

Stochastic Calculus

Nathanaël Berestycki

Cambridge, Lent 2009

These notes and other information about the course are available on
www.statslab.cam.ac.uk/~beresty/teach/StoCal/stocal.html

Christina Goldschmidt, Stefan Grosskinsky, Gregory Miermont, and James Norris have all contributed to these notes. However, errors are the responsibility of the author only. Comments and corrections are most welcome and should be sent to `N.Berestycki AT statslab.cam.ac.uk`

Date last modified: March 11, 2009

Contents

Introduction	3
1 Brownian Motion: definition and first properties	5
1.1 Introduction.	5
1.2 Wiener's theorem	6
1.3 Continuity and Hölder continuity of Brownian paths	9
1.4 Basic properties	13
2 Preliminaries	18
2.1 The bigger picture, in a nutshell	18
2.2 Finite variation integrals	19
2.3 Previsible processes	22
2.4 Local martingales	24
3 The stochastic integral	29
3.1 Simple integrands and L^2 properties	29
3.2 Quadratic variation	32
3.3 Itô integrals	36
3.4 Covariation	42
3.5 Itô's formula	44
4 Applications to Brownian motion and Martingales	48
4.1 Brownian martingales	48
4.2 Martingales characterized by their quadratic variation	51
4.3 Recurrence and Transience of Brownian motion	55
4.4 Donsker's invariance principle	57
4.5 Brownian motion and the Dirichlet problem	61
4.6 Girsanov's theorem	67
5 Stochastic Differential Equations	73
5.1 General definitions	73
5.2 Lipschitz coefficients	75
5.3 Strong Markov property and diffusion processes	79
5.4 Some links with PDEs	82
6 Diffusion approximation	86
6.1 Martingale problems.	86
6.2 Notions of weak convergence of processes	87
6.3 Markov chains and diffusions	89
6.4 Example.	96

Introduction

In this course we will develop the tools needed to handle continuous-time Markov processes in \mathbb{R}^d . We will restrict our attention to continuous processes, although the theory we develop is also well-suited for dealing with processes that exhibit jumps. The most basic example of such a process is Brownian motion $(W_t, t \geq 0)$ with constant diffusivity a . That is, for all $t \geq 0$, $W_t \sim \mathcal{N}(0, at)$. In fact, by the *Lévy-Khinchin Theorem* (see Advanced Probability, Chapter 8)

$$(X_t)_{t \geq 0} \text{ defined by } X_t = W_t + bt$$

is the most general spatially homogeneous Markov process (meaning, a Markov process whose transition probabilities are a shift of the transition probabilities starting from the origin).

Subject to reasonable regularity conditions, the most general inhomogeneous continuous-time Markov process can be characterized by the diffusivity $a(x)$ and the drift $b(x)$. The interpretation of these coefficients is that the process follows the solution of an ordinary differential equation in \mathbb{R}^d , $\dot{x} = b(x)$, which is perturbed by a Brownian noise with intensity $a(x)$ when the process is in x . Martingales play an important role in the description of the time evolution of such processes, which is demonstrated in the following non-rigorous calculation. Suppose that $d = 1$ and $(X_t)_{t \geq 0}$ is characterized by $a(x)$ and $b(x)$. Conditional on $X_s = x_s$ for all $s \leq t$

$$X_{t+dt} = x_t + b(x_t) dt + G_t,$$

where $G_t \sim N(0, a(x_t) dt)$. Take $f \in C_b^2(\mathbb{R})$ (bounded and two times continuously differentiable). Then by Taylor expansion

$$\mathbb{E}(f(X_{t+dt}) - f(X_t) \mid X_s = x_s, s \leq t) = \mathbb{E}\left(f'(x_t)\left(b(x_t) dt + G_t + \frac{1}{2}f''(x_t) G_t^2\right)\right).$$

Up to order dt this includes the second derivative f'' , since $G_t \sim \sqrt{a(x_t) dt} N(0, 1) = O(\sqrt{dt})$. This fact will be revisited later in a rigorous way as one of the main results of stochastic calculus (*Itô's formula*). Calculating the expectations we get

$$\mathbb{E}(f(X_{t+dt}) - f(X_t) \mid X_s = x_s, s \leq t) = L f(x_t) dt,$$

where

$$L f(x) = \frac{1}{2}a(x) f''(x) + b(x) f'(x),$$

which is also called the *generator* of the process $(X_t)_{t \geq 0}$. It describes the infinitesimal expected time evolution of the process in the following sense. Consider

$$M_t^f = f(X_t) - f(X_0) - \int_0^t L f(X_s) ds.$$

We have calculated that $\mathbb{E}(M_{t+dt}^f - M_t^f \mid X_s = x_s, s \leq t) = 0$ for all t , and so the process $(M_t^f)_{t \geq 0}$ is a martingale for all f . This property will often serve to characterize $(X_t)_{t \geq 0}$.

Organization of the course. The course contains three main parts of unequal size. In the first part we introduce and study basic properties of Brownian motion, such as continuity and Markov properties. This serves as a guide throughout the study of more complex processes later in the course. In the second part, which is the largest, we define the stochastic integral with respect to semimartingales and use this construction to prove Itô's formula. This allows us to study in far more details the properties of Brownian motion: transience, recurrence, applications to harmonic analysis, and Girsanov's theorem. These deep results are an application of the martingale point of view on Brownian motion, as opposed to the results in the first part of the course, which are essentially based on the Markov property. Finally, in the third part of this course, we develop the theory of stochastic differential equations and diffusions. We use this theory to show that many simple stochastic discrete models can be effectively studied by taking a *diffusion approximation*.

Stochastic calculus has very important application in sciences (biology or physics) as well as mathematical finance. For example, we will develop all the necessary tools to rigorously prove results like the Black-Scholes formula. But it is also a very beautiful part of modern probability and has led to a considerable enrichment of the subject itself.

1 Brownian Motion: definition and first properties

1.1 Introduction.

This chapter is devoted to the construction and some properties of one of probability theory's most fundamental objects. Brownian motion earned its name after R. Brown, who observed around 1827 that tiny particles of pollen in water have an extremely erratic motion. It was observed by Physicists that this was due to an important number of random shocks undertaken by the particles from the (much smaller) water molecules in motion in the liquid. A. Einstein established in 1905 the first mathematical basis for Brownian motion, by showing that it must be an isotropic Gaussian process. The first rigorous mathematical construction of Brownian motion is due to N. Wiener in 1923, using Fourier theory.

