $\mathcal{F}_T$ to a probability measure on $\mathcal{F}_\infty$, and as we have mentioned before this is in general impossible. But if we content ourselves with a smaller filtration (much in the spirit of Proposition 9.5), something is possible.

Proposition 9.14 *Let W be a Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathbb{F}^W$ be the filtration generated by W and let X be a process that is progressively measurable w.r.t. $\mathbb{F}^W$ such that $\int_0^T X_s^2 \,\mathrm{d}s < \infty$ a.s. for all $T > 0$. Let $\mu = X \cdot W$, and assume that $Z = \mathcal{E}(\mu)$ is a martingale. Then there exists a probability measure $\mathbb{Q}$ on $\mathcal{F}_\infty^W$ such that $W^{\mathbb{Q}} = W - \int_0^\cdot X_s \,\mathrm{d}s$ is a Brownian motion on $(\Omega, \mathcal{F}_\infty^W, \mathbb{Q})$.*

Proof (sketchy) For all T, the probability measures $\mathbb{Q}_T$ induce probability measures $\mathbb{Q}'_T$ on $(\mathbb{R}^{[0,T]}, \mathcal{B}(\mathbb{R}^{[0,T]}))$, the laws of the Brownian motion $W^{\mathbb{Q}}$ up to the times T. Notice that for $T' > T$ the restriction of $\mathbb{Q}'_{T'}$ to $\mathcal{B}(\mathbb{R}^{[0,T]})$ coincides with $\mathbb{Q}'_T$. It follows that the finite dimensional distributions of $W^{\mathbb{Q}}$ form a consistent family. In view of Kolmogorov's theorem there exists a probability measure $\mathbb{Q}'$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R}^{[0,\infty)}))$ such that the restriction of $\mathbb{Q}'$ to $\mathcal{B}(\mathbb{R}^{[0,T]})$ coincides with $\mathbb{Q}'_T$. A typical set F in $\mathcal{F}_\infty^W$ has the form $F = \{\omega : W_\cdot(\omega) \in F'\}$, for some $F' \in \mathcal{B}(\mathbb{R}^{[0,\infty)})$. Here one needs measurability of the mapping $\omega \mapsto W_\cdot(\omega)$, take this for granted. So we define $\mathbb{Q}(F) = \mathbb{Q}'(F')$ and one can show that this definition is unambiguous. This results in a probability measure on $\mathcal{F}_\infty^W$. If $F \in \mathcal{F}_T^W$, then we have $\mathbb{Q}(F) = \mathbb{Q}'_T(F') = \mathbb{Q}_T(F)$, so this $\mathbb{Q}$ is the one we are after. Observe that $W^{\mathbb{Q}}$ is adapted to $\mathbb{F}^W$ and let $0 \leq t_1 < \cdots < t_n$ be a given arbitrary n-tuple. Then for $B \in \mathcal{B}(\mathbb{R}^n)$ we have $\mathbb{Q}((W_{t_1}^{\mathbb{Q}}, \ldots, W_{t_n}^{\mathbb{Q}}) \in B) = \mathbb{Q}_{t_n}((W_{t_1}^{\mathbb{Q}}, \ldots, W_{t_n}^{\mathbb{Q}}) \in B)$. Since $W^{\mathbb{Q}}$ is Brownian motion on $[0, t_n]$ (Corollary 9.13) the result follows. $\square$

Remark 9.15 Consider the situation of Proposition 9.14 and let $\mathbb{F}^W$ be the filtration generated by W. Let $\mathcal{F}_t = \mathcal{F}_t^W \vee \mathcal{N}_0$ $(t \geq 0)$, where $\mathcal{N}_0$ is the collection of all sets in $\mathcal{F}_\infty^W$ having probability zero under $\mathbb{P}$. Let $\mathbb{Q}$ be the probability measure as in Proposition 9.14 and $\mathbb{Q}_T$ be the probability measure of Corollary 9.13. It is tempting to think that the restriction of $\mathbb{Q}$ to the σ-algebra $\mathcal{F}_T$ coincides with $\mathbb{Q}_T$. This is in general not true. One can find sets F in $\mathcal{F}_T$ for which $\mathbb{P}(F) = 0$ and hence $\mathbb{Q}_T(F) = 0$, although $\mathbb{Q}(F) > 0$ (Exercise 9.9). It similarly follows that the measures $\mathbb{Q}$ and $\mathbb{P}$ as defined on $\mathcal{F}_\infty^W$ may be mutually singular, although for every finite T one has $\mathbb{Q}_T \sim \mathbb{P}_T$.

9.4 The Kazamaki and Novikov conditions

The results in the previous section were based on the fact that the density process Z had the martingale property. In the present section we will see two sufficient conditions in terms of properties of Z that guarantee this. The condition in Proposition 9.16 below is called *Kazamaki's condition*, the one in Proposition 9.18 is known as *Novikov's condition*. Recall that if $Z = \mathcal{E}(\mu)$ for some $\mu \in \mathcal{M}_c^{loc}$, then $Z \in \mathcal{M}_c^{loc}$ as well and by nonnegativity it is a supermartingale, and hence $\mathbb{E}\, Z_t \leq 1$ for all $t \geq 0$. The two propositions that follow guarantee that $\mathbb{E}\, Z_t = 1$, equivalently, that Z is a true martingale.

Proposition 9.16 *Let μ be a local martingale with $\mu_0 = 0$ and suppose that $S = \exp(\frac{1}{2}\mu)$ is a uniformly integrable submartingale. Then $\mathcal{E}(\mu)$ is a uniformly integrable martingale.*

Proof Let $a \in (0, 1)$, put $S(a) = \exp(a\mu/(1 + a))$ and consider $\mathcal{E}(a\mu)$. In what follows we take $a < 1$, note that $\frac{a}{1+a} < \frac{1}{2}$. Verify that $S(a)_t = S_t^{2a/(1+a)}$ and $\mathcal{E}(a\mu) = \mathcal{E}(\mu)^{a^2} S(a)^{1-a^2}$. For any set $F \in \mathcal{F}$ we have by Hölder's inequality for any finite stopping time τ

$$\mathbb{E}\, \mathbf{1}_F \mathcal{E}(a\mu)_\tau \leq (\mathbb{E}\, \mathcal{E}(\mu)_\tau)^{a^2} (\mathbb{E}\, \mathbf{1}_F S(a)_\tau)^{1-a^2}.$$

Since $\mathcal{E}(\mu)$ is a nonnegative local martingale, it is a nonnegative supermartingale (Proposition 4.7 (iii)). Hence $\mathbb{E}\, \mathcal{E}(\mu)_\tau \leq 1$. Together with the easy to prove fact that $\{S(a)_\tau : \tau \text{ a stopping time}\}$ is uniformly integrable, we obtain that $\{\mathcal{E}(a\mu)_\tau : \tau \text{ finite stopping time}\}$ is uniformly integrable too (Exercise 9.11), hence $\mathcal{E}(a\mu)$ belongs to class D. This property combined with $\mathcal{E}(a\mu)$ being a local martingale yields that it is actually a uniformly integrable martingale (see Proposition 4.7). Hence it has a limit $\mathcal{E}(a\mu)_\infty$ with expectation equal to one. Using Hölder's inequality again and noting that $\mathcal{E}(\mu)_\infty$ exists as an a.s. limit of a nonnegative supermartingale, we obtain

$$1 = \mathbb{E}\, \mathcal{E}(a\mu)_\infty \leq \big(\mathbb{E}\, \mathcal{E}(\mu)_\infty\big)^{a^2} \big(\mathbb{E}\, S(a)_\infty\big)^{1-a^2}. \tag{9.2}$$

By uniform integrability of $S = \exp(\frac{1}{2}\mu)$, the expectations $\mathbb{E}\, \exp(\frac{1}{2}\mu_t)$ are bounded. It then follows by Fatou's lemma that $\mathbb{E}\, S_\infty < \infty$. The trivial bound (consider the cases $S(a)_\infty \leq 1$ and $S(a)_\infty > 1$)

$$S(a)_\infty \leq 1 + S_\infty$$

yields

$$\sup_{a<1} \mathbb{E}\, S(a)_\infty < 1 + \mathbb{E}\, S_\infty < \infty,$$

and thus $\limsup_{a \to 1} (\mathbb{E}\, S(a)_\infty)^{1-a^2} = 1$. But then we obtain from (9.2), that $\mathbb{E}\, \mathcal{E}(\mu)_\infty \geq 1$. Using the already known inequality $\mathbb{E}\, \mathcal{E}(\mu)_\infty \leq 1$, we conclude that $\mathbb{E}\, \mathcal{E}(\mu)_\infty = 1$, from which the assertion follows. □

In the proof of Proposition 9.18 we will use the following lemma.

Lemma 9.17 *Let M be a uniformly integrable martingale with the additional property that $\mathbb{E}\, \exp(M_\infty) < \infty$. Then $\exp(M)$ is a uniformly integrable submartingale.*

Proof Exercise 9.12. □

Proposition 9.18 *Let μ be a continuous local martingale with $\mu_0 = 0$ such that*

$$\mathbb{E}\,\exp(\frac{1}{2}\langle\mu\rangle_\infty) < \infty \tag{9.3}$$

Then $\mathcal{E}(\mu)$ is a uniformly integrable martingale and $\mathbb{Q}(F) = \mathbb{E}\,\mathbf{1}_F\mathcal{E}(\mu)_\infty$ defines a probability measure on $\mathcal{F}_\infty$.

Proof First we show that μ is of class D. It follows from the hypothesis and Jensen's inequality that $\mathbb{E}\,\langle\mu\rangle_\infty < \infty$. Let T be a finite stopping time. Then $\mathbb{E}\,\mu_T^2 \leq \mathbb{E}\,\langle\mu\rangle_\infty < \infty$ (see Exercise 4.5). It follows that the collection of these μ_T is uniformly integrable and therefore μ is of Class D and thus even a martingale in view of Proposition 4.7. Consequently μ_∞ exists as an a.s. and L^1-limit.

We apply the Cauchy-Schwarz inequality to

$$\exp(\frac{1}{2}\mu_\infty) = \big(\mathcal{E}(\mu)_\infty\big)^{1/2}\exp(\frac{1}{4}\langle\mu\rangle_\infty)$$

to get

$$\mathbb{E}\,\exp(\frac{1}{2}\mu_\infty) \leq \big(\mathbb{E}\,\mathcal{E}(\mu)_\infty\big)^{1/2}\big(\mathbb{E}\,\exp(\frac{1}{2}\langle\mu\rangle_\infty)\big)^{1/2} \leq \big(\mathbb{E}\,\exp(\frac{1}{2}\langle\mu\rangle_\infty)\big)^{1/2}.$$

Hence $\mathbb{E}\,\exp(\frac{1}{2}\mu_\infty) < \infty$. It follows from Lemma 9.17 that $\exp(\frac{1}{2}\mu)$ is a uniformly integrable submartingale and by Proposition 9.16 we conclude that $\mathcal{E}(\mu)$ is uniformly integrable on $[0, \infty]$. □

Remark 9.19 There also exist versions of Propositions 9.16 and 9.18 for a finite time horizon $T < \infty$. These can easily be obtained from the presented versions by replacing μ with the stopped process μ^T and making the appropriate changes.

9.5 Exercises

9.1 Prove Proposition 9.3.

9.2 Let X be a semimartingale on $(\Omega, \mathcal{F}, \mathbb{P})$. Let Y be a locally bounded process. Show that $Y \cdot X$ is invariant under an absolutely continuous change of measure.

9.3 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measurable space on which is defined a Brownian motion W. Let $\{\mathbb{P}^\theta : \theta \in \mathbb{R}\}$ be a family of probability measures on $\mathcal{F}_\infty^W$ such that for all θ the process W^θ defined by $W_t^\theta = W_t - \theta t$ is a Brownian motion under $\mathbb{P}^\theta$. Show that this family of measures can be defined, and how to construct it. Suppose that we consider θ as an unknown parameter (value) and that we observe X that under each $\mathbb{P}^\theta$ has the semimartingale decomposition $X_t = \theta t + W_t^\theta$. Take $\theta_t = \frac{X_t}{t}$ as estimator of θ when observations up to time t have been made. Show that θ_t is consistent: $\theta_t \to \theta$, $\mathbb{P}^\theta$-a.s. for all θ. Conclude that the $\mathbb{P}^\theta$ are mutually singular on $\mathcal{F}_\infty^W$.

9.4 Finish the proof of Proposition 9.7.

9.5 Let $\mathbb{P}$ and $\mathbb{Q}$ be measure on the space $(\Omega, \mathcal{F})$ with a filtration $\mathbb{F}$ and assume that $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$ with density process Z. Let T be a stopping time. Let $\Omega' = \Omega \cap \{T < \infty\}$ and $\mathcal{F}'_T = \{F \cap \{T < \infty\} : F \in \mathcal{F}_T\}$. Show that with $\mathbb{P}'_T$ the restriction of $\mathbb{P}_T$ to $\mathcal{F}'_T$ (and likewise we have $\mathbb{Q}'_T$) that $\mathbb{Q}'_T \ll \mathbb{P}'_T$ and $\frac{\mathrm{d}\mathbb{Q}'_T}{\mathrm{d}\mathbb{P}'_T} = Z_T$.

9.6 Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on the space $(\Omega, \mathcal{F})$ and assume that $\mathbb{Q}_t \ll \mathbb{P}_t$ for all $t \geq 0$. Let Z be the density process. Let $T^n = \inf\{t : Z_t < \frac{1}{n}\}$. Show that $\mathbb{Q}(T^n < \infty) \leq \frac{1}{n}$ and deduce that $\mathbb{Q}(\inf\{Z_t : t > 0\} = 0) = 0$ and that $\mathbb{Q}(\lim_n T^n = \infty) = 1$.

9.7 Show that the $M^{T^n} Z$ in the proof of Theorem 9.9 are $\mathbb{P}$-local martingales. *Hint: Verify that* $M_t^{T^n} Z_t = M_t^{T^n} Z_t^{T^n} + \int_0^t M_{T^n} \mathbf{1}_{\{T_n < s\}} \, \mathrm{d}Z_s$, with a well defined stochastic integral.

9.8 Prove the claims made in Remark 9.12.

9.9 Consider the situation of Remark 9.15. Consider the density process $Z = \mathcal{E}(W)$. Let $F = \{\lim_{t\to\infty} \frac{W_t}{t} = 1\}$. Use this set to show that the probability measures $\mathbb{Q}_T$ and $\mathbb{Q}$ restricted to $\mathcal{F}_T$ are different.

9.10 Show that uniqueness holds for the local martingale μ of Proposition 9.10.

9.11 Show that the process $S(a)$ in the proof of Proposition 9.16 is uniformly integrable and that consequently the same holds for $\mathcal{E}(a\mu)$.

9.12 Prove Lemma 9.17.

10 Stochastic differential equations

By a *stochastic differential equation* (sde) we mean an equation like

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t \tag{10.1}$$

or

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{10.2}$$

which is sometimes abbreviated by

$$\mathrm{d}X = b(X)\,\mathrm{d}t + \sigma(X)\,\mathrm{d}W.$$

The meaning of (10.1) is nothing else but shorthand notation for the following *stochastic integral equation*

$$X_t = X_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s,\ (t \geq 0). \tag{10.3}$$

Here W is a Brownian motion and b and σ are Borel-measurable functions on $\mathbb{R}^2$ with certain additional requirements to be specified later on. The first integral in (10.3) is a pathwise Lebesgue-Stieltjes integral and the second one a stochastic integral. Of course, both integrals should be well-defined. Examples of stochastic differential equations have already been met in previous sections. For instance, if $X = \mathcal{E}(W)$, then

$$X_t = 1 + \int_0^t X_s\,\mathrm{d}W_s,$$

which is of the above type.

We give an infinitesimal interpretation of the functions b and σ. Suppose that a process X can be represented as in (10.3) and that the stochastic integral is a square integrable martingale. Then we have that $\mathbb{E}\,[X_{t+h} - X_t|\mathcal{F}_t] = \mathbb{E}\,[\int_t^{t+h} b(s, X_s)\,\mathrm{d}s|\mathcal{F}_t]$. For small h, this should 'approximately' be equal to $b(t, X_t)h$. The conditional variance of $X_{t+h} - X_t$ given $\mathcal{F}_t$ is

$$\mathbb{E}\,[(\int_t^{t+h} \sigma(s, X_s)\,\mathrm{d}W_s)^2|\mathcal{F}_t] = \mathbb{E}\,[\int_t^{t+h} \sigma^2(s, X_s)\,\mathrm{d}s|\mathcal{F}_t],$$

which is approximated for small h by $\sigma^2(t, X_t)h$. Hence the coefficient b in Equation (10.1) tells us something about the direction in which X_t changes and σ something about the variance of the displacement. We call b the *drift* coefficient and σ the *diffusion* coefficient.

A process X should be called a solution with initial condition ξ, if it satisfies Equation (10.3) and if $X_0 = \xi$. But this phrase is insufficient for a proper definition of the concept of a solution. In this course we will treat two different concepts, one being *strong solution*, the other one *weak solution*, with the emphasis on the former.

10.1 Strong solutions

In order that the Lebesgue-Stieltjes integral and the stochastic integral in (10.3) exist, we have to impose (technical) regularity conditions on the process X that is involved. Suppose that the process X is defined on some probability space with a filtration. These conditions are (of course) that X is progressive (which is the case if X is adapted and continuous) and that

$$\int_0^t (|b(s, X_s)| + \sigma(s, X_s)^2)\,\mathrm{d}s < \infty, \text{a.s.}\,\forall t \geq 0. \tag{10.4}$$

In all what follows below, we assume that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is such that it supports a Brownian motion W and a random variable ξ that is independent of W. The filtration $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ that we will mainly work with in the present section is the filtration generated by W and ξ and then augmented with the null sets. More precisely. We set $\mathcal{F}_t^\circ = \mathcal{F}_t^W \vee \sigma(\xi)$, $\mathcal{N}$ the $(\mathbb{P}, \mathcal{F}_\infty^\circ)$-null sets and $\mathcal{F}_t = \mathcal{F}_t^\circ \vee \mathcal{N}$. By previous results we know that this filtration is right-continuous and that W is also Brownian w.r.t. it.

Definition 10.1 Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a Brownian motion W defined on it as well as a random variable ξ independent of W, a process X defined on this space is called a *strong solution* of Equation (10.1) with *initial condition* ξ if $X_0 = \xi$ a.s. and X

(i) has continuous paths a.s.
(ii) is $\mathbb{F}$ adapted.
(iii) satisfies Condition (10.4)
(iv) satisfies (10.1) a.s.

The main result of this section concerns existence and uniqueness of a strong solution of a stochastic differential equation.

Theorem 10.2 *Assume that the coefficients b and σ are Lipschitz continuous in the second variable, i.e. there exists a constant $K > 0$ such that*

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|,\ \forall t \in [0, \infty), \forall x, y \in \mathbb{R},$$

and that $b(\cdot, 0)$ and $\sigma(\cdot, 0)$ are locally bounded functions (i.e. they are bounded on compacta). Then, for any initial condition ξ with $\mathbb{E}\,\xi^2 < \infty$ the Equation (10.1) has a unique strong solution. Moreover, $\mathbb{E}\,X_t^2 < \infty$ for all $t \geq 0$.

Proof For given processes X and Y that are such that (10.4) is satisfied for both of them, we define

$$U_t(X) = \xi + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}W_s$$

and $U_t(Y)$ likewise. Note that Equation (10.3), valid for all $t \geq 0$ with $X_0 = \xi$ can now be written as $X = U(X)$ and a solution of this equation can be considered as some kind of a fixed point of U. By U^k we denote the k-fold composition of U. Our aim is to show that the paths of $U^k(\xi)$ converge uniformly in $C[0,T]$ with the sup norm a.s. Thereto we introduce the notation $||Z||_T = \sup\{|Z_t| : t \leq T\}$, for any process Z. *Note that this notation holds for the present proof only.* We first prove that for every $T > 0$ there exists a constant C such that

$$\mathbb{E}\,||U^k(X) - U^k(Y)||_t^2 \leq \frac{C^k t^k}{k!}\mathbb{E}\,||X - Y||_t^2,\ t \leq T. \tag{10.5}$$

Fix $T > 0$. Since for any real numbers p, q it holds that $(p+q)^2 \leq 2(p^2+q^2)$, we obtain for $t \leq T$

$$\begin{aligned}&|U_t(X) - U_t(Y)|^2 \leq \\ &2\Big(\int_0^t |b(s,X_s) - b(s,Y_s)|\,\mathrm{d}s\Big)^2 + 2\Big(\int_0^t (\sigma(s,X_s) - \sigma(s,Y_s))\,\mathrm{d}W_s\Big)^2,\end{aligned}$$

and thus

$$\begin{aligned}&||U(X) - U(Y)||_t^2 \leq \\ &2\Big(\int_0^t |b(s,X_s) - b(s,Y_s)|\,\mathrm{d}s\Big)^2 + 2\sup_{u\leq t}\Big(\int_0^u (\sigma(s,X_s) - \sigma(s,Y_s))\,\mathrm{d}W_s\Big)^2\end{aligned}$$

Take expectations and use the Cauchy-Schwarz inequality as well as a variation on Doob's L^2-inequality (Exercise 10.1), to obtain

$$\begin{aligned}&\mathbb{E}\,||U(X) - U(Y)||_t^2 \leq \\ &2\mathbb{E}\,\Big(t\int_0^t |b(s,X_s) - b(s,Y_s)|^2\,\mathrm{d}s\Big) + 8\mathbb{E}\,\Big(\int_0^t (\sigma(s,X_s) - \sigma(s,Y_s))^2\,\mathrm{d}s\Big).\end{aligned}$$

Now use the Lipschitz condition to get

$$\begin{aligned}\mathbb{E}\,||U(X) - U(Y)||_t^2 &\leq 2\mathbb{E}\,\Big(tK^2\int_0^t (X_s - Y_s)^2\,\mathrm{d}s\Big) \\ &\qquad + 8\mathbb{E}\,\Big(K^2\int_0^t (X_s - Y_s)^2\,\mathrm{d}s\Big) \\ &= 2K^2(t+4)\int_0^t \mathbb{E}\,(X_s - Y_s)^2\,\mathrm{d}s \\ &\leq 2K^2(T+4)\int_0^t \mathbb{E}\,||X - Y||_s^2\,\mathrm{d}s.\end{aligned} \tag{10.6}$$

Let $C = 2K^2(T+4)$. Then we have $\mathbb{E}\,||U(X) - U(Y)||_t^2 \leq Ct\,\mathbb{E}\,||X - Y||_t^2$, which establishes (10.5) for $k = 1$. Suppose that we have proved (10.5) for some

k. Then we use induction to get from (10.6)

$$\begin{aligned}\mathbb{E}\,||U^{k+1}(X)-U^{k+1}(Y)||_t^2 &\le C\int_0^t \mathbb{E}\,||U^k(X)-U^k(Y)||_s^2\,\mathrm{d}s\\ &\le C\int_0^t \frac{C^k s^k}{k!}\,\mathrm{d}s\,\mathbb{E}\,||X-Y||_t^2\\ &= \frac{C^{k+1}t^{k+1}}{(k+1)!}\mathbb{E}\,||X-Y||_t^2.\end{aligned}$$

This proves (10.5). We continue the proof by using Picard iteration, i.e. we are going to define recursively a sequence of processes X^n that will have a limit in a suitable sense on any time interval $[0,T]$ and that limit will be the candidate solution. We set $X^0=\xi$ and define $X^n=U(X^{n-1})=U^n(X^0)$ for $n\ge 1$. Notice that by construction all the X^n have continuous paths and are $\mathbb{F}$-adapted. It follows from Equation (10.5) that we have

$$\mathbb{E}\,||X^{n+1}-X^n||_T^2 \le \frac{C^nT^n}{n!}\mathbb{E}\,||X^1-X^0||_T^2 \tag{10.7}$$

From the conditions on b and σ we can conclude that $B:=\mathbb{E}\,||X^1-X^0||_T^2<\infty$ (Exercise 10.10). By Chebychev's inequality we have that

$$\mathbb{P}(||X^{n+1}-X^n||_T>2^{-n})\le B\frac{(4CT)^n}{n!}.$$

It follows from the Borel-Cantelli lemma that the set $\Omega'=\liminf_{n\to\infty}\{||X^{n+1}-X^n||_T\le 2^{-n}\}$ has probability one. On this set we can find for all ω an integer n big enough such that for all $m\in\mathbb{N}$ one has $||X^{n+m}-X^n||_T(\omega)\le 2^{-n}$. In other words, on Ω' the sample paths of the processes X^n form a Cauchy sequence in $C[0,T]$ w.r.t. the sup-norm, and thus all of them have continuous limits. We call the limit process X (outside Ω' we define it as zero) and show that this process is the solution of (10.1) on $[0,T]$.

Since $X_t^{n+1}=(U(X^n)_t-U(X)_t)+U(X)_t$ for each $t\ge 0$ and certainly $X_t^{n+1}\to X_t$ in probability, it is sufficient to show that for each t we have convergence in probability (or stronger) of $U(X^n)_t-U(X)_t$ to zero. We look at

$$U(X^n)_t-U(X)_t=\int_0^t(b(s,X_s^n)-b(s,X_s))\,\mathrm{d}s \tag{10.8}$$

$$+\int_0^t(\sigma(s,X_s^n)-\sigma(s,X_s))\,\mathrm{d}W_s. \tag{10.9}$$

We first consider the integral of (10.8). Notice that on Ω' we have that $||X-X^n||_T^2(\omega)\to 0$. One thus has (again ω-wise on Ω')

$$\begin{aligned}|\int_0^t(b(s,X_s^n)-b(s,X_s))\,\mathrm{d}s| &\le K\int_0^T|X_s^n-X_s|\,\mathrm{d}s\\ &\le KT||X^n-X||_T\to 0.\end{aligned}$$

Hence we have a.s. convergence to zero of the integral in (10.8). Next we look at the stochastic integral of (10.9). One has

$$\begin{aligned}\mathbb{E}\,\Big(\int_0^t (\sigma(s,X_s)-\sigma(s,X_s^n))\,\mathrm{d}W_s\Big)^2 &= \mathbb{E}\int_0^t (\sigma(s,X_s)-\sigma(s,X_s^n))^2\,\mathrm{d}s\\ &\leq K^2\int_0^t \mathbb{E}\,(X_s-X_s^n)^2\,\mathrm{d}s\\ &\leq K^2T\mathbb{E}\,\|X-X^n\|_T^2\end{aligned}$$

and we therefore show that $\mathbb{E}\,\|X-X^n\|_T^2 \to 0$, which we do as follows.

Realizing that $(\mathbb{E}\,\|X^{n+m}-X^n\|_T^2)^{1/2}$ is the L^2-norm of $\|X^{n+m}-X^n\|_T$, and noting the triangle inequality $\|X^{n+m}-X^n\|_T \leq \sum_{k=n}^{n+m-1}\|X^{k+1}-X^k\|_T$, we apply Minkowski's inequality and obtain

$$(\mathbb{E}\,\|X^{n+m}-X^n\|_T^2)^{1/2} \leq \sum_{k=n}^{n+m-1}(\mathbb{E}\,\|X^{k+1}-X^k\|_T^2)^{1/2}.$$

Then, from (10.7) we obtain

$$\mathbb{E}\,\|X^{n+m}-X^n\|_T^2 \leq \Big(\sum_{k=n}^{\infty}\big(\frac{(CT)^k}{k!}\big)^{1/2}\Big)^2\mathbb{E}\,\|X^1-X^0\|_T^2. \tag{10.10}$$

Combined with the already established a.s. convergence of X^n to X in the sup-norm, we get by application of Fatou's lemma from (10.10) that $\mathbb{E}\,\|X-X^n\|_T^2$ is also bounded from above by the right hand side of this equation and thus converges to zero (it is the tail of a convergent series) for $n\to\infty$.

What is left is the proof of unicity on $[0,T]$. Suppose that we have two solutions X and Y. Then, using the same arguments as those that led us to (10.6), we obtain

$$\begin{aligned}\mathbb{E}\,(X_t-Y_t)^2 &= \mathbb{E}\,(U_t(X)-U_t(Y))^2\\ &\leq C\int_0^t \mathbb{E}\,(X_s-Y_s)^2\,\mathrm{d}s.\end{aligned}$$

It follows from Gronwall's inequality, Exercise 10.9, that $\mathbb{E}\,(X_t-Y_t)^2=0$, for all t and hence, by continuity, X and Y are indistinguishable on $[0,T]$. Extension of the solution X to $[0,\infty)$ is then established by the unicity of solutions on any interval $[0,T]$. □

Theorem 10.2 is a classical result obtained by Itô. Many refinements are possible by relaxing the conditions on b and σ. One that is of particular interest concerns the diffusion coefficient. Lipschitz continuity can be weakened to some variation on Hölder continuity.

Proposition 10.3 *Consider Equation (10.1) and assume that b satisfies the Lipschitz condition of Theorem 10.2, whereas for σ we assume that*

$$|\sigma(t,x) - \sigma(t,y)| \leq h(|x-y|), \tag{10.11}$$

where $h : [0,\infty) \to [0,\infty)$ is a strictly increasing function with $h(0) = 0$ and the property that

$$\int_0^1 \frac{1}{h(u)^2}\,\mathrm{d}u = \infty.$$

Then, given an initial condition ξ, Equation (10.1) admits at most one strong solution.

Proof Take a_n ($n \geq 1$) such that

$$\int_{a_n}^1 \frac{1}{h(u)^2}\,\mathrm{d}u = \frac{1}{2}n(n+1).$$

Then $\int_{a_n}^{a_{n-1}} \frac{1}{nh(u)^2}\,\mathrm{d}u = 1$. We can take a smooth disturbance of the integrand in this integral, a nonnegative continuous function ρ_n with support in (a_n, a_{n-1}) and bounded by $\frac{2}{nh(\cdot)^2}$ such that $\int_{a_n}^{a_{n-1}} \rho_n(u)\,\mathrm{d}u = 1$. We consider *even* functions ψ_n defined for all $x > 0$ by

$$\psi_n(x) = \psi_n(-x) = \int_0^x \int_0^y \rho_n(u)\,\mathrm{d}u\,\mathrm{d}y.$$

Then the sequence (ψ_n) is increasing, $|\psi_n'(x)| \leq 1$ and $\lim_n \psi_n(x) = |x|$, by the definition of the ρ_n. Notice also that the ψ_n are in C^2, since the ρ_n are continuous.

Suppose that there are two strong solutions X and Y with $X_0 = Y_0$. Assume for a while the properties $\mathbb{E}\,|X_t| < \infty$, $\mathbb{E}\int_0^t \sigma(s,X_s)^2\,\mathrm{d}s < \infty$ for all $t > 0$ and the analogous properties for Y. Consider the difference process $V = X - Y$. Apply Itô's rule to $\psi_n(V)$ and take expectations. Then the martingale term vanishes and we are left with (use the above mentioned properties of the ψ_n)

$$\begin{aligned}
\mathbb{E}\,\psi_n(V_t) &= \mathbb{E}\int_0^t \psi_n'(V_s)(b(s,X_s) - b(s,Y_s))\,\mathrm{d}s \\
&\quad + \frac{1}{2}\mathbb{E}\int_0^t \psi_n''(V_s)(\sigma(s,X_s) - \sigma(s,Y_s))^2\,\mathrm{d}s \\
&\leq K\int_0^t \mathbb{E}\,|V_s|\,\mathrm{d}s + \frac{1}{2}\mathbb{E}\int_0^t \psi_n''(V_s)h^2(V_s)\,\mathrm{d}s \\
&\leq K\int_0^t \mathbb{E}\,|V_s|\,\mathrm{d}s + \frac{t}{n}.
\end{aligned}$$

Letting $n \to \infty$, we obtain $\mathbb{E}\,|V_t| \leq K\int_0^t \mathbb{E}\,|V_s|\,\mathrm{d}s$ and the result follows from the Gronwall inequality and sample path continuity under the temporary integrability assumptions on X and Y. As usual, these assumptions can be removed by a localization procedure. In this case one works with the stopping times

$$T^n = \inf\{t > 0 : |X_t| + |Y_t| + \int_0^t (\sigma(s, X_s)^2 + \sigma(s, Y_s)^2)\, \mathrm{d}s > n\}$$

and the stopped processes X^{T^n} and Y^{T^n} have all desired integrability properties. The preceding shows that the stopped processes V^{T^n} are indistinguishable from zero, and the same is then true for V. □

With the techniques in the above proof one can also prove a *comparison result* for solutions to two stochastic differential equations with the same diffusion coefficient.

Proposition 10.4 *Consider the equations*

$$\mathrm{d}X_t = b_j(t, X_t)\, \mathrm{d}t + \sigma(t, X_t)\, \mathrm{d}W_t,\ j = 1, 2,$$

and let X^j be the unique two strong solutions. Assume further that the functions b_j and σ are continuous, b_1 or b_2 satisfies the Lipschitz condition of Theorem 10.2 and the function σ satisfies Condition (10.11). *Assume further*

(i) $X_0^1 \leq X^2$,

(ii) $b_1(t, x) \leq b_2(t, x)$ *for all* $(t, x) \in [0, \infty) \times \mathbb{R}$.

Then $\mathbb{P}(X_t^1 \leq X_t^2, \forall t \geq 0) = 1$.