In order to motivate the introduction of this object, we first begin by a "microscopical" depiction of Brownian motion. Suppose $(X_n, n \geq 0)$ is a sequence of \mathbb{R}^d valued random variables with mean 0 and covariance matrix $\sigma^2 I_d$, which is the identity matrix in d dimensions, for some $\sigma^2 > 0$. Namely, if $X_1 = (X_1^1, \dots, X_1^d)$,

$$E[X_1^i] = 0, \quad E[X_1^i X_1^j] = \sigma^2 \delta_{ij}, \quad 1 \leq i, j \leq d.$$

We interpret X_n as the spatial displacement resulting from the shocks due to water molecules during the n -th time interval, and the fact that the covariance matrix is scalar stands for an isotropy assumption (no direction of space is privileged).

From this, we let $S_n = X_1 + \dots + X_n$ and we embed this discrete-time process into continuous time by letting

$$B_t^{(n)} = n^{-1/2} S_{[nt]}, \quad t \geq 0.$$

Let $|\cdot|$ be the Euclidean norm on \mathbb{R}^d and for $t > 0$ and $X, y \in \mathbb{R}^d$, define

$$p_t(x) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x|^2}{2t}\right),$$

which is the density of the Gaussian distribution $\mathcal{N}(0, tI_d)$ with mean 0 and covariance matrix tI_d . By convention, the Gaussian law $\mathcal{N}(m, 0)$ is the Dirac mass at m .

Proposition 1.1 *Let $0 = t_1 \geq t_2 < \dots < t_k$. Then the finite marginal distributions of $B^{(n)}$ with respect to times t_1, \dots, t_k converge weakly as $n \rightarrow \infty$. More precisely, if F is a bounded continuous function, and letting $x_0 = 0, t_0 = 0$,*

$$\mathbb{E} \left[F(B_{t_1}^{(n)}, \dots, B_{t_k}^{(n)}) \right] \xrightarrow{n \rightarrow \infty} \int_{(\mathbb{R}^d)^k} F(x_1, \dots, x_k) \prod_{1 \leq i \leq k} p_{\sigma^2(t_i - t_{i-1})}(x_i - x_{i-1}) dx_i.$$

Otherwise said, $(B_{t_1}^{(n)}, \dots, B_{t_k}^{(n)})$ converges in distribution to (G_1, G_2, \dots, G_k) , which is a random vector whose law is characterized by the fact that $(G_1, G_2 - G_1, \dots, G_k - G_{k-1})$ are independent centered Gaussian random variables with respective covariance matrices $\sigma^2(t_i - t_{i-1})I_d$.

Proof. With the notations of the theorem, we first check that $(B_{t_1}^{(n)}, B_{t_2}^{(n)} - B_{t_1}^{(n)}, \dots, B_{t_k}^{(n)} - B_{t_{k-1}}^{(n)})$ is a sequence of independent random variables. Indeed, one has for $1 \leq i \leq k$

$$B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)} = \frac{1}{\sqrt{n}} \sum_{j=[nt_{i-1}]+1}^{[nt_i]} X_j,$$

and the independence follows by the fact that $(X_j, j \geq 0)$ is an i.i.d. family. Even better, we have the identity in distribution for the i -th increment

$$B_{t_i}^{(n)} - B_{t_{i-1}}^{(n)} \stackrel{d}{=} \frac{\sqrt{[nt_i] - [nt_{i-1}]}}{\sqrt{n}} \frac{1}{\sqrt{[nt_i] - [nt_{i-1}]}} \sum_{j=1}^{[nt_i] - [nt_{i-1}]} X_j$$

and the central limit theorem shows that this converges in distribution to a Gaussian law $\mathcal{N}(0, \sigma^2(t_i - t_{i-1})I_d)$. Summing up our study, and introducing characteristic functions, we have shown that for every $\xi = (\xi_j, 1 \leq j \leq k)$,

$$\begin{aligned} \mathbb{E} \left[\exp \left(i \sum_{j=1}^k \xi_j (B_{t_j}^{(n)} - B_{t_{j-1}}^{(n)}) \right) \right] &= \prod_{j=1}^k \exp \left(i \xi_j (B_{t_j}^{(n)} - B_{t_{j-1}}^{(n)}) \right) \\ &\xrightarrow{n \rightarrow \infty} \prod_{j=1}^k \exp (i \xi_j (G_j - G_{j-1})) \\ &= \mathbb{E} \left[\exp \left(i \prod_{j=1}^k \xi_j (G_j - G_{j-1}) \right) \right], \end{aligned}$$

where G_1, \dots, G_k is distributed as in the statement of the proposition. By Lévy's convergence theorem we deduce that increments of $B^{(n)}$ between times t_i converge to increments of the sequence G_i , which is easily equivalent to the statement.

This suggests that $B^{(n)}$ should converge to a process B whose increments are independent and Gaussian with covariances dictated by the above formula. This will be set in a rigorous way later in the course, with Donsker's invariance theorem.

1.2 Wiener's theorem

We now start to define and study Brownian motion.

Definition 1.1 An \mathbb{R}^d -valued stochastic process $(B_t, t \geq 0)$ is called a standard Brownian motion if it is a continuous process, that satisfies the following conditions:

- (i) $B_0 = 0$ a.s.,
- (ii) for every $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_k$, the increments $(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_k} - B_{t_{k-1}})$ are independent, and
- (iii) for every $t, s \geq 0$, the law of $B_{t+s} - B_t$ is Gaussian with mean 0 and covariance sId .

The term “standard” refers to the fact that B_1 is normalized to have variance I_d , and the choice $B_0 = 0$.

The characteristic properties (i), (ii), (iii) exactly amount to say that the finite-dimensional marginals of a Brownian motion are given by the formula of Proposition 1.1. Therefore the law of the Brownian motion is uniquely determined. We now show Wiener’s theorem that Brownian motion exists!

Theorem 1.2 (Wiener) *There exists a Brownian motion on some probability space.*

We will first prove the theorem in dimension $d = 1$ and construct a process $(B_t, 0 \leq t \leq 1)$ satisfying the properties of a Brownian motion. This proof is essentially due to P. Lèvy in 1948. Before we start, we will need the following lemma, which is left as an exercise.