Proof As in the preceding proof we assume w.l.o.g. that $\mathbb{E} \int_0^t \sigma(s, X_s)^2\, \mathrm{d}s < \infty$ for all $t > 0$. Consider the functions ψ_n as in that proof, and change them into $\phi_n = \psi_n \mathbf{1}_{[0,\infty)}$. Then $\phi_n(x) \to x^+$, $0 \leq \phi_n'(x) = \psi_n'(x)\mathbf{1}_{[0,\infty)}(x) \leq 1$. Instead of the estimate on $\mathbb{E}\, \psi_n(V_t)$, we now find

$$\mathbb{E}\, \phi_n(V_t) \leq \mathbb{E} \int_0^t \phi_n'(V_s)(b_1(s, X_s^1) - b^2(s, X_s^2))\, \mathrm{d}s + \frac{t}{n}.$$

We now make the choice that b_1 satisfies the Lipschitz condition. Then, on the set $\{X_s^1 \geq X_s^2\}$,

$$\begin{aligned} b_1(s, X_s^1) - b^2(s, X_s^2) &= b_1(s, X_s^1) - b_1(s, X_s^2) + b_1(s, X_s^2) - b_2(s, X_s^2) \\ &\leq K(X_s^1 - X_s^2) = KV_s. \end{aligned}$$

It then follows that

$$\mathbb{E}\, \phi_n(V_t) \leq \mathbb{E}\, K \int_0^t \phi_n'(V_s)\mathbf{1}_{[0,\infty)}(V_s)V_s\, \mathrm{d}s + \frac{t}{n}.$$

Letting $n \to \infty$ we obtain

$$\mathbb{E}\, V_t^+ \leq K \int_0^t V_s^+\, \mathrm{d}s,$$

and another application of Gronwall's lemma yields V_t^+ indistinguishable from zero, which we wanted to prove. □

A strong solution X by definition satisfies the property that for each t the random variable X_t is $\mathcal{F}_t$-measurable. Sets in $\mathcal{F}_t$ are typically determined by certain realizations of ξ and the paths of W up to time t. This suggests that there is a causal relationship between X and (ξ and) W, that should formally take the form that $X_t = f(t, \xi, W_{[0,t]})$, where $W_{[0,t]}$ stands for the map that takes the ω's to the paths $\{W_s(\omega), s \leq t\}$. One would like the map f to have the appropriate measurability properties. This can be accomplished when one is working with the *canonical set-up*. By the canonical set-up we mean the following. We take $\Omega = \mathbb{R} \times C[0, \infty)$. Let $\omega = (u, f) \in \Omega$. We define $W_t(\omega) = f(t)$ and $\xi(\omega) = u$. A filtration on Ω is obtained by setting $\mathcal{F}_t^0 = \sigma(\xi, W_s, s \leq t)$. Let μ be a probability measure on $\mathbb{R}$, $\mathbb{P}^W$ be the Wiener measure on $C[0, \infty)$ (the unique probability measure that makes the coordinate process a Brownian motion) and $\mathbb{P} = \mathbb{P}^W \times \mu$. Finally, we get the filtration $\mathbb{F}$ by augmenting the $\mathcal{F}_t^0$ with the $\mathbb{P}$-null sets of $\mathcal{F}_\infty^0$. Next to the filtration $\mathbb{F}$ we also consider the filtration $\mathbb{H}$ on $C[0, \infty)$ that consists of the σ-algebras $\mathcal{H}_t = \sigma(h_s, s \leq t)$, where $h_t(f) = f(t)$. We state the next two theorems without proof, but see also Exercise 10.13 for a simpler version of these theorems.

Theorem 10.5 *Let the canonical set-up be given. Assume that (10.1) has a strong solution with initial condition ξ (so that in particular Condition (10.4) is satisfied). Let μ be the law of ξ. Then there exists a functional $F_\mu : \Omega \to C[0, \infty)$ such that for all $t \geq 0$*

$$F_\mu^{-1}(\mathcal{H}_t) \subset \mathcal{F}_t \tag{10.12}$$

and such that $F_\mu(\xi, W)$ and X are $\mathbb{P}$-indistinguishable. Moreover, if we work on another probability space on which all the relevant processes are defined, a strong solution of (10.1) with an initial condition ξ is again given by $F_\mu(\xi, W)$ with the same functional F_μ as above.

If we put $F(x, f) := F_{\delta_x}(x, f)$ (where δ_x is the Dirac measure at x), we would like to have $X = F(\xi, W)$. There is however in general a measurability problem with the map $(x, f) \mapsto F(x, f)$. By putting restrictions on the coefficients this problem disappears.

Theorem 10.6 *Suppose that b and σ are as in Theorem 10.2. Then a strong solution X may be represented as $X = F(\xi, W)$, where F satisfies the measurability property of (10.12) and moreover, for each $f \in C[0, \infty)$, the map $x \mapsto F(x, f)$ is continuous. Moreover, on any probability space that supports a Brownian motion W and a random variable ξ, a strong solution X is obtained as $X = F(\xi, W)$, with the same mapping F.*

10.2 Weak solutions

Contrary to strong solutions, that have the interpretation of X as an output process of a machine with inputs W and ξ, weak solutions are basically processes defined on a suitable space that can be represented by a stochastic differential equation. This is formalized in the next definition.

Definition 10.7 A *weak solution* of Equation (10.1) by definition consists of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $\mathbb{F}$ and a pair of continuous adapted processes (X, W) such that W is Brownian motion relative to this filtration and moreover

(i) Condition (10.4) is satisfied
(ii) Equation (10.1) is satisfied a.s.

The law of X_0, given a certain weak solution, is called the initial distribution. Notice that it follows from this definition that X_0 and W are independent. Related to weak solutions there are different concepts of uniqueness. For us the most relevant one is *uniqueness in law*, which is defined as follows. Uniqueness in law is said to hold if any two weak solutions $((\Omega, \mathcal{F}, \mathbb{P}), \mathbb{F}, (X, W))$ and $((\Omega', \mathcal{F}', \mathbb{P}'), \mathbb{F}', (X', W'))$ with X_0 and X_0' identically distributed are such that also the processes X and X' have the same distribution.

Next to this there exists the notion of *pathwise uniqueness*. Suppose that the two weak solutions above are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that $W = W'$. It is then said that pathwise uniqueness holds for (weak solutions to) Equation (10.1) if $\mathbb{P}(X_0 = X_0') = 1$ implies $\mathbb{P}(X_t = X_t',\ \forall t \geq 0)$. It is intuitively clear that pathwise uniqueness implies uniqueness in law, but it is far from easy to rigorously prove this. Surprising is the fact that the existence of weak solutions and pathwise uniqueness imply the existence (and uniqueness) of strong solutions. This can be used to show that Equation (10.1) admits existence of a strong solution under the conditions of Proposition 10.3.

The proposition below shows that one can guarantee the existence of weak solutions under weaker conditions than those under which existence of strong solutions can be proved, as should be the case. The main relaxation is that we drop the Lipschitz condition on b.

Proposition 10.8 *Consider the equation*

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \mathrm{d}W_t,\ t \in [0, T].$$

Assume that b satisfies the growth condition $|b(t,x)| \leq K(1 + |x|)$, for all $t \geq 0$, $x \in \mathbb{R}$. Then, for any initial law μ, this equation has a weak solution.

Proof We start out with a measurable space $(\Omega, \mathcal{F})$ and a family of probability measures $\mathbb{P}^x$ on it. Let X be a continuous process on this space such that under each $\mathbb{P}^x$ the process $X - x$ is a standard Brownian motion w.r.t. some filtration $\mathbb{F}$. The $\mathbb{P}^x$ can be chosen to be a *Brownian family*, which entail that the maps $x \mapsto \mathbb{P}^x(F)$ are Borel measurable for each $F \in \mathcal{F}$. Take this for granted. Let $Z_T = \exp(\int_0^T b(s, X_s)\,\mathrm{d}X_s - \frac{1}{2}\int_0^T b(s, X_s)^2\,\mathrm{d}s)$. At the end of the proof we show that $\mathbb{E}_{\mathbb{P}^x} Z_T = 1$ for all x. Assuming that this is the case, we can define a probability measure $\mathbb{Q}^x$ on $\mathcal{F}_T$ by $\mathbb{Q}^x(F) = \mathbb{E}_{\mathbb{P}^x} Z_T \mathbf{1}_F$. Put

$$W_t = X_t - X_0 - \int_0^t b(s, X_s)\,\mathrm{d}s.$$

It follows from Girsanov's theorem that under each of the measures $\mathbb{Q}^x$ the process W is a standard Brownian motion w.r.t. $\{\mathcal{F}_t\}_{t\leq T}$. Define the probability measure $\mathbb{Q}$ by $\mathbb{Q}(F) = \int \mathbb{Q}^x(F)\,\mu(\mathrm{d}x)$. Then W is a Brownian motion under $\mathbb{Q}$ as well (you check why!) and $\mathbb{Q}(X_0 \in B) = \mu(B)$. It follows that the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ together with the filtration $\{\mathcal{F}_t\}_{t\leq T}$ and the processes X and W constitute a weak solution.

We now show that $\mathbb{E}_{\mathbb{P}^x} Z_T = 1$ and we simply write $\mathbb{E}\, Z_T$ for the expectation in the rest of the proof. Our goal is to conclude by application of a variant of Novikov's condition. The idea is to consider the process Z over a number of small time intervals, to check that Novikov's condition is satisfied on all of them and then to collect the conclusions in the appropriate way.

Let $N > 0$ (to be specified later on), put $\delta = T/N$ and consider the processes b^n given by $b^n_t = b(t, X_t)\mathbf{1}_{[(n-1)\delta, n\delta)}(t)$ as well as the associated processes $Z^n = \mathcal{E}(b^n \cdot X)$. Suppose these are all martingales. Then $\mathbb{E}\,[Z^n_{n\delta}|\mathcal{F}_{(n-1)\delta}] = Z^n_{(n-1)\delta}$, which is equal to one. Since we have

$$Z_{N\delta} = \prod_{n=1}^{N} Z^n_{n\delta},$$

we obtain

$$\mathbb{E}\, Z_{N\delta} = \mathbb{E}\,[\prod_{n=1}^{N-1} Z^n_{n\delta}\mathbb{E}\,[Z^N_{N\delta}|\mathcal{F}_{(N-1)\delta}]] = \mathbb{E}\,\prod_{n=1}^{N-1} Z^n_{n\delta} = \mathbb{E}\, Z_{(N-1)\delta},$$

which can be seen to be equal to one by an induction argument. To show that the Z^n are martingales we use Novikov's condition. This condition is

$$\mathbb{E}\,\exp\big(\frac{1}{2}\int_0^T (b^n_t)^2\,\mathrm{d}t\big) = \mathbb{E}\,\exp\big(\frac{1}{2}\int_{(n-1)\delta}^{n\delta} b(t, X_t)^2\,\mathrm{d}t\big) < \infty.$$

By the assumption on b this follows as soon as we know that

$$\mathbb{E}\,\exp\big(\frac{1}{2}\delta K^2(1 + ||X||_T)^2\big) < \infty,$$

where $||X||_T = \sup\{|X_t| : t \leq T\}$, as in the proof of Theorem 10.2. The latter inequality follows from

$$\mathbb{E}\,\exp\big(\delta K^2 ||X||_T^2\big) < \infty, \tag{10.13}$$

which we establish now. Consider the process V given by $V_t = \exp(\frac{1}{2}\delta K^2 X_t^2)$. It follows that $\mathbb{E}\, V_T^2 = \mathbb{E}\,\exp(\delta K^2 X_T^2)$. Since X_T has a normal $N(x, T)$ distribution, the expectation is finite for $\delta < 1/2K^2T$, equivalently $N > 2K^2T^2$. Take such N. Since V is a positive square integrable submartingale on $[0, T]$, we can apply Doob's L^2-inequality to get $\mathbb{E}\,||V||_T^2 \leq 4\mathbb{E}\, V_T^2 = 4\mathbb{E}\,\exp(\delta K^2 X_T^2) < \infty$. Noting that $||V||_T^2 = \exp\big(\delta K^2 ||X||_T^2\big)$, we obtain (10.13). This finishes the proof of $\mathbb{E}\, Z_T = 1$. □

We now give an example of a stochastic differential equation that has a weak solution that is unique in law, but that doesn't admit a strong solution. This equation is

$$X_t = \int_0^t \operatorname{sgn}(X_s)\,\mathrm{d}W_s, \tag{10.14}$$

where

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

First we show that a weak solution exists. Take a $(\Omega, \mathcal{F}, \mathbb{P})$ on which is defined a Brownian motion X and define $W_t = \int_0^t \operatorname{sgn}(X)\,\mathrm{d}X$. It follows from Lévy's characterization that also W is a Brownian motion. Moreover, one easily sees that (10.14) is satisfied. However, invoking Lévy's characterization again, any weak solution to this equation must be a Brownian motion. Hence we have uniqueness in law. Computations show that along with X also $-X$ is a weak solution, whose paths are just the reflected versions of the paths of X. We see that *pathwise uniqueness* doesn't hold, a first indication that existence of a strong solution is problematic. Of course, also $-X$ is a Brownian motion, in agreement with uniqueness in law.

We proceed by showing that assuming existence of a strong solution leads to an absurdity. The proper argument to be used for this involves *local time*, a process that we don't treat in this course. We hope to convince the reader with a heuristic argumentation. For the process W we have

$$W_t = \int_0^t \mathbf{1}_{\{X_s \neq 0\}} \frac{X_s}{|X_s|}\,\mathrm{d}X_s + \int_0^t \mathbf{1}_{\{X_s = 0\}}\,\mathrm{d}X_s.$$

The quadratic variation of the second stochastic integral in the display has expectation equal to $\mathbb{E}\int_0^t \mathbf{1}_{\{X_s=0\}}\,\mathrm{d}s = \int_0^t \mathbb{E}\,\mathbf{1}_{\{X_s=0\}}\,\mathrm{d}s = 0$, in view of properties of Brownian motion. Hence we can ignore the second term and recast the expression for W_t as

$$W_t = \frac{1}{2}\int_0^t \mathbf{1}_{\{X_s \neq 0\}} \frac{1}{|X_s|}\,\mathrm{d}(X_s^2 - s).$$

All the processes involved in the right hand side of this equation are $\mathbb{F}^{|X|}$-adapted, $t \mapsto X_t^2 - t$ is martingale w.r.t. $\mathbb{F}^{|X|}$, and so W is adapted to $\mathbb{F}^{|X|}$ as well. But a strong solution satisfies $\mathcal{F}_t^X \subset \mathcal{F}_t$, which would lead to $\mathcal{F}_t^X \subset \mathcal{F}_t^{|X|}$, an inclusion which is absurd for the Brownian motion X.

If we *change* the definition of sgn into $\operatorname{sgn}(x) = \frac{x}{|x|}\mathbf{1}_{\{x \neq 0\}}$, then with this definition, Equation (10.14) admits a trivial *strong* solution (which one?). The above reasoning also now applies and it follows that there is no other strong solution. Moreover, X a Brownian motion is again a weak solution and uniqueness in law doesn't hold anymore.

10.3 Markov solutions

First we define a *transition kernel*. It is a function $k : \mathbb{R}\times\mathcal{B}(\mathbb{R}) \to [0,1]$ satisfying the properties

$$x \mapsto k(x, A) \text{ is Borel-measurable for all } A \in \mathcal{B}(\mathbb{R})$$

and

$$A \mapsto k(x, A) \text{ is a probability measure for all } x \in \mathbb{R}.$$

Definition 10.9 A process X is called *Markov* w.r.t. a filtration $\mathbb{F}$ if there exists a family of transition kernels $\{P_{t,s} : t \geq s \geq 0\}$ such that for all $t \geq s$ and for all bounded continuous functions f it holds that

$$\mathbb{E}\left[f(X_t)|\mathcal{F}_s\right] = \int f(y) P_{t,s}(X_s, \mathrm{d}y). \tag{10.15}$$

A Markov process X is called homogeneous, if there is a family of transition kernels $\{P_u : u \geq 0\}$ such that for all $t \geq s$ one has $P_{t,s} = P_{t-s}$. Such a process is called strong Markov if for every a.s. finite stopping time T one has

$$\mathbb{E}\left[f(X_{T+t})|\mathcal{F}_T\right] = \int f(y) P_t(X_T, \mathrm{d}y).$$

We are interested in Markov solutions to stochastic differential equations. Since the Markov property (especially if the involved filtration is the one generated by the process under consideration itself) is mainly a property of the distribution of a process, it is natural to consider weak solutions to stochastic differential equations. However, showing (under appropriate conditions) that a solution to a stochastic differential equation enjoys the Markov property is much easier for strong solutions, on which we put the emphasis, and therefore we confine ourselves to this case. The canonical approach to show that weak solutions enjoy the Markov property is via what is known as the *Martingale problem*. For showing that strong solutions have the Markov property, we don't need this concept. The main result of this section is Theorem 10.11 below.

First some additional notation and definitions. If X is a (strong) solution to (10.1), then

$$X_t = \xi + \int_s^t b(u, X_u)\,\mathrm{d}u + \int_s^t \sigma(u, X_u^s)\,\mathrm{d}W_u, \tag{10.16}$$

with $\xi = X_s$. By $X^{s,x} = \{X_t^{s,x} : t \geq s\}$ we denote the unique strong solution (assuming that it exists) to (10.16) with $X_s^{s,x} = x \in \mathbb{R}$. Likewise, $X^{s,\xi}$ denotes the strong solution to (10.16) with (possibly) random initial value $X_s^{s,\xi} = \xi$. Note that a strong solution $X^{s,\xi}$ is adapted to the filtration $\mathbb{F}^s = \{\mathcal{F}_t^s, t \geq s\}$, generated by ξ, the increments of W after time s and augmented with the null sets. More precisely, with $\mathcal{F}_t^{s,\circ} = \sigma(\xi) \vee \sigma(W_u - W_s : s \leq u \leq t)$ and $\mathcal{N}$ the $(\mathbb{P}, \mathcal{F}_\infty^{s,\circ})$-null sets we put $\mathcal{F}_t^s = \mathcal{F}_t^{s,\circ} \vee \mathcal{N}$.

In what follows we will work with a bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ and for such an f we define the function $v_{t,s}^f$ by $v_{t,s}^f(x) = \mathbb{E}\, f(X_t^{s,x})$.

Lemma 10.10 *Assume that the conditions of Theorem 10.2 are satisfied. Let $X^{s,\xi}$ denote the strong solution to (10.16) with initial value $X^{s,\xi}_s = \xi$, where the random variable ξ is $\mathcal{F}_s$-measurable and satisfies $\mathbb{E}\,\xi^2 < \infty$. Then $v^f_{t,s}(\cdot)$ is continuous and $\mathbb{E}\,[f(X^{s,\xi}_t)|\mathcal{F}_s] = v^f_{t,s}(\xi)$.*

Proof Let x be a fixed initial condition and consider $X^{s,x}$. Since $X^{s,x}$ is adapted to the augmented filtration generated by $W_u - W_s$ for $u > s$ it is independent of $\mathcal{F}_s$. Hence $\mathbb{E}\,[f(X^{s,x}_t)|\mathcal{F}_s] = \mathbb{E}\,f(X^{s,x}_t) = v^f_{t,s}(x)$. If ξ assumes countably many values ξ_j, we compute

$$\begin{aligned}
\mathbb{E}\,[f(X^{s,\xi}_t)|\mathcal{F}_s] &= \sum_j \mathbf{1}_{\{\xi=\xi_j\}}\mathbb{E}\,[f(X^{s,\xi_j}_t)|\mathcal{F}_s] \\
&= \sum_j \mathbf{1}_{\{\xi=\xi_j\}} v^f_{t,s}(\xi_j) \\
&= v^f_{t,s}(\xi).
\end{aligned}$$

The general case follows by approximation. For arbitrary ξ we consider for every n a countably valued random variable ξ^n, such that $\xi^n \to \xi$. One can take for instance $\xi^n = 2^{-n}[2^n\xi]$, in which case $\xi^n \leq \xi \leq \xi^n + 2^{-n}$. We compute, using Jensen's inequality for conditional expectations,

$$\begin{aligned}
&\mathbb{E}\,\big(\mathbb{E}\,[f(X^{s,\xi^n}_t)|\mathcal{F}_s] - \mathbb{E}\,[f(X^{s,\xi}_t)|\mathcal{F}_s]\big)^2 \\
&= \mathbb{E}\,\big(\mathbb{E}\,[f(X^{s,\xi^n}_t) - f(X^{s,\xi}_t)|\mathcal{F}_s]\big)^2 \\
&\leq \mathbb{E}\,\mathbb{E}\,[(f(X^{s,\xi^n}_t) - f(X^{s,\xi}_t))^2|\mathcal{F}_s] \\
&= \mathbb{E}\,(f(X^{s,\xi^n}_t) - f(X^{s,\xi}_t))^2. \qquad (10.17)
\end{aligned}$$

Now we apply Exercise 10.14, that tells us that $\mathbb{E}\,(X^{s,\xi^n}_t - X^{s,\xi}_t)^2 \to 0$ for $n \to \infty$. But then we also have the convergence in probability and since f is bounded and continuous also the expression in (10.17) tends to zero. Hence we have L^2-convergence of $\mathbb{E}\,[f(X^{s,\xi^n}_t)|\mathcal{F}_s]$ to $\mathbb{E}\,[f(X^{s,\xi}_t)|\mathcal{F}_s]$. The above reasoning also applies to a deterministic sequence (x^n) with $x^n \to x$, and we obtain continuity of the function $v^f_{t,s}(\cdot)$. From the L^2-convergence we obtain a.s. convergence along a suitably chosen subsequence. Recall that we already proved that

$$v^f_{t,s}(\xi^n) = \mathbb{E}\,[f(X^{s,\xi^n}_t)|\mathcal{F}_s].$$

Apply now the continuity result to the left hand side of this equation and the a.s. convergence (along some subsequence) to the right hand side to arrive at the desired conclusion. □

Theorem 10.11 *Let the coefficients b and σ satisfy the conditions of Theorem 10.2. Then the solution process is Markov. Under the additional assumption that the coefficients b and σ are functions of the space variable only, the solution process is strong Markov.*

Proof Lemma 10.10 has as a corollary that $\mathbb{E}\,[f(X_t)|\mathcal{F}_s] = v^f_{t,s}(X_s)$ for $t \geq s$, for all bounded and continuous f, since $X_t = X_t^{s,X_s}$. Applying this to functions f of the form $f(x) = \exp(\mathrm{i}\lambda x)$, we see that the conditional characteristic function of X_t given $\mathcal{F}_s$ is a measurable function of X_s, as $v^f_{t,s}(\cdot)$ is continuous. It then follows that for $t > s$ and Borel sets A the conditional probabilities $\mathbb{P}(X_t \in A|\mathcal{F}_s)$ are of the form $P_{t,s}(X_s, A)$, where the functions $\mathbb{P}_{t,s}(\cdot,\cdot)$ are transition kernels. But then $\int f(y)P_{t,s}(X_s, dy) = v^f_{t,s}(X_s) = \mathbb{E}\,[f(X_t)|\mathcal{F}_s]$. Hence X is Markov.

To prove the strong Markov property for time homogeneous coefficients we follow a similar procedure. If T is an a.s. finite stopping time, we have instead of Equation (10.16)

$$X_{T+t} = X_T + \int_T^{T+t} b(X_u)\,\mathrm{d}u + \int_T^{T+t} \sigma(X_u)\,\mathrm{d}W_u.$$

By the strong Markov property of Brownian motion, the process $u \mapsto W_{T+u} - W_T$, $u \geq 0$, is a Brownian motion independent of $\mathcal{F}_T$. The function $v^f_{t,s}(\cdot)$ introduced above now only depends on the time difference $t - s$ and we write $v^f_{t-s}(x)$ instead of $v^f_{t,s}(x)$. Hence we can copy the above analysis to arrive at $\mathbb{E}\,[f(X_{T+t})|\mathcal{F}_T] = v^f_t(X_T)$, from which the strong Markov property follows. $\square$

10.4 Exercises

10.1 Prove that for a continuous local martingale M it holds that

$$\mathbb{E}\,(\sup_{s\leq t}|M_s|)^2 \leq 4\mathbb{E}\,\langle M\rangle_t.$$

10.2 Let $X_0, \varepsilon_1, \varepsilon_2, \ldots$ be a sequence of independent random variables. Suppose that we generate a (discrete time) random process by the recursion

$$X_t = f(X_{t-1}, \varepsilon_t, t),\ (t \geq 1),$$

where f is a measurable function. Show that the process X is Markov: $\mathbb{P}(X_{t+1} \in B|\mathcal{F}^X_t) = \mathbb{P}(X_{t+1} \in B|X_t)$. Show also the stronger statement: for any bounded and measurable function h we have

$$\mathbb{E}\,[h(X_{t+1})|\mathcal{F}^X_t] = \int h(f(X_t, u, t+1))\,F_{t+1}(\mathrm{d}u),$$

where F_{t+1} is the distribution function of ε_{t+1}.

10.3 Consider the stochastic differential equation

$$\mathrm{d}X_t = 3X_t^{1/3}\,\mathrm{d}t + 3X_t^{2/3}\,\mathrm{d}W_t,$$

with initial condition $X_0 = 0$. Give two different strong solutions to this equation.

10.4 Consider the one-dimensional equation

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t,\; X_0,$$

where a and σ are real constants and X_0 and W are independent.

(a) Show that a strong solution is given by $X_t = e^{-at}X_0 + \sigma e^{-at}\int_0^t e^{as}\,\mathrm{d}W_s$.

(b) Let X_0 have mean μ_0 and variance $\sigma_0^2 < \infty$. Compute $\mu_t = \mathbb{E}\,X_t$ and $\sigma_t^2 = \mathrm{Var}\,X_t$.

(c) If X_0 moreover has a normal distribution, show that all X_t have a normal distribution as well.

(d) X_0 is said to have the invariant distribution if all X_t have the same distribution as X_0. Find this distribution (for existence you also need a condition on a).

10.5 Let X be defined by $X_t = x_0 \exp\big((b-\frac{1}{2}\sigma^2)t+\sigma W_t\big)$, where W is a Brownian motion and b and σ are constants. Write down a stochastic differential equation that X satisfies. Having found this sde, you make in the sde the coefficients b and σ depending on time. How does the solution of this equation look now?

10.6 Consider the equation

$$\mathrm{d}X_t = (\theta - aX_t)\,\mathrm{d}t + \sigma\sqrt{X_t \vee 0}\,\mathrm{d}W_t,\; X_0 = x_0 \geq 0. \tag{10.18}$$

Assume that $\theta \geq 0$. It admits a unique strong solution X. Show that it is nonnegative for all $t \geq 0$ (and hence $\mathrm{d}X_t = (\theta - aX_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t$). The equation is known as the Cox-Ingersoll-Ross model, used in interest rate theory.

10.7 Let W^i, $i = 1,\ldots,d$ be independent Brownian motions and let $\mathrm{d}X_t^i = -aX_t^i\,\mathrm{d}t + \rho\,\mathrm{d}W_t^i$. Put $X_t = \sum_{i=1}^d (X_t^i)^2$ and note that $X_t \geq 0$.

(a) Show that $\mathrm{d}X_t = (d\rho^2 - 2aX_t)\,\mathrm{d}t + 2\rho\Sigma_{i=1}^d X_t^i\,\mathrm{d}W_t^i$.

(b) Put $\mathrm{d}W_t = X_t^{-1/2}\sum_{i=1}^d X_t^i\,\mathrm{d}W_t^i$, $W_0 = 0$. Show that W is a Brownian motion.

(c) Show that X satisfies Equation (10.18) by suitably choosing the parameters.

10.8 Consider again Equation (10.18) and *assume* $2\theta \geq \sigma^2 > 0$ *and* $x_0 > 0$. In this exercise you show that $\mathbb{P}(X_t > 0,\, \forall t \geq 0) = 1$. Define for $x \in (0,\infty)$ (what is called the *scale function* for this SDE)

$$s(x) = \int_1^x \exp(2ay/\sigma^2)y^{-2\theta/\sigma^2}\,\mathrm{d}y.$$

Note that $s(x) < 0$ if $x \in (0,1)$ and that $\lim_{x\to 0} s(x) = -\infty$. Below we also need the stopping times $T^x = \inf\{t \geq 0 : X_t = x\}$ for $x = 0$, $x = \varepsilon > 0$ and $x = M > 0$, where $\varepsilon < x_0 < M$, and $T = T^\varepsilon \wedge T^M$.

(a) Show that $(\theta - bx)s'(x) + \frac{1}{2}\sigma^2 x s''(x) = 0$ and that

$$s(X_t^T) = s(x_0) + \sigma\int_0^{T\wedge t} s'(X_u)\sqrt{X_u}\,\mathrm{d}W_u.$$

(b) Take expectations of the squares in the above display and conclude that $\mathbb{E}\,T < \infty$. *Hint:* $s(X_t^T)$ *is bounded from above and* $s'(X_t^T)$ *is bounded from below.*

(c) Show that $s(x_0) = s(\varepsilon)\mathbb{P}(T^\varepsilon < T^M) + s(M)\mathbb{P}(T^\varepsilon > T^M)$.

(d) Deduce that $\mathbb{P}(T^0 < T^M) = 0$ for all large M and that $\mathbb{P}(T^0 < \infty) = 0$.

10.9 Let g be a nonnegative Borel-measurable function, that is locally integrable on $[0,\infty)$. Assume that g satisfies for all $t \geq 0$ the inequality $g(t) \leq a + b\int_0^t g(s)\,\mathrm{d}s$, where $a, b \geq 0$.

(a) Show that $g(t) \leq ae^{bt}$. *Hint (not obliged to follow):* Solve the inhomogeneous integral equation

$$g(t) = a + b\int_0^t g(s)\,\mathrm{d}s - p(t)$$

for a nonnegative function p.

(b) Give an example of a discontinuous g that satisfies the above inequalities.

10.10 Show that (cf. the proof of Theorem 10.2) $\mathbb{E}\,||X^1 - X^0||_T^2 < \infty$.

10.11 Let $T > 0$. Show that under the assumptions of Theorem 10.2 it holds that for all $T > 0$ there is a constant C (depending on T) such that $\mathbb{E}\,X_t^2 \leq C(1 + \mathbb{E}\,\xi^2)\exp(Ct)$, for all $t \leq T$.

10.12 Endow $C[0,\infty)$ with the metric d defined by

$$d(x,y) = \sum_{n=1}^{\infty} 2^{-n}(1 \wedge \max_{t\in[0,n]} |x(t) - y(t)|).$$

Show that the Borel σ-algebra $\mathcal{B}(C[0,\infty))$ coincides with the smallest σ-algebra that makes all finite dimensional projections measurable.

10.13 Let X be strong solution to (10.1) with initial value ξ. Show that for each $t > 0$ there is a map $f : \mathbb{R} \times C[0,t] \to \mathbb{R}$ that is $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(C[0,t])/\mathcal{B}(\mathbb{R})$-measurable such that $X_t(\omega) = f(\xi(\omega), W_{[0,t]}(\omega))$ for almost all ω.

10.14 Let X and Y be strong solutions with possibly different initial values and assume that the conditions of Theorem 10.2 are in force. Show that for all T there is a constant D such that (notation as in the proof of the theorem)

$$\mathbb{E}\,||X - Y||_t^2 \leq D\,\mathbb{E}\,|X_0 - Y_0|^2, \forall t \leq T.$$

Hint: Look at the first part of the proof of Theorem 10.2. First you use that $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, mimic the proof and finally you use Gronwall's inequality.

10.15 Show that under the assumptions of Theorem 10.2 also Equation (10.16) admits a unique strong solution. What can you say about this when we replace s with a finite stopping time T?

10.16 If X is a strong solution of (10.1) and the assumptions of Theorem 10.2 are in force, then the bivariate process $\{(X_t, t), t \geq 0\}$ is strong Markov. Show this.

10.17 If X is a (weak or strong) solution to (10.1) with b and σ locally bounded measurable functions, then for all $f \in C^2(\mathbb{R})$, the process $M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{L}_s f(X_s)\, \mathrm{d}s$ is a local martingale. Here the operators $\mathcal{L}_s$ are defined by $\mathcal{L}_s f(x) = b(s,x)f'(x) + \frac{1}{2}\sigma^2(s,x)f''(x)$ for $f \in C^2(\mathbb{R})$. If we restrict f to be a C_K^∞-function, then the M^f become martingales. Show these statements.

10.18 Consider the equation

$$X_t = 1 + \int_0^t X_s \, \mathrm{d}W_s.$$

(a) Apply the Picard-Lindelöf iteration scheme of the proof of Theorem 10.2 (so $X_t^0 \equiv 1$, etc.).

$$X_t^n = \sum_{k=0}^n \frac{1}{k!} H_k(W_t, t),$$

where the functions H_k are those of Exercise 7.1.

(b) Conclude *by a direct argument* that $X_t^n \to \mathcal{E}(W)_t$ a.s. for $n \to \infty$, for every $t \geq 0$.

(c) Do we also have a.s. convergence, uniform over compact time intervals?

10.19 Let X be a Markov process with associated transition kernels $P_{t,s}$ as in (10.15). Show the validity of the Chapman-Kolmogorov equations

$$P_{u,s}(x, A) = \int P_{u,t}(y, A) P_{t,s}(x, \mathrm{d}y),$$

valid for all Borel sets A and $s \leq t \leq u$.

10.20 The use of conditional characteristic functions in the proof of Theorem 10.11 can be avoided. Show first that it follows from Lemma 10.10 that for $x \in \mathbb{R}$ the conditional probabilities $\mathbb{P}(X_t \leq x|\mathcal{F}_s)$ are measurable functions of X_s and deduce from this that the same holds for conditional probabilities $\mathbb{P}(X_t \in A|\mathcal{F}_s)$, for Borel sets A.

10.21 Consider the SDE (10.1) and assume that b and σ are bounded. Let $X^{t,x}$ be the unique weak solution that starts at t in x. Let $T > t$ and assume that $X_T^{t,x}$ has a density denoted $p(t,T;x,\cdot)$ which possesses all the required differentiability properties needed below. Furthermore we consider functions $h \in C^2(\mathbb{R})$ with bounded support, contained in some (a,b).

(a) Use the Itô formula to give a semimartingale representation for $X_T^{t,x}$ (the integrals are of the type $\int_t^T$).

(b) Show, by taking expectations and by giving some additional arguments that

$$\int_a^b h(y)p(t,T;x,y)\,\mathrm{d}y = h(x) + \int_t^T \int_a^b h(y)\mathcal{L}_t^* p(t,u;x,y)\,\mathrm{d}y\,\mathrm{d}u,$$

where

$$\mathcal{L}_t^* f(y) = -\frac{\partial}{\partial y}\big(b(t,y)f(y)\big) + \frac{\partial^2}{\partial y^2}\big(\sigma(t,y)^2 f(y)\big),$$

and in $\mathcal{L}_t^* p(t,u;x,y)$ the t,u,x are considered as 'parameters'.

(c) Show that $\int_a^b h(y)\big(\frac{\partial}{\partial T}p(t,T;x,y) - \mathcal{L}_t^* p(t,T;x,y)\big)\,\mathrm{d}y$.

(d) Conclude that $\frac{\partial}{\partial T}p(t,T;x,y) = \mathcal{L}_t^* p(t,u;x,y)$.

The last equation is known as Kolmogorov's forward equation and the operator $\mathcal{L}_t^*$ is the formal adjoint *of the operator $\mathcal{L}_t$, the generator, that we will encounter in* (11.2). *So, with $\langle\cdot,\cdot\rangle$ the usual inner product in $L^2(\mathbb{R})$, it holds that $\langle \mathcal{L}_t^* f, g\rangle = \langle g, \mathcal{L}_t g\rangle$ for all f and g that are in $C^2(\mathbb{R})$ and have compact support.*

11 Partial differential equations

The simplest example of a partial differential equation whose solutions can be expressed in terms of a diffusion process is the *heat equation*

$$u_t - \frac{1}{2}u_{xx} = 0. \tag{11.1}$$

The fundamental solution of this equation for $t > 0$ and $x \in \mathbb{R}$ is the density of the $N(0,t)$ distribution,

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

This solution is clearly not unique, $u(t,x) = t + x^2$ is another one. For uniqueness one needs e.g. an initial value, $u(0,x) = f(x)$ say, with some pre-specified function f. Leaving technicalities aside for the moment, one may check by scrupulously interchanging differentiation and integration that

$$u(t,x) = \int_{\mathbb{R}} f(y)p(t,x-y)\,\mathrm{d}y$$

satisfies Equation (11.1) for $t > 0$. Moreover, we can write this function u as $u(t,x) = \mathbb{E}\,f(x + W_t)$, where we extend the domain of the t-variable to $t \geq 0$. Notice that for $t = 0$ we get $u(0,x) = f(x)$. Under appropriate conditions on f one can show that this gives the unique solution to (11.1) with the initial condition $u(0,x) = f(x)$.