Lemma 1.3 *Let \mathcal{N} be a standard Gaussian random variable. Then*

$$\frac{x^{-1} - x^{-3}}{\sqrt{2\pi}} e^{-x^2/2} \leq P(\mathcal{N} > x) \leq \frac{x^{-1}}{\sqrt{2\pi}} e^{-x^2/2}. \quad (1.1)$$

Let $D_0 = \{0, 1\}$, $D_n = \{k2^{-n}, 0 \leq k \leq 2^n\}$ for $n \geq 1$, and $D = \cup_{n \geq 0} D_n$ be the set of dyadic rational numbers in $[0, 1]$. On some probability space (Ω, \mathcal{F}, P) , let $(Z_d, d \in D)$ be a collection of i.i.d. random variables all having a Gaussian distribution $\mathcal{N}(0, 1)$ with mean 0 and variance 1. It is a well-known and important fact that if the random variables X_1, X_2, \dots are linear combinations of independent centered Gaussian random variables, then X_1, X_2, \dots are independent if and only if they are pairwise uncorrelated, namely $\text{Cov}(X_i, X_j) = E[X_i X_j] = 0$ for every $i \neq j$.

We set $X_0 = 0$ and $X_1 = Z_1$. Inductively, given $(X_d^{n-1}, d \in D_{n-1})$, we build $(X_d^n, d \in D_n)$ in such a way that $(X_d^n, d \in D_n)$ satisfies (i), (ii), (iii) in the definition of the Brownian motion (where the instants under consideration are taken in D_n).

To this end, take $d \in D_n \setminus D_{n-1}$, and let $d_- = d - 2^{-n}$ and $d_+ = d + 2^{-n}$ so that d_-, d_+ are consecutive dyadic numbers in D_{n-1} . Then define:

$$X_d^n = \frac{X_{d_-}^{n-1} + X_{d_+}^{n-1}}{2} + \frac{Z_d}{2^{(n+1)/2}}.$$

and put $X_{d_-}^n = X_{d_-}^{n-1}$ and $X_{d_+}^n = X_{d_+}^{n-1}$. Note that with these definitions,

$$X_d^n - X_{d_-}^n = (X_{d_+}^{n-1} - X_{d_-}^{n-1})/2 + Z_d/2^{(n+1)/2}$$

and

$$X_{d_+}^n - X_d^n = (X_{d_+}^{n-1} - X_{d_-}^{n-1})/2 - Z_d/2^{(n+1)/2}.$$

It follows that

$$N_d := (X_{d_+}^{n-1} - X_{d_-}^{n-1})/2$$

and

$$N'_d = Z_d/2^{(n+1)/2}$$

are by the induction hypothesis two independent centered Gaussian random variables with variance 2^{-n-1} . From this, one deduces

$$\text{Cov}(N_d + N'_d, N_d - N'_d) = \text{Var}(N_d) - \text{Var}(N'_d) = 0,$$

so that the increments $X_d^n - X_{d_-}^n$ and $X_{d_+}^n - X_d^n$ are independent with variance 2^{-n} , as should be. Moreover, these increments are independent of the increments $X_{d'+2^{-n-1}}^n - X_{d'}^n$ for $d' \in D_{n-1}, d' \neq d_-$ and of $Z_{d'}, d' \in D_n \setminus D_{n-1}, d' \neq D$ so they are independent of the increments $X_{d''+2^{-n}} - X_{d''}$ for $d'' \in D_n, d'' \notin \{d_-, d\}$. This allows the induction argument to proceed one step further.

We have thus defined a process $(X_d^n, d \in D_n)$ which satisfies properties (i), (ii) and (iii) for all dyadic times $t_1, t_2, \dots, t_k \in D_n$. Observe that if $D \in D_n, X_d^m = X_d^n$ for all $m \geq n$. Hence for all $d \in D$,

$$B_d = \lim_{m \rightarrow \infty} X_d^m$$

is well-defined and the process $(B_d, d \in D)$ obviously satisfies (i), (ii) and (iii). To extend this to a process defined on the entire interval $[0, 1]$, we proceed as follows. Define, for each $n \geq 0$, a process $X_n(t), 0 \leq t \leq 1$ to be the linear interpolation of the values $(B_d, d \in D_n)$ the dyadic times at level n . Note that if $d \in D$, say $d \in D_m$ with $m \geq 0$, then for any $n \geq m$, $X_n(d) = X_m(d) = B_d$. Furthermore, define an event A_n by

$$A_n = \left\{ \sup_{0 \leq t \leq 1} |X_n(t) - X_{n-1}(t)| > 2^{-n/4} \right\}.$$

We then have, by Lemma 1.1, if N is a standard gaussian random variable:

$$\begin{aligned} \mathbb{P}(A_n) &= \mathbb{P} \left(\bigcup_{j=0}^{2^{n-1}} \sup_{t \in [(2j)2^{-n}, (2j+2)2^{-n}]} |X_n(t) - X_{n-1}(t)| > 2^{-n/4} \right) \\ &\leq \sum_{j=0}^{2^{n-1}} \mathbb{P} \left(\frac{|Z_{(2j+1)2^{-n}}|}{2^{(n+1)/2}} > 2^{-n/4} \right) \\ &\leq \sum_{k=0}^{2^{n-1}} \mathbb{P}(|\mathcal{N}| > 2^{(n+2)/4}) \\ &\leq \pi^{-1/2} 2^{3n/4} \exp(-2^{n/2}) \end{aligned}$$

We conclude that

$$\sum_{n=0}^{\infty} \mathbb{P}(A_n) < \infty$$

and by Borel-Cantelli, the events A_n occur only finitely often. We deduce immediately that the sequence of functions X_n is almost surely Cauchy in $C(0, 1)$ equipped with the topology of uniform convergence, and hence X_n converges toward a continuous limit function $(\tilde{X}(t), 0 \leq t \leq 1)$ uniformly, almost surely. Since $X_n(t)$ is constantly equal to X_t for $t \in D$ and for n large enough, it must be that $X_t = \tilde{X}(t)$ for all $t \in D$. Thus \tilde{X} is a continuous extension of X , and we still denote this extension by X . We now deduce properties (i), (ii) and (iii) for X by continuity and the fact that X_n satisfies these properties. Indeed, let $k \geq 1$ and let $0 < t_1 < t_2 \cdots < t_k < 1$. Fix $\alpha_1, \alpha_2, \dots, \alpha_k \geq 0$. For every $1 \geq i \geq k$, fix a sequence $(d_i^{(n)})_t^\infty$ such that $\lim_{n \rightarrow \infty} d_i^{(n)} = t_i$, and assume (since D is dense in $[0, 1]$) that $d^{(n)} \in D$ and $t_{i-1} < d_i^{(n)} \leq t_i$. Then by Lebesgue's dominated convergence theorem:

$$\begin{aligned}
& \mathbb{E}(e^{i\alpha_1 X_{t_1} + i\alpha_2 (X_{t_2} - X_{t_1}) + \dots + i\alpha_k (X_{t_k} - X_{t_{k-1}})}) \\
&= \lim_{n \rightarrow \infty} \mathbb{E}(e^{i\alpha_1 X_{d_1^{(n)}} + i\alpha_2 (X_{d_2^{(n)}} - X_{d_1^{(n)}}) + \dots + i\alpha_k (X_{d_k^{(n)}} - X_{d_{k-1}^{(n)}})}) \\
&= \lim_{n \rightarrow \infty} \exp\left(-\frac{\alpha_1^2}{2} d_1^{(n)} - \dots - \frac{\alpha_k^2}{2} (d_k^{(n)} - d_{k-1}^{(n)})\right) \\
&= \exp\left(-\frac{\alpha_1^2}{2} t_1 - \dots - \frac{\alpha_k^2}{2} (t_k - t_{k-1})\right)
\end{aligned}$$