If we fix a terminal time T, we can define $v(t,x) = u(T-t,x)$ for $t \in [0,T]$, with u a solution of the heat equation. Then v satisfies the *backward heat equation* equation

$$v_t + \frac{1}{2}v_{xx} = 0,$$

with terminal condition $v(T,x) = f(x)$ if $u(0,x) = f(x)$. It follows that we have the representation $v(t,x) = \mathbb{E}\,f(x + W_{T-t}) = \mathbb{E}\,f(x + W_T - W_t)$. Denote by $X^{t,x}$ a process defined on $[t,\infty)$ that starts at t in x and whose increments have the same distribution as those of Brownian motion. Then we can identify this process as $X^{t,x}_s = x + W_s - W_t$ for $s \geq t$. Hence we have $v(t,x) = \mathbb{E}\,f(X^{t,x}_T)$. The notation is the similar to the one in Section 10.3, note that $X^{t,x}_s = x + \int_t^s \mathrm{d}W_u$.

In this section we will look at partial differential equations that are more general than the heat equation, with initial conditions replaced by a terminal condition. The main result is that solutions to such equations can be represented as functionals of solutions to stochastic differential equations.

11.1 Feynman-Kaç formula

Our starting point is the stochastic differential Equation (10.1). Throughout this section we assume that the coefficients b and σ are continuous and that the linear growth condition

$$|b(t,x)| + |\sigma(t,x)| \leq K(1+|x|)$$

is satisfied. Moreover we assume that (10.1) allows for each pair (t,x) a weak solution involving a process $(X_s^{t,x})_{s \geq t}$ that is unique in law and satisfies $X_t^{t,x} = x$ a.s. We also need the family of operators $\{\mathcal{L}_t, t > 0\}$ (called *generator* in Markov process language) acting on functions $f \in C^{1,2}([0,\infty) \times \mathbb{R})$ defined by

$$\mathcal{L}_t f(t,x) = b(t,x) f_x(t,x) + \frac{1}{2}\sigma^2(t,x) f_{xx}(t,x). \tag{11.2}$$

The main result of this section is Theorem 11.2 below, in the proof of which we need the following lemma on moments of a solution to a stochastic differential equation.

Lemma 11.1 *Let X be (part of) a weak solution of (10.1). Then for any finite time T and $p \geq 2$ there is a constant C such that*

$$\mathbb{E} \sup_{t \leq T} |X_t|^p \leq C e^{CT}(1 + \mathbb{E}\,|X_0|^p).$$

Proof The proof is based on an application of the Burkholder-Davis-Gundy inequality of Proposition 7.14. See Exercise 11.3. □

We now consider the *Cauchy problem.* Let $T > 0$ and let functions $f : \mathbb{R} \to \mathbb{R}$, $g, k : [0,T] \times \mathbb{R} \to \mathbb{R}$ be given. Find a (unique) solution $v : [0,T] \times \mathbb{R}$ that belongs to $C^{1,2}([0,T) \times \mathbb{R})$ and that is continuous on $[0,T] \times \mathbb{R}$ such that

$$v_t + \mathcal{L}_t v = kv - g, \tag{11.3}$$

and

$$v(T,\cdot) = f. \tag{11.4}$$

Next *assumptions* are imposed on f, g and k. They are all continuous on their domain, k is nonnegative and f, g satisfy the following growth condition. There exist constants $L > 0$ and $\lambda \geq 1$ such that

$$|f(x)| + \sup_{0 \leq t \leq T} |g(t,x)| \leq L(1 + |x|^{2\lambda}) \tag{11.5}$$

Theorem 11.2 *Let under the stated assumptions the Equation (11.3) with terminal Condition (11.4) have a solution v satisfying the growth condition*

$$\sup_{0\le t\le T} |v(t,x)| \le M(1+|x|^{2\mu}),$$

for some $M > 0$ and $\mu \ge 1$. Let $X^{t,x} = (X^{t,x}_s,\, s \ge t)$ be the weak solution to (10.1) starting at t in x that is unique in law. Then v admits the stochastic representation (Feynman-Kaç formula)

$$\begin{aligned} v(t,x) &= \mathbb{E}\,[f(X^{t,x}_T)\exp(-\int_t^T k(u,X^{t,x}_u)\,\mathrm{d}u)] \\ &\quad + \mathbb{E}\,[\int_t^T g(r,X^{t,x}_r)\exp(-\int_t^r k(u,X^{t,x}_u)\,\mathrm{d}u)\,\mathrm{d}r]. \end{aligned} \tag{11.6}$$

on $[0,T]\times\mathbb{R}$ and is thus unique.

Proof In the proof we simply write X instead of $X^{t,x}$. Let

$$Y_s = v(s,X_s)\exp(-\int_t^s k(u,X_u)\,\mathrm{d}u) \text{ for } s \ge t.$$

An application of Itô's formula combined with the fact that v solves (11.3) yields

$$\begin{aligned} Y_s - Y_t &= \int_t^s v_x(r,X_r)\sigma(r,X_r)\exp(-\int_t^r k(u,X_u)\,\mathrm{d}u)\,\mathrm{d}W_r \\ &\quad - \int_t^s g(r,X_r)\exp(-\int_t^r k(u,X_u)\,\mathrm{d}u)\,\mathrm{d}r. \end{aligned} \tag{11.7}$$

Notice that $Y_t = v(t,x)$ and that $Y_T = f(X_T)\exp(-\int_t^T k(u,X_u)\,\mathrm{d}u)$. If the stochastic integral in (11.7) would be a martingale, then taking expectations for $s = T$ would yield the desired result. Since this property is not directly guaranteed under the prevailing assumptions, we will reach our goal by stopping. Let $T^n = \inf\{s \ge t : |X_s| \ge n\}$. Consider the stochastic integral in (11.7) at $s = T\wedge T^n$. It can be written as

$$\int_t^T \mathbf{1}_{\{r\le T^n\}} v_x(r,X^{T^n}_r)\sigma(r,X^{T^n}_r)\exp(-\int_t^r k(u,X^{T^n}_u)\,\mathrm{d}u)\,\mathrm{d}W_r.$$

Since v_x and σ are bounded on compact sets, $|X^{T^n}|$ is bounded by n and $k \ge 0$, the integrand in the above stochastic integral is bounded and therefore the stochastic integral has zero expectation. Therefore, if we evaluate (11.7) at $s = T\wedge T^n$ and take expectations we obtain

$$\mathbb{E}\,Y_{T\wedge T^n} - v(t,x) = -\mathbb{E}\int_t^{T\wedge T^n} g(r,X_r)\exp(-\int_t^r k(u,X_u)\,\mathrm{d}u)\,\mathrm{d}r. \tag{11.8}$$

Consider first the left hand side of (11.8). It can be written as the sum of

$$\mathbb{E}\left(f(X_T)\exp(-\int_t^T k(u,X_u)\,\mathrm{d}u)\mathbf{1}_{\{T\leq T^n\}}\right) \tag{11.9}$$

and

$$\mathbb{E}\left(v(T^n,X_{T^n})\exp(-\int_t^{T^n} k(u,X_u)\,\mathrm{d}u)\mathbf{1}_{\{T>T^n\}}\right). \tag{11.10}$$

The expression (11.9) is bounded in absolute value by $L\mathbb{E}\,(1+|X_T|^{2\lambda})$in view of (11.5). It holds that $\mathbb{E}\,|X_T|^{2\lambda} \leq Ce^{C(T-t)}(1+|x|^{2\lambda}) < \infty$ in view of Lemma 11.1, since we start the diffusion at t in x. Hence we can apply the dominated convergence theorem to show that the limit of (11.9) for $n\to\infty$ is equal to

$$\mathbb{E}\,f(X_T)\exp(-\int_t^T k(u,X_u)\,\mathrm{d}u).$$

The absolute value of (11.10) is bounded from above by

$$M(1+n^{2\mu})\mathbb{P}(T^n\leq T). \tag{11.11}$$

Now,

$$\mathbb{P}(T^n\leq T)=\mathbb{P}(\sup_{t\leq s\leq T}|X_s|\geq n)\leq n^{-2p}\mathbb{E}\sup_{t\leq s\leq T}|X_s|^{2p}.$$

The expectation here is in view of Lemma 11.1 less than or equal to $C(1+|x|^{2p})e^{C(T-t)}$. Hence we can bound (11.11) from above by $M(1+n^{2\mu})n^{-2p}C(1+|x|^{2p})e^{C(T-t)}$. By choosing $p>\mu$, we see that the contribution of (11.10) vanishes for $n\to\infty$. Summing up,

$$\mathbb{E}\,Y_{T\wedge T^n}\to\mathbb{E}\,f(X_T)\exp(-\int_t^T k(u,X_u)\,\mathrm{d}u).$$

Next we turn to the right hand side of (11.8). Write it as

$$-\mathbb{E}\int_t^T \mathbf{1}_{\{r\leq T^n\}}g(r,X_r)\exp(-\int_t^r k(u,X_u)\,\mathrm{d}u)\,\mathrm{d}r.$$

The absolute value of the integrand we can bound by $L(1+|X_r|^{2\lambda})$, whose expectation is finite by another application of Lemma 11.1. Hence the dominated convergence theorem yields the result. □

Remark 11.3 The nonnegative function k that appears in the Cauchy problem is connected with *exponential killing*. Suppose that we have a process $X^{t,x}$ starting at time t in x and that there is an independent random variable Y with a standard exponential distribution. Let $\partial \notin \mathbb{R}$ be a so-called *coffin* or *cemetery state*. Then we define the process $X^{\partial,t,x}$ by ($s \geq t$)

$$X_s^{\partial,t,x} = \begin{cases} X_s^{t,x} & \text{if } \int_t^s k(u, X_u^{t,x})\,\mathrm{d}u < Y \\ \partial & \text{if } \int_t^s k(u, X_u^{t,x})\,\mathrm{d}u \geq Y \end{cases}$$

Functions f defined on $\mathbb{R}$ will be extended to $\mathbb{R} \cup \{\partial\}$ by setting $f(\partial) = 0$. If $X^{t,x}$ is a Markov process (solving Equation (10.1) for instance), then $X^{\partial,t,x}$ is a Markov process as well. If, in the terminology of the theory of Markov processes, $X^{t,x}$ has generator $\mathcal{L}$, then $X^{\partial,t,x}$ has generator $\mathcal{L}^k$ defined by $\mathcal{L}_t^k f(t,x) = \mathcal{L}_t f(t,x) - k(t,x)f(t,x)$. Furthermore we have

$$\mathbb{E}\, f(X_s^{\partial,t,x}) = \mathbb{E}\, f(X_s^{t,x}) \exp(-\int_t^u k(u, X_u^{t,x})\,\mathrm{d}u.$$

The above considerations enable one to connect the theory of solving the Cauchy problem with $k = 0$ to solving the problem with arbitrary k by jumping in the representation from the process $X^{t,x}$ to $X^{\partial,t,x}$.

11.2 Exercises

11.1 Show that the growth conditions on f and g are not needed in order to prove Theorem 11.2, if we assume instead that next to k also f and g are nonnegative. *It is sufficient to point out how to modify the proof of Theorem 11.2.*

11.2 Consider Equation (11.3) with $k = 0$ and $g = 0$. The equation is then called *Kolmogorov's backward equation*. Let f be continuous with compact support. Show that $v^{T,f}(t,x) = \mathbb{E}\, f(X_T^{t,x})$ satisfies Kolmogorov's backward equation for all $t < T$. Suppose that there exists a function p of four variables t, x, r, y such that for all f that are continuous with compact support one has $v^{r,f}(t,x) = \int_{\mathbb{R}} f(y)p(t,x;r,y)\,\mathrm{d}y$ and $\lim_{t\uparrow r} v^{r,f}(t,x) = f(x)$. (Such a function is then called *fundamental solution*.) Show that for fixed r, y the function $(t,x) \to p(t,x;r,y)$ satisfies Kolmogorov's backward equation. What is the interpretation of the function p? Show that the solution to the Cauchy problem (11.3), (11.4) with $k = 0$ takes the form

$$v(t,x) = \int_{\mathbb{R}} p(t,x;T,y)f(y)\,\mathrm{d}y + \int_t^T \int_{\mathbb{R}} p(t,x;r,y)g(r,y)\,\mathrm{d}y\,\mathrm{d}r.$$

11.3 Prove Lemma 11.1. *Hint:* Proceed as in Exercise 10.14 and use the Doob and Burkholder-Davis-Gundy inequalities.

11.4 The Black-Scholes partial differential equation is

$$v_t(t,x) + \frac{1}{2}\sigma^2 x^2 v_{xx}(t,x) + rxv_x(t,x) = rv(t,x),$$

with constants $r, \sigma > 0$. Let X be the price process of some financial asset and $T > 0$ a finite time horizon. A simple contingent claim is a nonnegative measurable function of X_T, $f(X_T)$ say, representing the value at time T of some *derived* financial product. The pricing problem is to find the right price at any time t prior to T that should be charged to trade this claim. Clearly, at time $t = T$, this price should be equal to $f(X_T)$. The theory of Mathematical Finance dictates that (under the appropriate assumptions) this price is equal to $v(t, X_t)$, with v a solution to the Black-Scholes equation. Give an explicit solution for the case of a *European call option*, i.e. $f(x) = \max\{x - K, 0\}$, where $K > 0$ is some positive constant.

11.5 Here we consider the Cauchy problem with an initial condition. We have the partial differential equation

$$u_t + ku = \mathcal{L}_t u + g,$$

with the initial condition $u(0, \cdot) = f$. Formulate sufficient conditions such that this problem has a unique solution which is given by

$$\begin{aligned} u(t,x) = {} & \mathbb{E}\, f(X_t^x) \exp(-\int_0^t k(u, X_u^x)\,\mathrm{d}u) \\ & + \mathbb{E} \int_0^t g(s, X_s^x) \exp(-\int_0^s k(u, X_u^x)\,\mathrm{d}u)\,\mathrm{d}s, \end{aligned}$$

where X^x is the solution to (10.1) with $X_0^x = x$.

A Brownian motion

In this section we prove the existence of Brownian motion. The technique that is used in the existence proof is based on linear interpolation properties for continuous functions.

A.1 Interpolation of continuous functions

Let a continuous function $f : [0,1] \to \mathbb{R}$ be given with $f(0) = 0$. We will construct an approximation scheme of f, consisting of continuous piecewise linear functions. To that end we make use of the dyadic numbers in $[0,1]$. Let for each $n \in \mathbb{N}$ the set D_n be equal to $\{k2^{-n} : k = 0, \ldots, 2^n\}$. The dyadic numbers in $[0,1]$ are then the elements of $\cup_{n=1}^{\infty} D_n$. To simplify the notation we write t_k^n for $k2^{-n} \in D_n$.

The interpolation starts with $f_0(t) \equiv tf(1)$ and then we define the other f_n recursively. Suppose f_{n-1} has been constructed by prescribing the values at the points t_k^{n-1} for $k = 0, \ldots, 2^{n-1}$ and by linear interpolation between these points. Look now at the points t_k^n for $k = 0, \ldots, 2^n$. For the even integers $2k$ we take $f_n(t_{2k}^n) = f_{n-1}(t_k^{n-1})$. Then for the odd integers $2k-1$ we define $f_n(t_{2k-1}^n) = f(t_{2k-1}^n)$. We complete the construction of f_n by linear interpolation between the points t_k^n. Note that for $m \geq n$ we have $f_m(t_k^n) = f(t_k^n)$.

The above interpolation scheme can be represented in a more compact way (to be used in Section A.2) by using the so-called *Haar* functions H_k^n. These are defined as follows. $H_1^0(t) \equiv 1$ and for each n we define H_k^n for $k \in I(n) = \{1, \ldots, 2^{n-1}\}$ by

$$H_k^n(t) = \begin{cases} \frac{1}{2\sigma_n} & \text{if } t_{2k-2}^n \leq t < t_{2k-1}^n \\ -\frac{1}{2\sigma_n} & \text{if } t_{2k-1}^n \leq t < t_{2k}^n \\ 0 & \text{elsewhere} \end{cases} \tag{A.1}$$

where $\sigma_n = 2^{-\frac{1}{2}(n+1)}$. Next we put $S_k^n(t) = \int_0^t H_k^n(u)\,\mathrm{d}u$. Note that the support of S_k^n is the interval $[t_{2k-2}^n, t_{2k}^n]$ and that the graphs of the S_k^n are tent shaped with peaks of height σ_n at t_{2k-1}^n.