It is now easy to construct a Brownian motion indexed by \mathbb{R}_+ . Simply take independent standard Brownian motions $(B_t^i, 0 \leq t \leq 1), i \geq 0$ as we just constructed, and let

$$B_t = \sum_{i=0}^{\lfloor t \rfloor - 1} B_1^i + B_{t - \lfloor t \rfloor}^{\lfloor t \rfloor}, \quad t \geq 0.$$

It is easy to check that this has the wanted properties. Finally, it is straightforward to build a Brownian motion in \mathbb{R}_d , by taking d independent copies B^1, \dots, B^d of B and checking that $((B_t^1, \dots, B_t^d), t \geq 0)$ is a Brownian motion in \mathbb{R}^d .

Remark. The extension of $(B_d, d \in D)$ could have been obtained by appealing to the existence of a continuous modification, whose existence is provided by Kolmogorov's criterion below.

1.3 Continuity and Hölder continuity of Brownian paths

In the last section we gave a construction of Brownian motion which directly yields a random process satisfying the three properties defining a Brownian motion, and which was at the same time continuous. In fact, and that is the reason why continuity is part of Definition 1.1, the next theorem will imply that any process satisfying (i), (ii) and (iii) can be slightly modified so that its trajectories are *a.s* continuous. The result is in fact much more general than that. As a consequence, we establish stronger regularity properties for Brownian motion than mere continuity: we prove that the path is almost surely Hölder with exponent $1/2 - \varepsilon$ for all $\varepsilon > 0$. To start with, we need to introduce the concept of version (modification) and indistinguishable versions.

Definition 1.2 If X and X' are two processes defined on some common probability space (Ω, \mathcal{F}, P) , we say that X' is a version of X if for every $t, P(X_t(\omega) = X'_t(\omega)) = 1$.

In particular, two versions X and X' of the same process share the same finite-dimensional distribution, however, this does *not* say that there exists an ω so that $X_t(\omega) = X'_t(\omega)$ for every t . This becomes true if both X and X' are *a priori* known to be continuous or cadlag, for instance. When the two trajectories coincide almost surely for all $t \geq 0$, we say that X and X' are indistinguishable:

Definition 1.3 If X and X' are two processes defined on some common probability space (Ω, \mathcal{F}, P) , we say that X' is an indistinguishable version of X if $P(X_t(\omega) = X'_t(\omega)) = 1$ for all t .

Note that, up to indistinguishability, there exists at most one continuous modification of a given process $(X_t, t \geq 0)$. Kolmogorov's criterion is a fundamental result which guarantees the existence of a continuous version (but not necessarily indistinguishable version) based solely on an L^p control of the two-dimensional distributions. We will apply to Brownian motion below, but it is useful in many other contexts.

Theorem 1.4 (Kolmogorov's continuity criterion) *Let $(X_t, 0 \leq t \leq 1)$ be a stochastic process with real values. Suppose there exist $p > 0, c > 0, \varepsilon > 0$ so that for every $s, t \geq 0$,*

$$E[|X_t - X_s|^p] \leq c|t - s|^{1+\varepsilon}.$$

Then, there exists a modification \tilde{X} of X which is a.s. continuous, and even α -Hölder continuous for any $\alpha \in (0, \varepsilon/p)$.

Proof. Let $D_n = \{k \cdot 2^{-n}, 0 \leq k \leq 2^n\}$ denote the dyadic numbers of $[0, 1]$ with level n , so D_n increases as n increases. Then letting $\alpha \in (0, \varepsilon/p)$, Markov's inequality gives for $0 \leq k < 2^n$,

$$P(|X_{k2^{-n}} - X_{(k+1)2^{-n}}| > 2^{-n\alpha}) E[|X_{k2^{-n}} - X_{(k+1)2^{-n}}|^p] \leq c2^{np\alpha}2^{-n-n\varepsilon} \leq c2^{-n}2^{-(\varepsilon-p\alpha)n}.$$

Summing over D_n we obtain

$$P\left(\sup_{0 \leq k < 2^n} |X_{k2^{-n}} - X_{(k+1)2^{-n}}| > 2^{-n\alpha}\right) \leq c2^{-n(\varepsilon-p\alpha)},$$

which is summable. Therefore, the Borel-Cantelli lemma shows that for a.a. ω , there exists N_ω so that if $n \geq N_\omega$, the supremum under consideration is $\leq 2^{-n\alpha}$. Otherwise said, a.s.,

$$\sup_{n \geq 0} \sup_{k \in \{0, \dots, 2^n - 1\}} \frac{|X_{k2^{-n}} - X_{(k+1)2^{-n}}|}{2^{-n\alpha}} \leq M(\omega) < \infty, \text{ a.s.}$$

We claim that this implies that for every $s, t \in D = \cup_{n \geq 0} D_n$, $|X_s - X_t| \leq M'(\omega)|t - s|^\alpha$, for some $M'(\omega) < \infty$ a.s.

Indeed, if $s, t \in D$, $s < t$, and let r is the least integer such that $t - s > 2^{-r-1}$. Then there exists $0 \leq k \leq 2^r$ and integers $l, m \geq 0$ such that

$$s = k2^{-r} - \varepsilon_1 2^{-r-1} - \dots - \varepsilon_l 2^{-r-l}$$

and

$$t = k2^{-r} + \varepsilon'_0 2^{-r} + \varepsilon'_1 2^{-r-1} + \dots + \varepsilon'_m 2^{-r-m}$$

with $\varepsilon_i, \varepsilon'_i \in \{0, 1\}$. For $0 \leq i \leq l$, let

$$s_i = k2^{-r} - \varepsilon_1 2^{-r-1} - \dots - \varepsilon_i 2^{-r-i}.$$

By the triangular inequality

$$\begin{aligned} |X_t - X_s| &= |X_{t_m} - X_{s_l}| \leq |X_{t_0} - X_{s_0}| + \sum_{i=1}^l |X_{t_i} - X_{t_{i-1}}| + \sum_{j=1}^m |X_{s_j} - X_{s_{j-1}}| \\ &\leq M(\omega)2^{-r\alpha} + \sum_{i=1}^l M(\omega)2^{-(r+i)\alpha} + \sum_{j=1}^m 2^{-(r+j)\alpha} M(\omega) \\ &\leq M(\omega)2^{-r\alpha}(1 + 2(1 - 2^{-\alpha})^{-1}) \\ &\leq M'(\omega)|t - s|^\alpha \end{aligned}$$

where $M'(\omega) = M(\omega)2^\alpha(1 + 2(1 - 2^{-\alpha})^{-1})$. Therefore, the process $(X_t, t \in D)$ is a.s. uniformly continuous (and even α -Hölder continuous). Since D is an everywhere dense set in $[0, 1]$, the latter process a.s. admits a unique continuous extension \tilde{X} on $[0, 1]$, which is also α -Hölder continuous (it is consistently defined by $\tilde{X}_t = \lim_n X_{t_n}$, where $(t_n, n \geq 0)$ is any D -valued sequence converging to t). On the exceptional set where $(X_d, d \in D)$ is not uniformly continuous, we let $\tilde{X}_t = 0, 0 \leq t \leq 1$, so \tilde{X} is continuous. It remains to show that \tilde{X} is a version of X . To this end, we estimate by Fatou's lemma

$$E[|X_t - \tilde{X}_t|^p] \leq \liminf_n E[|X_t - X_{t_n}|^p],$$

where $(t_n, n \geq 0)$ is any D -valued sequence converging to t . But since $E[|X_t - X_{t_n}|^p] \leq c|t - t_n|^{1+\varepsilon}$, this converges to 0 as $n \rightarrow \infty$. Therefore, $X_t = \tilde{X}_t$ a.s. for every t .

From now on we will consider exclusively a continuous modification of Brownian motion, which is unique up to indistinguishability. As a corollary to Kolmogorov's criterion, we obtain the aforementioned result on the Hölder properties of Brownian motion:

Corollary 1.5 *Let $(B_t, t \geq 0)$ be a standard Brownian motion in dimension 1. Almost surely, B is Hölder-continuous of order α for any $0 < \alpha < 1/2$. More precisely, with probability 1, for*

$$\sup_{n \leq t, s \leq n+1} \frac{|B_t - B_s|}{|t - s|^\alpha} < \infty. \quad (1.2)$$

Proof. Let $s \leq t \in D$, and notice that for every $p > 0$, since $B_t - B_s$ has the same law as $\sqrt{(t-s)N}$, (where N is a standard Gaussian random variable), we have $\mathbb{E}(|B_t - B_s|^p) \leq M|t-s|^{1+\varepsilon}$ with $\varepsilon = p/2 - 1$ and $M = \mathbb{E}(|N|^p) < \infty$. For $p > 2$, $\varepsilon > 0$ and thus X is Hölder of order α for $\alpha < \varepsilon/p = 1/2 - 1/p$. For instance, (1.2) holds with $\alpha = 1/2 - 1/p = 2p$ for all $p > 2$, and hence for any $\alpha < 1/2$ by taking a countable intersection of sets of probability 1.

Notice that the above corollary does not say anything about higher-order Hölder continuity: all we know is that the path is a.s. Hölder of order $\alpha < 1/2$. The next result tells us that this is, in some sense, sharp.

Theorem 1.6 *Let B be a continuous modification of Brownian motion. Let $\gamma > 1/2$. Then it holds:*

$$\mathbb{P}\left(\forall t \geq 0 : \limsup_{h \rightarrow 0^+} \frac{|B_{t+h} - B_t|}{h^\gamma} = +\infty\right) = 1$$

Proof. We first observe that

$$\begin{aligned} & \left\{ \exists t \geq 0 : \limsup_{h \rightarrow 0^+} \frac{|B_{t+h} - B_t|}{h^\gamma} < \infty \right\} \\ & \subseteq \bigcup_{p=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \left\{ \exists t \in [0, m] : |B_{t+h} - B_t| \leq ph^\gamma, \forall h \in (0, 1/k) \right\}. \end{aligned}$$

Therefore, it suffices to show that

$$\mathbb{P}(\exists t \in [0, m] : |B_{t+h} - B_t| \leq ph^\gamma, \forall h \in (0, \delta)) = 0$$

for all $p \geq 1, m \geq 1, \delta > 0$. For $n \geq 1, 1 \leq i \leq mn - 1$, define:

$$A_{i,n} = \left\{ \exists s \in \left[\frac{i}{n}, \frac{i+1}{n} \right] : |B_{s+h} - B_s| \leq ph^\gamma, \forall h \in (0, \delta) \right\}.$$

It suffices to show:

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{mn-1} \mathbb{P}(A_{i,n}) = 0 \quad (1.3)$$

Fix a large constant $K > 0$ to be chosen suitably later. We wish to exploit the fact that on the event $A_{i,n}$ many increments must be small. The trick is to be able to fix in advance the times at which these increments will be too small. More precisely, on $A_{i,n}$, as long as $n \geq (K+1)/\delta$, for all $1 \leq j \leq K$:

$$0 \leq \frac{i+j}{n} - s \leq \frac{K+1}{n} \leq \delta$$

where s is as in the definition of $A_{i,n}$. Thus, taking $h = (i+j)/n - s$, on $A_{i,n}$:

$$\left| B_{\frac{i+j}{n}} - B_s \right| \leq p \left(\frac{i+j}{n} - s \right)^\gamma \leq p \left(\frac{K+1}{n} \right)^\gamma$$

If $2 \leq j \leq K$, by the triangular inequality:

$$\left| B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}} \right| \leq 2p \left(\frac{K+1}{n} \right)^\gamma$$