Next we will show how to cast the interpolating scheme in such a way that the Haar functions, or rather the *Schauder* functions S_k^n, are involved. Observe that not only the S_k^n are tent shaped, but also the consecutive differences $f_n - f_{n-1}$ on each of the intervals $(t_{k-1}^{n-1}, t_k^{n-1})$! Hence they are multiples of each other and to express the interpolation in terms of the S_k^n we only have to determine the multiplication constant. The height of the peak of $f_n - f_{n-1}$ on $(t_{k-1}^{n-1}, t_k^{n-1})$ is the value η_k^n at the midpoint t_{2k-1}^n. So $\eta_k^n = f(t_{2k-1}^n) - \frac{1}{2}(f(t_{k-1}^{n-1}) + f(t_k^{n-1}))$. Then we have for $t \in (t_{2k-2}^n, t_{2k}^n)$ the simple formula

$$f_n(t) - f_{n-1}(t) = \frac{\eta_k^n}{\sigma_n} S_k^n(t),$$

and hence we get for all t

$$f_n(t) = f_{n-1}(t) + \sum_{k \in I(n)} \frac{\eta_k^n}{\sigma_n} S_k^n(t). \tag{A.2}$$

Summing Equation (A.2) over n leads with $I(0) = \{1\}$ to the following representation of f_n on the whole interval $[0, 1]$:

$$f_n(t) = \sum_{m=0}^{n} \sum_{k \in I(m)} \frac{\eta_k^m}{\sigma_m} S_k^m(t). \tag{A.3}$$

Proposition A.1 *Let f be a continuous function on $[0, 1]$. With the f_n defined by (A.3) we have $||f - f_n|| \to 0$, where $|| \cdot ||$ denotes the sup norm.*

Proof Let $\varepsilon > 0$ and choose N such that we have $|f(t) - f(s)| \leq \varepsilon$ as soon as $|t - s| < 2^{-N}$. It is easy to see that then $|\eta_k^n| \leq \varepsilon$ if $n \geq N$. On the interval $[t_{2k-2}^n, t_{2k}^n]$ we have that

$$|f(t) - f_n(t)| \leq |f(t) - f(t_{2k-1}^n)| + |f_n(t_{2k-1}^n) - f_n(t)| \leq \varepsilon + \eta_k^n \leq 2\varepsilon.$$

This bound holds on any of the intervals $[t_{2k-2}^n, t_{2k}^n]$. Hence $||f - f_n|| \to 0$. □

Corollary A.2 *For arbitrary $f \in C[0, 1]$ we have*

$$f = \sum_{m=0}^{\infty} \sum_{k \in I(m)} \frac{\eta_k^m}{\sigma_m} S_k^m, \tag{A.4}$$

where the infinite sum converges in the sup norm.

A.2 Existence of Brownian motion

Recall the definition of Brownian motion.

Definition A.3 A standard Brownian motion, also called Wiener process, is a stochastic process W, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with time index set $\mathcal{T} = [0, \infty)$ with the following properties.

(i) $W_0 = 0$,

(ii) The increments $W_t - W_s$ have a normal $N(0, t - s)$ distribution for all $t > s$.

(iii) The increments $W_t - W_s$ are independent of all $W(u)$ with $u \leq s < t$.

(iv) The paths of W are continuous functions.

A process W is called a Brownian motion relative to a filtration $\mathbb{F}$ if it is adapted to $\mathbb{F}$ and a standard Brownian motion with (iii) replaced with

(iii') The increments $W_t - W_s$ are independent of $\mathcal{F}_s$ for all $0 \leq s < t$.

Remark A.4 One can show that part (iii) of Definition A.3 is equivalent to the following. For all finite sequences $0 \leq t_0 \leq \ldots \leq t_n$ the random variables $W_{t_k} - W_{t_{k-1}}$ $(k = 1, \ldots, n)$ are independent.

The method we use to prove existence of Brownian motion (and of a suitable probability space on which it is defined) is a kind of converse of the interpolation scheme of Section A.1. We will define what is going to be Brownian motion recursively on the time interval $[0,1]$ by attributing values at the dyadic numbers in $[0,1]$. A crucial part of the construction is the following fact. Supposing that we have shown that Brownian motion exists we consider the random variables W_s and W_t with $s < t$. Draw independent of these random variables a random variable ξ with a standard normal distribution and define $Z = \frac{1}{2}(W_s + W_t) + \frac{1}{2}\sqrt{t-s}\,\xi$. Then Z also has a normal distribution, whose expectation is zero and whose variance can be shown to be $\frac{1}{2}(t+s)$ (this is Exercise A.1). Hence Z has the same distribution as $W_{\frac{1}{2}(t+s)}$! This fact lies at the heart of the construction of Brownian motion by a kind of 'inverse interpolation' that we will present below.

Let, as in Section A.1, $I(0) = \{1\}$ and $I(n)$ be the set $\{1, \ldots, 2^{n-1}\}$ for $n \geq 1$. Take a sequence of independent standard normally distributed random variables ξ^n_k that are all defined on some probability space Ω with $k \in I(n)$ and $n \in \mathbb{N} \cup \{0\}$ (it is a result in probability theory that one can take for Ω a countable product of copies of $\mathbb{R}$, endowed with a product σ-algebra and a product measure). With the aid of these random variables we are going to construct a sequence of continuous processes W^n as follows. Let, also as in Section A.1, $\sigma_n = 2^{-\frac{1}{2}(n+1)}$. Put

$$W^0_t = t\xi^0_1.$$

For $n \geq 1$ we get the following recursive scheme

$$W^n_{t^n_{2k}} = W^{n-1}_{t^{n-1}_k} \tag{A.5}$$

$$W^n_{t^n_{2k-1}} = \frac{1}{2}\left(W^{n-1}_{t^{n-1}_{k-1}} + W^{n-1}_{t^{n-1}_k}\right) + \sigma_n \xi^n_k. \tag{A.6}$$

For other values of t we define W^n_t by linear interpolation between the values of W^n at the points t^n_k. As in Section A.1 we can use the Schauder functions for a compact expression of the random functions W^n. We have

$$W^n_t = \sum_{m=0}^{n} \sum_{k \in I(m)} \xi^m_k S^m_k(t). \tag{A.7}$$

Note the similarity of this equation with (A.3). The main result of this section is

Theorem A.5 *For almost all ω the functions $t \mapsto W^n_t(\omega)$ converge uniformly to a continuous function $t \mapsto W_t(\omega)$ and the process $W : (\omega, t) \to W_t(\omega)$ is Brownian motion on $[0,1]$.*

Proof We start with the following result. If Z has a standard normal distribution and $x > 0$, then (Exercise A.2)

$$\mathbb{P}(|Z| > x) \leq \sqrt{\frac{2}{\pi}}\frac{1}{x}\exp(-\frac{1}{2}x^2). \tag{A.8}$$

Let $\beta_n = \max_{k \in I(n)} |\xi_k^n|$. Then $b_n := \mathbb{P}(\beta_n > n) \leq 2^{n-1}\sqrt{\frac{2}{\pi}}\frac{1}{n}\exp(-\frac{1}{2}n^2)$. Observe that $\sum_n b_n$ is convergent and that hence in virtue of the Borel-Cantelli lemma $\mathbb{P}(\limsup\{\beta_n > n\}) = 0$. Hence we can find a subset $\tilde{\Omega}$ of Ω with $\mathbb{P}(\tilde{\Omega}) = 1$, such that for all $\omega \in \tilde{\Omega}$ there exists a natural number $n(\omega)$ with the property that all $|\xi_k^n(\omega)| \leq n$ if $n \geq n(\omega)$ and $k \in I(n)$.
Consequently, for $p > n \geq n(\omega)$ we have

$$\sup_t |W_t^n(\omega) - W_t^p(\omega)| \leq \sum_{m=n+1}^{\infty} m\sigma_m < \infty. \tag{A.9}$$

This shows that the sequence $W_\cdot^n(\omega)$ with $\omega \in \tilde{\Omega}$ is Cauchy in $C[0,1]$, so that it converges to a continuous limit, which we call $W_\cdot(\omega)$. For ω's not in $\tilde{\Omega}$ we define $W_\cdot(\omega) = 0$. So we now have continuous functions $W_\cdot(\omega)$ for all ω with the property $W_0(\omega) = 0$.

As soon as we have verified properties (ii) and (iii) of Definition A.3 we know that W is a Brownian motion. We will verify these two properties at the same time by showing that all increments $\Delta_j := W_{t_j} - W_{t_{j-1}}$ with $t_j > t_{j-1}$ are independent $N(0, t_j - t_{j-1})$ distributed random variables. Thereto we will prove that the characteristic function $\mathbb{E}\exp(\mathrm{i}\sum_j \lambda_j \Delta_j)$ is equal to $\exp(-\frac{1}{2}\sum_j \lambda_j^2(t_j - t_{j-1}))$.

The Haar functions form a Complete Orthonormal System of $L^2[0,1]$ (see Exercise A.3). Hence every function $f \in L^2[0,1]$ has the representation $f = \sum_{n,k}\langle f, H_k^n\rangle H_k^n = \sum_{n=0}^{\infty}\sum_{k \in I(n)}\langle f, H_k^n\rangle H_k^n$, where $\langle \cdot,\cdot \rangle$ denotes the inner product of $L^2[0,1]$ and where the infinite sum is convergent in $L^2[0,1]$. As a result we have for any two functions f and g in $L^2[0,1]$ the Parseval identity $\langle f, g\rangle = \sum_{n,k}\langle f, H_k^n\rangle\langle g, H_k^n\rangle$. Taking the specific choice $f = 1_{[0,t]}$ and $g = 1_{[0,s]}$ results in $\langle 1_{[0,t]}, H_k^n\rangle = S_k^n(t)$ and $t \wedge s = \langle 1_{[0,t]}, 1_{[0,s]}\rangle = \sum_{n,k} S_k^n(t) S_k^n(s)$.
We use this property to compute the limit of $\mathrm{Cov}(W_t^n, W_s^n)$. We have

$$\begin{aligned}
\mathrm{Cov}(W_t^n, W_s^n) &= \mathbb{E}\left(\sum_{m=0}^{n}\sum_{k \in I(m)} \xi_k^m S_k^m(t) \sum_{p=0}^{n}\sum_{l \in I(p)} \xi_l^p S_l^p(s)\right) \\
&= \sum_{m=0}^{n}\sum_{k \in I(m)} S_k^m(t) S_k^m(s)),
\end{aligned}$$

which converges to $s \wedge t$ for $n \to \infty$. Since for all fixed t we have $W_t^n \to W_t$ a.s., we have $\mathbb{E}\exp(\mathrm{i}\sum_j \lambda_j \Delta_j) = \lim_{n\to\infty}\mathbb{E}\exp(\mathrm{i}\sum_j \lambda_j \Delta_j^n)$ with $\Delta_j^n = W_{t_j}^n - W_{t_{j-1}}^n$.
We compute

$$\begin{aligned}
\mathbb{E}\exp(\mathrm{i}\sum_j \lambda_j \Delta_j^n) &= \mathbb{E}\exp(\mathrm{i}\sum_{m \leq n}\sum_k\sum_j \lambda_j \xi_k^m (S_k^m(t_j) - S_k^m(t_{j-1}))) \\
&= \prod_{m \leq n}\prod_k \mathbb{E}\exp(\mathrm{i}\sum_j \lambda_j \xi_k^m (S_k^m(t_j) - S_k^m(t_{j-1}))) \\
&= \prod_{m \leq n}\prod_k \exp(-\frac{1}{2}(\sum_j \lambda_j (S_k^m(t_j) - S_k^m(t_{j-1}))^2)
\end{aligned}$$

Working out the double product, we get in the exponential the sum over the four variables $i, j, m \leq n, k$ of $-\frac{1}{2}\lambda_i\lambda_j$ times

$$S_k^m(t_j)S_k^m(t_i) - S_k^m(t_{j-1})S_k^m(t_i) - S_k^m(t_j)S_k^m(t_{i-1}) + S_k^m(t_{j-1})S_k^m(t_{i-1})$$

and this quantity converges to $t_j - t_{j-1}$ as $n \to \infty$. Hence the expectation in the display converges to $\exp(-\frac{1}{2}\sum_j \lambda_j^2(t_j - t_{j-1}))$ as $n \to \infty$. This completes the proof. $\square$

Having constructed Brownian motion on $[0,1]$, we proceed to show that it also exists on $[0,\infty)$. Take for each $n \in \mathbb{N}$ a probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ that supports a Brownian motion W^n on $[0,1]$. Consider then $\Omega = \prod_n \Omega_n$, $\mathcal{F} = \prod_n \mathcal{F}_n$ and $\mathbb{P} = \prod_n \mathbb{P}_n$. Note that these Brownian motions are independent by construction. Let $\omega = (\omega_1, \omega_2, \ldots)$ and define then

$$W_t(\omega) = \sum_{n\geq 0} 1_{[n,n+1)}(t)\left(\sum_{k=1}^{n} W_1^k(\omega_k) + W_{t-n}^{n+1}(\omega_{n+1})\right). \tag{A.10}$$

This obviously defines a process with continuous paths and for all t the random variable W_t is the sum of independent normal random variables. It is not hard to see that the thus defined process has independent increments. It is immediate that $\mathbb{E}\, W_t = 0$ and that $\operatorname{Var} W_t = t$.

A.3 Properties of Brownian motion

Although Brownian motion is a process with continuous paths, the paths are very irregular. For instance, they are (almost surely) of unbounded variation over nonempty intervals. This follows from the Doob-Meyer decomposition (Theorem 2.16) of W^2, or from Proposition 3.8, see also Exercise 3.9.

We can say something more precise about the continuity of the sample paths of Brownian motion.

Proposition A.6 *The paths of Brownian motion are a.s. Hölder continuous of any order γ with $\gamma < \frac{1}{2}$, i.e. for almost every ω and for every $\gamma \in (0, \frac{1}{2})$ there exists a $C > 0$ such that $|W_t(\omega) - W_s(\omega)| \leq C|t-s|^\gamma$ for all t and s.*

Proof (Exercise A.4). Use that $|S_k^m(t) - S_k^m(s)| \leq 2^{\frac{1}{2}(m-1)}|t-s|$ and an inequality similar to (A.9). $\square$

Having established a result on the continuity of the paths of Brownian motion, we now turn to the question of differentiability of these paths. Proposition A.7 says that they are nowhere differentiable.

Proposition A.7 *Put $D = \{\omega : t \mapsto W_t(\omega)$ is differentiable at some $s \in (0,1)\}$. Then D is contained in a set of zero probability.*

Proof Let $A_{jk}(s)$ be the set

$$\{\omega : |W_{s+h}(\omega) - W_s(\omega)| \leq j|h|, \text{ for all } h \text{ with } |h| \leq 1/k\}$$

and $A_{jk} = \bigcup_{s\in(0,1)} A_{jk}(s)$. Then we have the inclusion $D \subset \bigcup_{jk} A_{jk}$. Fix j, k and s for a while, pick $n \geq 4k$, choose $\omega \in A_{jk}(s)$ and choose i such that $s \in (\frac{i-1}{n}, \frac{i}{n}]$. Note first for $l = 1, 2, 3$ the trivial inequalities $\frac{i+l}{n} - s \leq \frac{l+1}{n} \leq \frac{1}{k}$. The triangle inequality and $\omega \in A_{jk}(s)$ gives for $l = 1, 2, 3$

$$\begin{aligned}
&|W_{\frac{i+l}{n}}(\omega) - W_{\frac{i+l-1}{n}}(\omega)| \\
&\leq |W_{\frac{i+l}{n}}(\omega) - W_s(\omega)| + |W_s(\omega) - W_{\frac{i+l-1}{n}}(\omega)| \\
&\leq \frac{l+1}{n}j + \frac{l}{n}j = \frac{2l+1}{n}j.
\end{aligned}$$

It then follows that

$$A_{jk} \subset \bigcup_{i=1}^{n} \bigcap_{l=1,2,3} \{\omega : |W_{\frac{i+l}{n}}(\omega) - W_{\frac{i+l-1}{n}}(\omega)| \leq \frac{2l+1}{n}j\}. \tag{A.11}$$

We proceed by showing that the right hand side of this inclusion has probability zero. We use the following auxiliary result: if X has a $N(0, \sigma^2)$ distribution, then $\mathbb{P}(|X| \leq x) < x/\sigma$ (Exercise A.5). By the independence of the increments of Brownian motion the probability of the intersection in (A.11) is the product of the probabilities of each of the terms and this product is less then $105j^3n^{-3/2}$. Hence $\mathbb{P}(A_{jk}) \leq 105j^3n^{-1/2}$, which tends to zero for $n \to \infty$. The conclusion now follows from $D \subset \bigcup_{jk} A_{jk}$. □

A.4 Exercises

A.1 Show that the random variable Z on page 97 has a normal distribution with mean zero and variance equal to $\frac{1}{2}(s + t)$.

A.2 Prove inequality (A.8).

A.3 The Haar functions form a Complete Orthonormal System in $L^2[0, 1]$. Show first that the Haar functions are orthonormal. To prove that the system is complete, you argue as follows. Let f be orthogonal to all $H_{k,n}$ and set $F = \int_0^{\cdot} f(u)du$. Show that F is zero in all $t = k2^{-n}$, and therefore zero on the whole interval. Conclude that $f = 0$.

A.4 Prove Proposition A.6.

A.5 Show that for a random variable X with a $N(0, \sigma^2)$ distribution it holds that $\mathbb{P}(|X| \leq x) < x/\sigma$, for $x, \sigma > 0$.

B Optional sampling in discrete time

Let $\mathbb{F}$ be a filtration in discrete time, an increasing sequence of σ-algebras $\mathcal{F}_n$ $(n = 0, 1, \ldots)$. Recall the definition of a stopping time T, a map $T : \Omega \to [0, \infty]$ such that $\{T \leq n\} \in \mathcal{F}_n$ for every n. Of course T is a stopping time iff $\{T = n\} \in \mathcal{F}_n$ for every n.