Therefore, there exists $C > 0$ such that for all $n \geq (K+1)/\delta$

$$\begin{aligned} \mathbb{P}(A_{i,n}) &\leq \mathbb{P} \left(\bigcap_{j=2}^K \left\{ \left| B_{\frac{i+j}{n}} - B_{\frac{i+j-1}{n}} \right| \leq 2p \left(\frac{K+1}{n} \right)^\gamma \right\} \right) \\ &\leq \prod_{j=2}^K \mathbb{P} \left(|\mathcal{N}(0, 1/n)| \leq 2p \left(\frac{K+1}{n} \right)^\gamma \right) \\ &\leq \left[\mathbb{P} \left(|\mathcal{N}(0, 1)| \leq 2p \left(\frac{K+1}{n} \right)^\gamma n^{1/2} \right) \right]^{K-1} \\ &\leq \left[2p \left(\frac{K+1}{n} \right)^\gamma n^{1/2} \right]^{K-1} = \frac{C}{n^{(\gamma-1/2)(K-1)}} \end{aligned}$$

It follows that for all $n \geq (K+1)/\delta$:

$$\sum_{i=0}^{mn-1} \mathbb{P}(A_{i,n}) \leq \frac{Cm}{n^{(\gamma-1/2)(K-1)-1}}$$

Thus if K is large enough that $(\gamma-1/2)(K-1) > 1$, the right-hand side tends to 0 for all $n \geq n_0 := \lceil (K+1/\delta) \rceil + 1$. This proves (1.3), and, as a consequence, Theorem 1.4.

As a corollary to the last result, we obtain the celebrated Paley-Wiener-Zygmund theorem:

Corollary 1.7 *Almost surely, $t \mapsto B_t$ is nowhere differentiable*

1.4 Basic properties

Let $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ be the ‘Wiener space’ of continuous functions, endowed with the product σ -algebra \mathcal{W} (or the Borel σ -algebra associated with the compact-open topology).

Definition 1.4 (Wiener’s measure) Let $(B_t, t \geq 0)$ be a Brownian motion, and let \mathbb{W} be the law on Ω of B : that is, for any $A \in \mathcal{W}$,

$$\mathbb{W}(A) = P((B_t, t \geq 0) \in A)$$

\mathbb{W} is called Wiener’s measure.

Of course, the point is that \mathbb{W} does not depend on B . We now think of Ω as our probability space. For $\omega \in \Omega$ define:

$$X_t(\omega) = \omega(t), t \geq 0$$

to be the canonical process. Then $(X_t, t \geq 0)$, under the probability measure \mathbb{W} , is a Brownian motion. This is the canonical construction of Brownian motion. (This construction will come particularly handy when we discuss applications to Girsanov’s theorem for change of measures.) For $x \in \mathbb{R}^d$ we also let $\mathbb{W}_x(dw)$ be the image measure of \mathbb{W} by $(w_t, t \geq 0) \mapsto (x + w_t, t \geq 0)$. A (continuous) process with law $\mathbb{W}_x(dw)$ is called a Brownian motion started at x . We let $(\mathcal{F}_t^B, t \geq 0)$ be the natural filtration of $(B_t, t \geq 0)$, completed by zero-probability events.

We now state some fundamental results, which are often referred to as the scaling properties of Brownian motion, or scale-invariance of Brownian motion. They are easy to prove and are left as an exercise.

Proposition 1.8 *Let B be a standard Brownian motion in \mathbb{R}^d .*

1. *Rotational invariance: If $U \in O(d)$ is an orthogonal matrix, then $UB = (UB_t, t \geq 0)$ is again a Brownian motion. In particular, $-B$ is a Brownian motion.*
2. *Scaling property: If $\lambda > 0$ then $(\lambda^{-1/2}B_{\lambda t}, t \geq 0)$ is a standard Brownian motion*
3. *Time-inversion: $(tB_{1/t}, t \geq 0)$ is also a Brownian motion (at $t = 0$ the process is defined by its value 0).*

We now start to discuss ideas revolving around the Markov property of Brownian motion and its applications to path properties. We begin with the simple Markov property, which takes a particularly nice form in this context.

Theorem 1.9 *Let $(B_t, t \geq 0)$ be a Brownian motion, and let $s > 0$. Then*

$$(\tilde{B}_t := B_{t+s} - B_s, t \geq 0)$$

is a Brownian motion, independent of the σ -field $\mathcal{F}_{s+}^B = \bigcap_{t>s} \mathcal{F}_t^B$.

Proof. Since \tilde{B} is continuous and $\tilde{B}_0 = 0$, to show that \tilde{B} is a Brownian motion it suffices to check that the increments have the correct distribution. However if $t \geq u$, $\tilde{B}_t - \tilde{B}_u = B_{s+t} - B_{s+u}$ so this follows directly from the fact that B itself is a Brownian motion. It remains to show that \tilde{B} is independent from \mathcal{F}_{s+} . We start by checking independence with respect to \mathcal{F}_s , for which we can assume $d = 1$. We will use this easy lemma, which is an important property worth remembering:

Lemma 1.10 *Let $s, t \geq 0$. Then*

$$\text{Cov}(B_s, B_t) = s \wedge t.$$

Now, to prove independence of \tilde{B} with respect to \mathcal{F}_s , it suffices to check that the finite-dimensional marginals are independent: i.e., if $s_1 \leq \dots \leq s_m \leq s$ and $t_1 \leq \dots \leq t_n$, we want to show that

$$(B_{s_1}, \dots, B_{s_m}) \text{ and } (\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n})$$

are independent. However, the $m+n$ -coordinate vector $(B_{s_1}, \dots, B_{s_m})$ and $(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n})$ is a Gaussian vector (since it is the image by a linear application of a Gaussian vector), and it suffices to check that the covariance of two distinct terms is 0. Since each term has zero expectation:

$$\begin{aligned} \text{Cov}(\tilde{B}_{t_i}, B_{s_j}) &= \mathbb{E}(\tilde{B}_{t_i} B_{s_j}) \\ &= \mathbb{E}(B_{s+t_i} B_{s_j}) - \mathbb{E}(B_s B_{s_j}) \\ &= s_j \wedge (s + t_i) - (s_j \wedge s) = s_j - s_j = 0 \end{aligned}$$

which proves the independence with respect to \mathcal{F}_s . If $A \in \mathcal{F}_{s^+}$, we wish to show that for every continuous functional $F : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ continuous and bounded,

$$\mathbb{E}(1_{\{A\}} F(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_k})) = \mathbb{P}(A) \mathbb{E}(F(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_k}))$$

Now, for any $\varepsilon > 0$, $A \in \mathcal{F}_{s^+} \subseteq \mathcal{F}_{s+\varepsilon}$, thus, using the property just proved:

$$\mathbb{E}(1_A F(B_{t_1+s+\varepsilon}-B_{s+\varepsilon}, \dots, B_{t_k+s+\varepsilon}-B_{s+\varepsilon})) = \mathbb{P}(A) \mathbb{E}(F(B_{t_1+s+\varepsilon}-B_{s+\varepsilon}, \dots, B_{t_k+s+\varepsilon}-B_{s+\varepsilon}))$$

Letting $\varepsilon \rightarrow 0$ in the above identity, since B is continuous and F is bounded and continuous, we have (by Lebesgue's dominated convergence theorem),

$$\mathbb{E}(1_{\{A\}} F(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_k})) = \mathbb{P}(A) \mathbb{E}(F(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_k}))$$

as required.

Theorem 1.11 (Blumenthal's zero-one law) *Let B be a standard Brownian motion. The σ -algebra $\mathcal{F}_{0^+}^B = \bigcap_{\varepsilon > 0} \mathcal{F}_\varepsilon^B$ is trivial i.e. constituted of events of probability 0 or 1.*

Proof. By the previous result, $(B_t, t \geq 0)$ is independent from \mathcal{F}_{0^+} . However \mathcal{F}_∞^B contains $\mathcal{F}_{0^+}^B$, so this implies that the σ -field \mathcal{F}_{0^+} is independent of itself, and $P(A) = P(A \cap A) = P(A)^2$ by independence. Thus $P(A)$ is solution to the equation $x = x^2$ whose roots are precisely 0 and 1.

Proposition 1.12 (i) *For $d = 1$ and $t \geq 0$, let $S_t = \sup_{0 \leq s \leq t} B_s$ and $I_t = \inf_{0 \leq s \leq t} B_s$ (these are random variables because B is continuous). Then almost surely, for every $\varepsilon > 0$, one has*

$$S_\varepsilon > 0 \quad \text{and} \quad I_\varepsilon < 0$$

In particular, a.s. there exists a zero of B in any interval of the form $(0, \varepsilon)$, $\varepsilon > 0$

(ii) A.s.,

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

(iii) Let C be an open cone in \mathbb{R}^d with non-empty interior and origin at 0 (i.e., a set of the form $\{tu : t > 0, u \in A\}$, where A is a non-empty open subset of the unit sphere of \mathbb{R}^d). If

$$H_C = \inf\{t > 0 : B_t \in C\}$$

is the first hitting time of C , then $H_C = 0$ a.s.

Proof. (i) The probability that $B_t > 0$ is $1/2$ for every t , so $P(S_t > 0) \geq 1/2$, and therefore if $t_n, n \geq 0$ is any sequence decreasing to 0, $P(\limsup_n \{B_{t_n} > 0\}) \geq \limsup_n P(B_{t_n} > 0) = 1/2$. Since the event $\limsup_n \{B_{t_n} > 0\}$ is in \mathcal{F}_{0+} , Blumenthal's law shows that its probability must be 1. The same is true for the infimum by considering the Brownian motion $-B$.

(ii) Let $S_\infty = \sup_{t \geq 0} B_t$. By scaling invariance, for every $\lambda > 0$, $\lambda S_\infty = \sup_{t > 0} \lambda B_t$ has same law as $\sup_{t \geq 0} B_{\lambda^2 t} = S_\infty$. This is possible only if $S_\infty \in \{0, \infty\}$ a.s., however, it cannot be 0 by (i).

(iii) The cone C is invariant by multiplication by a positive scalar, so that $P(B_t \in C)$ is the same as $P(B_1 \in C)$ for every t by the scaling invariance of Brownian motion. Now, if C has nonempty interior, it is straightforward to check that $P(B_1 \in C) > 0$, and one concludes similarly as above. Details are left to the reader.

We now want to prove an important analog of the simple Markov property, where deterministic times are replaced by stopping times. To begin with, we extend a little the definition of Brownian motion, by allowing it to start from a random location, and by working with filtrations that are (slightly) larger than the natural filtration of a standard Brownian motion.

Definition 1.5 We say that B is a Brownian motion (started at X) if $(B_t - X, t \geq 0)$ is a standard Brownian motion which is independent of X .

Otherwise said, it is the same as the definition as a standard Brownian motion, except that we do not require that $B_0 = 0$. If we want to express this on the Wiener space with the Wiener measure, we have for every measurable functional $F : \Omega_W \rightarrow \mathbb{R}_+$,

$$E[F(B_t, t \geq 0)] = E[F(B_t - B_0 + X, t \geq 0)],$$

and since $(B_t - B_0)$ has law \mathbb{W} and since X is independent from $(B_t - B_0)$, this is

$$\int_{\mathbb{R}^d} P(X \in dx) \int_{\Omega} \mathbb{W}(dw) F(x + w(t), t \geq 0) = \int_{\mathbb{R}^d} P(X \in dx) W_x(F) = E[\mathbb{W}_X(F)],$$

where $\mathbb{W}_X(F)$ is the random variable $\omega \mapsto W_{B_0(\omega)}(F)$.

Definition 1.6 Let $(\mathcal{F}_t, t \geq 0)$ be a filtration. We say that a Brownian motion B is an (\mathcal{F}_t) -Brownian motion if B is adapted to (\mathcal{F}_t) , and if $B^{(t)} = (B_{t+s} - B_t, s \geq 0)$ is independent of \mathcal{F}_t for every $t \geq 0$.

For instance, if (\mathcal{F}_t) is the natural filtration of a 2-dimensional Brownian motion $(B_t^1, B_t^2, t \geq 0)$, then $(B_t^1, t \geq 0)$ is an (\mathcal{F}_t) -Brownian motion. If B' is a standard Brownian motion and X is a random variable independent of B' , then $B = (X + B'_t, t \geq 0)$ is a Brownian motion (started at $B_0 = X$), and is an $(\mathcal{F}_t^B) = (\sigma(X) \vee \mathcal{F}_t^{B'})$ -Brownian motion. A Brownian motion is always an \mathcal{F}_t^B -Brownian motion. If B is a standard Brownian motion, then the completed filtration $\mathcal{F}_t = \mathcal{F}_t^B \vee \mathcal{N}$ (\mathcal{N} being the set of events of probability 0) can be shown to be right-continuous, i.e. $\mathcal{F}_{t+} = \mathcal{F}_t$ for every $t \geq 0$, and B is an (\mathcal{F}_t) -Brownian motion.

Definition 1.7 Let \mathcal{F} be a filtration and let T be a stopping time. The σ -field \mathcal{F}_T is defined by

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}$$

It is elementary (but tedious) that in the case of filtration generated by a process X , $\mathcal{F}_T = \sigma(X_{s \wedge T}, s \geq 0)$. In particular T and X_T are \mathcal{F}_T -measurable. This corroborates the intuition that \mathcal{F}_T is the σ -algebra generated by all the events occurring prior to time T . We may now state the strong Markov property.