For a stopping time T we define the σ-algebra

$$\mathcal{F}_T := \{F \subset \Omega : F \cap \{T \leq n\} \in \mathcal{F}_n \text{ for every } n\}.$$

If S and T are stopping times with $S \leq T$, then $\mathcal{F}_S \subset \mathcal{F}_T$. If X is a process with index set $\mathbb{N}$, we define $X_T = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{T=n\}}$ and so $X_T = X_T \mathbf{1}_{\{T<\infty\}}$. If also X_∞ is defined, we include $n = \infty$ in the last summation. In both cases X_T is a well-defined random variable and even $\mathcal{F}_T$-measurable (check!).

NB: All (sub)martingales and stopping times below are defined with respect to a given filtration $\mathbb{F}$.

Lemma B.1 *Let X be a submartingale and T a bounded stopping time, $T \leq N$ say for some $N \in \mathbb{N}$. Then $\mathbb{E}\,|X_T| < \infty$ and*

$$X_T \leq \mathbb{E}\,[X_N|\mathcal{F}_T] \quad \text{a.s.}$$

Proof Integrability of X_T follows from $|X_T| \leq \sum_{n=0}^{N} |X_n|$. Let $F \in \mathcal{F}_T$. Then, because $F \cap \{T = n\} \in \mathcal{F}_n$ and the fact that X is a submartingale, we have

$$\begin{aligned}
\mathbb{E}\,[X_N \mathbf{1}_F] &= \sum_{n=0}^{N} \mathbb{E}\,[X_N \mathbf{1}_F \mathbf{1}_{\{T=n\}}] \\
&\geq \sum_{n=0}^{N} \mathbb{E}\,[X_n \mathbf{1}_F \mathbf{1}_{\{T=n\}}] \\
&= \sum_{n=0}^{N} \mathbb{E}\,[X_T \mathbf{1}_F \mathbf{1}_{\{T=n\}}] \\
&= \mathbb{E}\,[X_T \mathbf{1}_F],
\end{aligned}$$

which is what we wanted to prove. □

Theorem B.2 *Let X be a uniformly integrable martingale with a last element X_∞, so $X_n = \mathbb{E}\,[X_\infty|\mathcal{F}_n]$ a.s. for every n. Let T and S be stopping times with $S \leq T$. Then X_T and X_S are integrable and*

(i) $X_T = \mathbb{E}\,[X_\infty|\mathcal{F}_T]$ a.s.

(ii) $X_S = \mathbb{E}\,[X_T|\mathcal{F}_S]$ a.s.

Proof First we show that X_T is integrable. Notice that $\mathbb{E}\,|X_T|\mathbf{1}_{\{T=\infty\}} = \mathbb{E}\,|X_\infty|\mathbf{1}_{\{T=\infty\}} \leq \mathbb{E}\,|X_\infty| < \infty$. Next, because $|X|$ is a submartingale with last element $|X_\infty|$,

$$\begin{aligned}\mathbb{E}\,|X_T|\mathbf{1}_{\{T<\infty\}} &= \sum_{n=0}^{\infty}\mathbb{E}\,|X_n|\mathbf{1}_{\{T=n\}}\\ &\le \sum_{n=0}^{\infty}\mathbb{E}\,|X_\infty|\mathbf{1}_{\{T=n\}}\\ &= \mathbb{E}\,|X_\infty|\mathbf{1}_{\{T<\infty\}} < \infty.\end{aligned}$$

We proceed with the proof of (i). Notice that $T\wedge n$ is a bounded stopping time for every n. But then by Lemma B.1 it holds a.s. that

$$\begin{aligned}\mathbb{E}\,[X_\infty|\mathcal{F}_{T\wedge n}] &= \mathbb{E}\,[\mathbb{E}\,[X_\infty|\mathcal{F}_n]|\mathcal{F}_{T\wedge n}]\\ &= \mathbb{E}\,[X_n|\mathcal{F}_{T\wedge n}]\\ &= X_{T\wedge n}.\end{aligned}$$

Let now $F\in\mathcal{F}_T$, then $F\cap\{T\le n\}\in\mathcal{F}_{T\wedge n}$ and by the above display, we have

$$\mathbb{E}\,[X_\infty\mathbf{1}_{F\cap\{T\le n\}}] = \mathbb{E}\,[X_{T\wedge n}\mathbf{1}_{F\cap\{T\le n\}} = \mathbb{E}\,[X_T\mathbf{1}_{F\cap\{T\le n\}}].$$

Let $n\to\infty$ and apply the Dominated convergence theorem to get

$$\mathbb{E}\,[X_\infty\mathbf{1}_F\mathbf{1}_{\{T<\infty\}}] = \mathbb{E}\,[X_T\mathbf{1}_F\mathbf{1}_{\{T<\infty\}}].$$

Together with the trivial identity $\mathbb{E}\,[X_\infty\mathbf{1}_F\mathbf{1}_{\{T=\infty\}}] = \mathbb{E}\,[X_T\mathbf{1}_F\mathbf{1}_{\{T=\infty\}}]$ this yields $\mathbb{E}\,[X_\infty\mathbf{1}_F] = \mathbb{E}\,[X_T\mathbf{1}_F]$ and (i) is proved.

For the proof of (ii) we use (i) two times and obtain

$$\mathbb{E}\,[X_T|\mathcal{F}_S] = \mathbb{E}\,[\mathbb{E}\,[X_\infty|\mathcal{F}_T]|\mathcal{F}_S] = \mathbb{E}\,[X_\infty|\mathcal{F}_S] = X_S.$$

□

Theorem B.3 *Let X be a submartingale such that $X_n\le 0$ for all $n=0,1,\ldots$. Let T and S be stopping times with $S\le T$. Then X_T and X_S are integrable and $X_S\le\mathbb{E}\,[X_T|\mathcal{F}_S]$ a.s.*

Proof Because of Lemma B.1 we have $\mathbb{E}\,[-X_{T\wedge n}]\le\mathbb{E}\,[-X_0]$ for every $n\ge 0$, which implies by virtue of Fatou's lemma $0\le\mathbb{E}\,[-X_T\mathbf{1}_{\{T<\infty\}}]<\infty$.

Let $E\in\mathcal{F}_S$, then $E\cap\{S\le n\}\in\mathcal{F}_{S\wedge n}$. Application of Lemma B.1 and non-positivity of X yields

$$\mathbb{E}\,[X_{S\wedge n}\mathbf{1}_E\mathbf{1}_{\{S\le n\}}]\le\mathbb{E}\,[X_{T\wedge n}\mathbf{1}_E\mathbf{1}_{\{S\le n\}}]\le\mathbb{E}\,[X_{T\wedge n}\mathbf{1}_E\mathbf{1}_{\{T\le n\}}]$$

and hence

$$\mathbb{E}\,[X_S\mathbf{1}_E\mathbf{1}_{\{S\le n\}}]\le\mathbb{E}\,[X_T\mathbf{1}_E\mathbf{1}_{\{T\le n\}}].$$

The Monotone convergence theorem yields $\mathbb{E}\,[X_S\mathbf{1}_E]\le\mathbb{E}\,[X_T\mathbf{1}_E]$. □

We finish this section with the Optional sampling theorem in discrete time.

Theorem B.4 *Let X be a submartingale with a last element X_∞, so $X_n \leq \mathbb{E}[X_\infty|\mathcal{F}_n]$ a.s. for every n. Let T and S be stopping times with $S \leq T$. Then X_T and X_S are integrable and*
(i) $X_T \leq \mathbb{E}[X_\infty|\mathcal{F}_T]$ a.s.
(ii) $X_S \leq \mathbb{E}[X_T|\mathcal{F}_S]$ a.s.

Proof Let $M_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$. By Theorem B.2, we get $M_S = \mathbb{E}[M_T|\mathcal{F}_S]$. Put then $Y_n = X_n - M_n$. Then Y is a submartingale with $Y_n \leq 0$. From Theorem B.3 we get $Y_S \leq \mathbb{E}[Y_T|\mathcal{F}_S]$. Since $X_T = M_T + Y_T$ and $X_S = M_S + Y_S$, the result follows. $\square$

C Functions of bounded variation and Stieltjes integrals

In this section we define functions of bounded variation and review some basic properties. Stieltjes integrals will be discussed subsequently. We consider functions defined on an interval $[a, b]$. Next to these we consider partitions Π of $[a, b]$, finite subsets $\{t_0, \ldots, t_n\}$ of $[a, b]$ with the convention $t_0 \leq \cdots \leq t_n$, and $\mu(\Pi)$ denotes the mesh of Π. Extended partitions, denoted Π^*, are partitions Π, together with additional points τ_i, with $t_{i-1} \leq \tau_i \leq t_i$. By definition $\mu(\Pi^*) = \mu(\Pi)$. Along with a function α, a partition Π, we define

$$V^1(\alpha; \Pi) := \sum_{i=1}^{n} |\alpha(t_i) - \alpha(t_{i-1})|,$$

the variation of α over the partition Π.

Definition C.1 A function α is said to be of bounded variation if $V^1(\alpha) := \sup_\Pi V^1(\alpha; \Pi) < \infty$, the supremum taken over all partitions Π. The variation function $v_\alpha : [a, b] \to \mathbb{R}$ is defined by $v_\alpha(t) = V^1(\alpha \mathbf{1}_{[a,t]})$.

A refinement Π' of a partition Π satisfies by definition the inclusion $\Pi \subset \Pi'$. In such a case, one has $\mu(\Pi') \leq \mu(\Pi)$ and $V^1(\alpha; \Pi') \geq V^1(\alpha; \Pi)$. It follows from the definition of $V^1(\alpha)$, that there exists a sequence (Π_n) of partitions (which can be taken as successive refinements) such that $V^1(\alpha; \Pi_n) \to V^1(\alpha)$.

Example C.2 Let α be continuously differentiable and assume $\int_a^b |\alpha'(t)|\, \mathrm{d}t$ finite. Then $V^1(\alpha) = \int_a^b |\alpha'(t)|\, \mathrm{d}t$. This follows, since $V^1(\alpha; \Pi)$ can be written as a Riemann sum

$$\sum_{i=1}^{n} |\alpha'(\tau_i)|\, (t_i - t_{i-1}),$$

where the τ_i satisfy $t_{i-1} \leq \tau_i \leq t_i$ and $\alpha'(\tau_i) = \frac{\alpha(t_i) - \alpha(t_{i-1})}{t_i - t_{i-1}}$.

Note that v_α is an increasing function with $v_\alpha(a) = 0$ and $v_\alpha(b) = V^1(\alpha)$. Any monotone function α is of bounded variation and in this case $V^1(\alpha) = |\alpha(b) - \alpha(a)|$ and $v_\alpha(t) = |\alpha(t) - \alpha(a)|$. Also the difference of two increasing functions is of bounded variation. This fact has a converse.

Proposition C.3 *Let α be of bounded variation. Then there exists increasing functions v_α^+ and v_α^- such that $v_\alpha^+(a) = v_\alpha^-(a) = 0$, $\alpha(t) - \alpha(a) = v_\alpha^+(t) - v_\alpha^-(t)$. Moreover, one can choose them such that $v_\alpha^+ + v_\alpha^- = v_\alpha$.*

Proof Define

$$v_\alpha^+(t) = \frac{1}{2}(v_\alpha(t) + \alpha(t) - \alpha(a))$$
$$v_\alpha^-(t) = \frac{1}{2}(v_\alpha(t) - \alpha(t) + \alpha(a)).$$

We only have to check that these functions are increasing, since the other statements are obvious. Let $t' > t$. Then $v_\alpha^+(t') - v_\alpha^+(t) = \frac{1}{2}(v_\alpha^+(t') - v_\alpha^+(t) + \alpha(t') - \alpha(t))$. The difference $v_\alpha^+(t') - v_\alpha^+(t)$ is the variation of α over the interval $[t, t']$, which is greater than or equal to $|\alpha(t') - \alpha(t)|$. Hence $v_\alpha^+(t') - v_\alpha^+(t) \geq 0$, and the same holds for $v_\alpha^-(t') - v_\alpha^-(t)$. □

The decomposition in this proposition enjoys a minimality property. If w^+ and w_- are increasing functions, $w^+(a) = w^-(a) = 0$ and $\alpha(t) - \alpha(a) = w^+(t) - w^-(t)$, then for all $t' > t$ one has $w^+(t') - w^+(t) \geq v_\alpha^+(t') - v_\alpha^+(t)$ and $w^-(t') - w^-(t) \geq v_\alpha^-(t') - v_\alpha^-(t)$. This property is basically the same as its counterpart for the Jordan decomposition of signed measures.

The following definition generalizes the concept of Riemann integral.

Definition C.4 Let $f, \alpha : [a, b] \to \mathbb{R}$ and Π^* be an extended partition of $[a, b]$. Write

$$S(f, \alpha; \Pi^*) = \sum_{i=1}^n f(\tau_i)\,(\alpha(t_i) - \alpha(t_{i-1})).$$

We say that $S(f, \alpha) = \lim_{\mu(\Pi^*) \to 0} S(f, a; \Pi^*)$, if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $\mu(\Pi^*) < \delta$ implies $|S(f, \alpha) - S(f, \alpha; \Pi^*)| < \varepsilon$. If this happens, we say that f is integrable w.r.t. α and we commonly write $\int f\,\mathrm{d}\alpha$ for $S(f, \alpha)$, and call it the Stieltjes integral of f w.r.t. α.

Proposition C.5 *Let $f, \alpha : [a, b] \to \mathbb{R}$, f continuous and α of bounded variation. Then f is integrable w.r.t. α. Moreover, the triangle inequality $|\int f\,\mathrm{d}\alpha| \leq \int |f|\,\mathrm{d}v_\alpha$ holds.*

Proof To show integrability of f w.r.t. α, the idea is to compare $S(f, \alpha; \Pi_1^*)$ and $S(f, \alpha; \Pi_2^*)$ for two extended partitions Π_1^* and Π_2^*. By constructing another extended partition Π^* that is a *refinement* of Π_1^* and Π_2^* in the sense that all t_i and t_i from Π_1^* and Π_2^* belong to Π^*, it follows from

$$|S(f, \alpha; \Pi_1^*) - S(f, \alpha; \Pi_2^*)| \leq |S(f, \alpha; \Pi_1^*) - S(f, \alpha; \Pi^*)| + |S(f, \alpha; \Pi^*) - S(f, \alpha; \Pi_2^*)|$$

that it suffices to show that $|S(f,\alpha;\Pi_1^*) - S(f,\alpha;\Pi_2^*)|$ can be made small for Π_2 a refinement of Π_1. Let $\varepsilon > 0$ and choose $\delta > 0$ such that $|f(t) - f(s)| < \varepsilon$ whenever $|t-s| < \delta$ (possible by uniform continuity of f). Assume that $\mu(\Pi_1) < \delta$, then also $\mu(\Pi_2) < \delta$. Consider first two extended partitions of Π_1, Π_1^* and Π_1', the latter with intermediate points τ_i'. Then

$$\begin{aligned}|S(f,\alpha;\Pi_1^*) - S(f,\alpha;\Pi_1')| &\leq \sum_i |f(\tau_i) - f(\tau_i')|\,|\alpha(t_i) - \alpha(t_{i-1})| \\ &\leq \varepsilon\, V^1(\alpha;\Pi_1) \leq \varepsilon\, V^1(\alpha).\end{aligned}$$

In the next step we assume that Π_2 is obtained from Π_1 by adding one point, namely τ_j for some τ_j from the extended partition Π_1^*. Further we assume that Π_2^* contains all the intermediate points τ_i from Π_1^*, whereas we also take the intermediate points from the intervals $[t_{j-1}, \tau_j]$ and $[\tau_j, t_j]$ both equal to τ_j. It follows that $S(f,\alpha;\Pi_1^*) = S(f,\alpha;\Pi_2^*)$. A combination of the two steps finishes the proof. The triangle inequality for the integrals holds almost trivially. $\square$

Proposition C.6 *Let $f, \alpha : [a,b] \to \mathbb{R}$, be continuous and of bounded variation. Then the following integration by parts formula holds.*

$$\int f \,\mathrm{d}\alpha + \int \alpha \,\mathrm{d}f = f(b)\alpha(b) - f(a)\alpha(a).$$

Proof Choose points $t_a \leq \cdots \leq t_n$ in $[a,b]$ and $\tau_i \in [t_{i-1}, t_i]$, $i = 1, \ldots, n$, to which we add $\tau_0 = a$ and $\tau_{n+1} = b$. By Abel's summation formula, we have

$$\begin{aligned}&\sum_{i=1}^n f(\tau_i)\,(\alpha(t_i) - \alpha(t_{i-1})) = \\ &\quad f(b)\alpha(b) - f(a)\alpha(a) - \sum_{i=1}^{n+1} \alpha(t_{i-1})\,(f(\tau_i) - f(\tau_{i-1})).\end{aligned}$$

The result follows by application of Proposition C.5. $\square$

This proposition can be used to *define* $\int \alpha \,\mathrm{d}f$ for functions α of bounded variation and continuous functions f, simply by putting

$$\int \alpha \,\mathrm{d}f := f(b)\alpha(b) - f(a)\alpha(a) - \int f \,\mathrm{d}\alpha.$$

Next we give an example that illustrates that the continuity assumption on f in Proposition C.5 cannot be omitted in general.

Example C.7 Let $\alpha : [-1,1] \to \mathbb{R}$ be given by $\alpha(t) = \mathbf{1}_{[0,1]}(t)$. Let $f : [-1,1] \to \mathbb{R}$ be discontinuous at zero, for instance $f = \alpha$; note that f and α share a point of discontinuity. Let Π be a partition of $[-1,1]$ whose elements are numbered in such a way that $t_{m-1} \leq 0 \leq t_m$, let $\tau = \tau_m \in [t_{m-1}, t_m]$. Then $S(f,\alpha;\Pi^*) = f(\tau)$, and this doesn't converge for $\tau \to 0$.

In the annoying case of the above example, the Stieltjes integral is not defined. By adjusting the concept of Stieltjes integral into the direction of a Lebesgue integral, we obtain that also in this case the integral is well defined.

It follows from Proposition C.3 that α admits finite left and right limits at all t in $[a,b]$. By α_+ we denote the function given by $\alpha_+(t) = \lim_{u \downarrow t} \alpha(u)$ for $t \in [a,b)$ and $\alpha_+(b) = \alpha(b)$. Note that α_+ is right-continuous. Let $\mu = \mu_\alpha$ be the signed measure on $([a,b], \mathcal{B}([a,b]))$ that is uniquely defined by $\mu((s,t]) = \alpha_+(t) - \alpha_+(s)$ for all $a \leq s < t \leq b$. For measurable f one can consider the Lebesgue integral $\int f \, \mathrm{d}\mu_\alpha$. We now present a result relating Stieltjes and Lebesgue integrals. The proposition below gives an example of such a connection, but can be substantially generalized.

Proposition C.8 *Let $f, \alpha : [a,b] \to \mathbb{R}$, f continuous and α of bounded variation. Then the Lebesgue integral $\int f \, \mathrm{d}\mu_\alpha$ and the Stieltjes integral $\int f \, \mathrm{d}\alpha$ are equal, $\int f \, \mathrm{d}\mu_\alpha = \int f \, \mathrm{d}\alpha$.*

Proof Exercise. □

For $\mu = \mu_\alpha$ as above, we call the integral $\int f \, \mathrm{d}\alpha$, defined as $\int f \, \mathrm{d}\mu_\alpha$, the Lebesgue-Stieltjes integral of f w.r.t. α. Consider again Example C.7. Now the Lebesgue-Stieltjes integral $\int \alpha \, \mathrm{d}\alpha$ is well defined and $\int \alpha \, \mathrm{d}\alpha = 1$.

D Dunford-Pettis criterion

In this section we present a proof of one of the two implications in the Dunford-Pettis characterization of uniform integrability, Lemma 2.14. Indeed, it concerns the implication that was needed in the proof of the Doob-Meyer decomposition. We formulate it as Proposition D.2 below. First some additional terminology.

Suppose that X is a Banach space. By X^* we denote the space of all continuous linear functionals on X, it is called the dual space of X. On says that a sequence $(x_n) \subset X$ converges weakly to $x \in X$ if $Tx_n \to Tx$, as $n \to \infty$, for all $T \in X^*$. The corresponding topology on X is called the *weak topology*, and one speaks of weakly open, weakly closed and weakly compact sets etc. This topology is defined by neighborhoods of $0 \in X$ of the form $\{x \in X : |T_i x| < \varepsilon,\, i = 1, \ldots, n\}$, with the $T_i \in X^*$, $\varepsilon > 0$ and $n \in \mathbb{N}$.

It is known that if $X = L^p(S, \Sigma, \mu)$, its dual space X^* can be identified with $L^q(S, \Sigma, \mu)$ (and we simply write $X^* = L^q(S, \Sigma, \mu)$), where $q = p/(p-1)$ for $p \geq 1$. In particular we have $X^* = X$, when $X = L^2(S, \Sigma, \mu)$, which is the Riesz-Fréchet theorem. First we present a lemma, which is a special case of *Alaoglu's theorem.*

Lemma D.1 *The weak closure of the unit ball $B = \{x \in X : ||x||_2 < 1\}$ in $X = L^2(S, \Sigma, \mu)$ is weakly compact.*

Proof The set B can be considered as a subset of $[-1, +1]^X$, since every $x \in X$ can be seen as a linear functional on X. Moreover, we can view the weak topology on X as induced by the product topology on $[-1,+1]^X$, if $[-1,+1]$ is endowed with the ordinary topology. By Tychonov's theorem, $[-1,+1]^X$ is compact in the product topology, and so is the weak closure of B as a closed subset of $[-1,+1]^X$. □

We now switch to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. From the definition of weak topology on $L^1(\Omega, \mathcal{F}, \mathbb{P})$ we deduce that a set $U \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$ is weakly sequentially compact, if every sequence in it has a subsequence with weak limit in U. Stated otherwise, U is weakly sequentially compact in $L^1(\Omega, \mathcal{F}, \mathbb{P})$, if every sequence $(X_n) \subset U$ has a subsequence (X_{n_k}) such that there exists $X \in U$ with $\mathbb{E}\, X_{n_k} Y \to \mathbb{E}\, XY$, for all bounded random variables Y. This follows since for $X = L^1$ its dual space $X^* = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$. We also say that a set $U \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$ is relatively sequentially compact if every sequence in U has a subsequence with weak limit in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ (not necessarily in U itself). This notion is equivalent to saying that the closure of U is weakly sequentially compact. We now formulate one half of the Dunford-Pettis criterion for uniform integrability.

Proposition D.2 *Let (X_n) be a uniformly integrable sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then (X_n) is relatively weakly sequentially compact.*

Proof Let $k > 0$ be an integer and put $X_n^k = X_n \mathbf{1}_{\{|X_n| \leq k\}}$. Then for fixed k the sequence (X_n^k) is bounded and thus also bounded in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. By Lemma D.1, it has a subsequence $(X_{n_j^k}^k)$ that converges weakly in L^2 to a limit X^k. We can even say more, there exists a sequence $(n_j) \subset \mathbb{N}$ such that for all k the sequence $(X_{n_j}^k)$ converges weakly to X^k. To see this, we argue as in the proof of Helly's theorem and utilize a diagonalization argument. Let $k = 1$ and find the convergent subsequence $(X_{n_j^1}^1)$. Consider then $(X_{n_j^1}^2)$ and find the convergent subsequence $(X_{n_j^2}^2)$. Note that also $(X_{n_j^2}^1)$ is convergent. Continue with $(X_{n_j^2}^3)$ etc. Finally, the subsequence in $\mathbb{N}$ that does the job for all k is (n_j^j), for which we write (n_j).

Observe that $(X_{n_j}^k - X_{n_j}^l)$ has $X^k - X^l$ as its weak limit in L^2, but then this is also the weak limit in L^1, since every bounded random variable Y is square integrable. Hence we have $\mathbb{E}\,(X_{n_j}^k - X_{n_j}^l)Y \to \mathbb{E}\,(X^k - X^l)Y$ for every bounded Y. Take $Y = \operatorname{sgn}(X^k - X^l)$. Since with this choice for Y we have $\mathbb{E}\,(X_{n_j}^k - X_{n_j}^l)Y \leq \mathbb{E}\,|X_{n_j}^k - X_{n_j}^l|$, we obtain $\liminf_j \mathbb{E}\,|X_{n_j}^k - X_{n_j}^l| \geq \mathbb{E}\,|X^k - X^l|$.

Write $\mathbb{E}\,|X_{n_j}^k - X_{n_j}^l| = \mathbb{E}\,|X_{n_j}(\mathbf{1}_{\{|X_{n_j}|>l\}} - \mathbf{1}_{\{|X_{n_j}|>k\}})|$. By uniform integrability this tends to zero for $k, l \to \infty$, uniformly in the n_j. It then follows that $\mathbb{E}\,|X^k - X^l| \to 0$ as $k, l \to \infty$, in other words, (X^k) is Cauchy in L^1 and thus has a limit X. We will now show that this X is the one we are after.

We have to show that $\mathbb{E}\,(X_{n_j} - X)Y \to 0$ for arbitrary bounded Y. Write

$$\mathbb{E}\,(X_{n_j} - X)Y = \mathbb{E}\,X_{n_j} \mathbf{1}_{\{|X_{n_j}|>k\}} Y + \mathbb{E}\,(X_{n_j}^k - X^k)Y + \mathbb{E}\,(X^k - X)Y.$$

The first of the three terms can be made arbitrary small by choosing k big enough, uniformly in the n_j, by uniform integrability. The second term tends to zero for $n_j \to \infty$ by the weak convergence in L^1 of the $X^k_{n_j}$ to X^k, whereas the third term can be made arbitrary small for large enough k by the L^1-convergence of X^k to X. $\square$

E Banach-Steinhaus theorem

Let S be a topological space. Recall that a subset of S is called nowhere dense (in S), if its closure has an empty interior. If $E \subset S$, then $\bar{E}$ denotes its closure and $\mathrm{int}E$ its interior.

Definition E.1 A subset of S is said to be of the first category (in S), if it is a countable union of nowhere dense sets. If a set is not of the first category, it is said to be of the second category.

The following theorem is known as Baire's category theorem.

Theorem E.2 *If S is a complete metric space, then the intersection of any countable collection of dense open sets is dense.*

Proof Let $E_1, E_2, \ldots$ be dense open sets and $D = \bigcap_{n=1}^{\infty} E_n$. Let B_0 be an arbitrary non-empty open set. We will show that $D \cap B_0 \neq \emptyset$. Select recursively open balls B_n with radius at most $\frac{1}{n}$ such that $\bar{B}_n \subset E_n \cap B_{n-1}$. This can be done since the E_n are dense subsets. Let c_n be the center of B_n. Since $B_n \subset B_{n-1}$ and the radii converge to zero, the sequence (c_n) is Cauchy. By completeness the sequence has a limit c and then $c \in \bigcap_{n=1}^{\infty} \bar{B}_n \subset D$. Since trivially $c \in B_0$, we have $D \cap B_0 \neq \emptyset$. $\square$

Remark E.3 Baire's theorem also holds true for a topological space S that is locally compact. The proof is almost the same.

Corollary E.4 *A metric space S is of the second category (in itself).*

Proof Let $E_1, E_2, \ldots$ be open nowhere dense subsets. Let $O_n = S \setminus \bar{E}_n$, an open set. Then $\bar{O}_n \supset S \setminus \mathrm{int}\bar{E}_n = S$, hence O_n is dense. It follows from Theorem E.2 that $\bigcap_n O_n \neq \emptyset$, so $\bigcup_n O_n^c \neq S$. But then $\bigcup_n E_n$ can't be equal to S either. $\square$

Let X and Y be Banach spaces and L a bounded linear operator from X into Y. Recall that boundedness is equivalent to continuity. The operator norm of L is defined by

$$||L|| = \sup\{||Lx|| : x \in X \text{ and } ||x|| = 1\}.$$

Note that we use the same symbol $||\cdot||$ for different norms. The following theorem is known as *the principle of uniform boundedness*, or as the Banach-Steinhaus theorem. Other, equivalent, formulations can be found in the literature.

Theorem E.5 *Let X and Y be Banach spaces and $\mathcal{L}$ be a family of bounded linear operators from X into Y. Suppose that for all $x \in X$ the set $\{||Lx|| : L \in \mathcal{L}\}$ is bounded. Then the set $\{||L|| : L \in \mathcal{L}\}$ is bounded as well.*

Proof Let $\varepsilon > 0$ be given and let $X_n = \{x \in X : \sup\{||Lx|| : L \in \mathcal{L}\} \leq n\varepsilon\}$. Since every L is continuous as well as $||\cdot||$, the set X_n is closed. In view of the assumption, we have $X = \bigcup_n X_n$. Since X is of the second category (Corollary E.4), it follows that some X_{n_0} must have nonempty interior. Hence there is $x_0 \in X_{n_0}$ and some $\delta > 0$ such that the closure of the ball $B = B(x_0, \delta)$ belongs to X_{n_0}. For every $x \in B$, we then have $||Lx|| \leq n_0\varepsilon$ for every $L \in \mathcal{L}$. Let $B' = B - x_0$. Then B' is a neighborhood of zero and every $y \in B'$ can be written as $y = x - x_0$ for some $x \in B$. This yields $||Ly|| \leq 2n_0\varepsilon$, valid for every $L \in \mathcal{L}$. Let now $v \in X$ be an arbitrary vector with norm one. Then we apply the above to $y := \delta v$ and obtain from this

$$||Lv|| \leq \frac{2n_0\varepsilon}{\delta},$$

valid for all $L \in \mathcal{L}$ and v with $||v|| = 1$. But then $\sup_{L\in\mathcal{L}} \sup_{v:||v||=1} ||Lv|| < \infty$, which is what we wanted to show. □

F The Radon-Nikodym theorem for absolutely continuous probability measures

Let us first state the Radon-Nikodym theorem 9.1 in the version we need it.

Theorem F.1 *Consider a measurable space $(\Omega, \mathcal{F})$ on which are defined two probability measures $\mathbb{Q}$ and $\mathbb{P}$. Assume $\mathbb{Q} \ll \mathbb{P}$. Then there exists a $\mathbb{P}$-a.s. unique random variable Z with $\mathbb{P}(Z \geq 0) = 1$ and $\mathbb{E}\, Z = 1$ such that for all $F \in \mathcal{F}$ one has*

$$\mathbb{Q}(F) = \mathbb{E}\,\mathbf{1}_F Z. \tag{F.12}$$

Proof Let $\hat{\mathbb{P}} = \frac{1}{2}(\mathbb{P} + \mathbb{Q})$ and consider $T : \mathcal{L}^2(\Omega, \mathcal{F}, \hat{\mathbb{P}}) \to \mathbb{R}$ defined by $TX = \mathbb{E}_{\mathbb{Q}}\, X$. Then

$$\begin{aligned}|TX| &\leq \mathbb{E}_{\mathbb{Q}}\,|X| \\ &\leq 2\mathbb{E}_{\hat{\mathbb{P}}}\,|X| \\ &\leq 2(\mathbb{E}_{\hat{\mathbb{P}}}\, X^2)^{1/2}.\end{aligned}$$

Hence T is a continuous linear functional on $\mathcal{L}^2(\Omega, \mathcal{F}, \hat{\mathbb{P}})$, and by the *Riesz-Fréchet theorem* there exist $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \hat{\mathbb{P}})$ such that $\mathbb{E}_{\mathbb{Q}}\, X = 2\,\mathbb{E}_{\hat{\mathbb{P}}}\, XY = \mathbb{E}_{\mathbb{Q}}\, XY + \mathbb{E}\, XY$. Hence we have

$$\mathbb{E}_{\mathbb{Q}}\, X(1 - Y) = \mathbb{E}\, XY, \tag{F.13}$$

for all $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \hat{\mathbb{P}})$. To find a convenient property of Y, we make two judicious choices for X in (F.13).

First we take $X = \mathbf{1}_{\{Y\geq 1\}}$. We then get

$$0 \geq \mathbb{E}_{\mathbb{Q}}\, \mathbf{1}_{\{Y\geq 1\}}(1 - Y) = \mathbb{E}\, \mathbf{1}_{\{Y\geq 1\}} Y \geq \mathbb{P}(Y \geq 1),$$

and hence $\mathbb{P}(Y \geq 1) = 0$. By absolute continuity also $\mathbb{Q}(Y \geq 1) = 0$.

Second, we choose $X = \mathbf{1}_{\{Y<0\}}$ and get, note that $X(1-Y) \geq 0$,

$$0 \leq \mathbb{E}_{\mathbb{Q}}\, \mathbf{1}_{\{Y<0\}}(1-Y) = \mathbb{E}\, \mathbf{1}_{\{Y<0\}} Y \leq 0,$$

and hence $\mathbb{P}(Y < 0) = 0$ and $\mathbb{Q}(Y < 0) = 0$. We conclude that $\mathbb{Q}(Y \in [0,1)) = \mathbb{P}(Y \in [0,1)) = 1$.

We now claim that the random variable Z in the assertion of the theorem is given by

$$Z = \frac{Y}{1-Y} \mathbf{1}_{[0,1)}(Y).$$

Note that Z takes its values in $[0, \infty)$. Write $Z = \lim_n Z_n$ with $Z_n = Y S_n$, where $S_n = \mathbf{1}_{[0,1)}(Y) \sum_{k=0}^{n-1} Y^k$ and note that $0 \leq S_n, Z_n \leq n$. Let $F \in \mathcal{F}$. We apply (F.13) with $X = \mathbf{1}_F S_n$ to obtain, using $(1-Y)S_n = \mathbf{1}_{[0,1)}(Y)(1-Y^n)$,

$$\mathbb{E}_{\mathbb{Q}}\, \mathbf{1}_F \mathbf{1}_{[0,1)}(Y)(1-Y^n) = \mathbb{E}\, \mathbf{1}_F Z_n.$$

Monotone convergence gives

$$\mathbb{E}_{\mathbb{Q}}\, \mathbf{1}_F \mathbf{1}_{[0,1)}(Y) = \mathbb{E}\, \mathbf{1}_F Z. \tag{F.14}$$

As $\mathbb{Q}(Y \in [0,1)) = 1$ one has $\mathbb{Q}(F) = \mathbb{Q}(F \cap \{Y \in [0,1)\}) = \mathbb{E}_{\mathbb{Q}}\, \mathbf{1}_F \mathbf{1}_{[0,1)}(Y)$, and (F.12) follows from (F.14). Moreover, as $\tilde{\mathbb{P}}$ is a probability measure, we get with $F = \Omega$ that $\mathbb{E}\, Z = 1$.

The final issue to address is the $\mathbb{P}$-a.s. uniqueness of Z. Suppose Z' is another random variable satisfying the assertion of the theorem. Then (F.12) is also valid for any $F \in \mathcal{F}$ with Z replaced with Z'. Take $F = \{Z > Z'\}$. By subtraction one obtains $\mathbb{E}\, \mathbf{1}_{\{Z>Z'\}}(Z - Z') = 0$ and hence $\mathbb{P}(Z \leq Z') = 1$. By swapping the roles of Z and Z', one concludes $\mathbb{P}(Z = Z') = 1$, which finishes the proof. $\square$

Index