Theorem 1.13 (Strong Markov property) *Let $(B_t, t \geq 0)$ be an (\mathcal{F}_t) -Brownian motion in \mathbb{R}^d and T be an (\mathcal{F}_t) -stopping time. We let $B_t^{(T)} = B_{T+t} - B_T$ for every $t \geq 0$ on the event $\{T < \infty\}$, and 0 otherwise. Conditionally on $\{T < \infty\}$, the process $B^{(T)}$ is a standard Brownian motion, which is independent of \mathcal{F}_T . Otherwise said, conditionally given \mathcal{F}_T and $\{T < \infty\}$, the process $(B_{T+t}, t \geq 0)$ is an (\mathcal{F}_{T+t}) -Brownian motion started at B_T .*

Proof. Suppose first that $T < \infty$ a.s. Let $A \in \mathcal{F}_T$, and consider times $t_1 < t_2 < \dots < t_p$. We want to show that for every bounded continuous function F on $(\mathbb{R}^d)^p$,

$$E[\mathbf{1}_{\{A\}} F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})] = P(A) E[F(B_{t_1}, \dots, B_{t_p})]. \quad (1.4)$$

Indeed, taking $A = \Omega$ entails that $B^{(T)}$ is a Brownian motion, while letting A vary in \mathcal{F}_T entails the independence of $(B_{t_1}^{(T)}, \dots, B_{t_k}^{(T)})$ and \mathcal{F}_T for every t_1, \dots, t_k , hence of $B^{(T)}$ and \mathcal{F}^T .

$$\begin{aligned} E[\mathbf{1}_{\{A\}} F(B_{t_1}^{(T)}, \dots, B_{t_p}^{(T)})] &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} E[\mathbf{1}_{\{A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}\}} F(B_{t_1}^{(k2^{-n})}, \dots, B_{t_p}^{(k2^{-n})})] \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} P(A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\}) E[F(B_{t_1}, \dots, B_{t_p})] \\ &= P(A) E[F(B_{t_1}, \dots, B_{t_p})]. \end{aligned}$$

where we used the simple Markov property and the fact that $A \cap \{(k-1)2^{-n} < T \leq k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$ by definition. Finally, if $P(T = \infty) > 0$, check that (1.4) remains true when replacing A by $A \cap \{T < \infty\}$, and divide by $P(\{T < \infty\})$.

An important example of application of the strong Markov property is the so-called reflection principle. Recall that $S_t = \sup_{0 \leq s \leq t} B_s$.

Theorem 1.14 (Reflection principle) *Let $0 < a$ and $b \leq a$, then for every $t \geq 0$,*

$$P(S_t \geq a, B_t \leq b) = P(B_t \geq 2a - b).$$

Proof. Let $T_a = \inf\{t \geq 0 : B_t \geq a\}$ be the first entrance time of B_t in $[a, \infty)$ for $a > 0$. Then T_a is an (\mathcal{F}_t^B) -stopping time for every a and $T_a < \infty$ a.s. since $S_\infty = \infty$ a.s. where $S_\infty = \lim_{t \rightarrow \infty} S_t$.

Now by continuity of B , $B_{T_a} = a$ for every a . We thus have:

$$\begin{aligned} P(S_t \geq a, B_t \leq b) &= P(T_a \leq t, B_t \leq b) \\ &= P(T_a \leq t, B_{t-T_a}^{(T_a)} \leq b - a) \\ &= P(T_a \leq t, -B^{(T_a)} \geq a - b). \end{aligned}$$

Now, by the strong Markov property at time T_a , B^{T_a} is a Brownian motion independent of \mathcal{F}_{T_a} and thus of T_a . In particular, we deduce that the joint law of $(T_a, B^{(T_a)})$ is identical to the joint law of $(T_a, -B^{(T_a)})$, by symmetry of Brownian motion. It follows that

$$\begin{aligned} P(S_t \geq a, B_t \leq b) &= P(T_a \leq t, -B^{(T_a)} \geq a - b) \\ &= P(T_a \leq t, B_t \geq 2a - b) \\ &= P(B_t \geq 2a - b). \end{aligned}$$

Corollary 1.15 *We have the following identities in distribution: for all $t \geq 0$,*

$$S_t \stackrel{d}{=} |B_t| \stackrel{d}{=} |\mathcal{N}(0, t)|.$$

Moreover, for every $x > 0$, the random time T_x has same law as $(x/B_1)^2$.

Proof. We write, for all $t \geq 0$ and all $a \geq 0$,

$$\begin{aligned} P(S_t \geq a) &= P(S_t \geq a, B_t \leq a) + P(S_t \geq a, B_t \geq a) \\ &= P(B_t \geq 2a - a) + P(B_t \geq a) \\ &= 2P(B_t \geq a) = P(|B_t| \geq a) \end{aligned}$$

since when $B_t \geq a$, $S_t \geq a$ automatically as well. We leave the computation of the distribution of T_x as an exercise (cf. Example Sheet 1).

We end with a famous result of P. Lévy on the quadratic variation of Brownian motion. This result plays a fundamental role in the development of the stochastic integral. Let $(B_t, t \geq 0)$ be a standard Brownian motion. Let $t > 0$ be fixed and for $n \geq 1$ let $\Delta_n = \{0 = t_0(n) < t_1(n) < \dots < t_{m_n}(n) := t\}$ be a subdivision of $[0, t]$, such that

$$\eta_n := \max_{1 \leq i \leq m_n} (t_i(n) - t_{i-1}(n)) \xrightarrow{n \rightarrow \infty} 0.$$

Theorem 1.16 (Lévy)

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{m_n} (B_{t_i} - B_{t_{i-1}})^2 = t$$

The proof is part of Example Sheet 1. (in fact, this convergence holds almost surely as soon as the subdivisions are nested, but this is much more difficult).

2 Preliminaries

2.1 The bigger picture, in a nutshell

We will now spend rather a lot of time to give a precise and rigorous construction of the stochastic integral, for as large a class of processes as possible, subject to continuity. This level of generality has a price, which is that the construction can appear quite technical without shedding any light on the sort of processes we are talking about. So before we embark on this journey, here are a few points which, in my opinion, guide the whole construction and should also guide your intuition throughout. What follows is only informal, and in particular, we do not describe issues related to measurability, and finiteness of the integral.

The real difficulty in the construction of the integral is in how to make sense of

$$\int_0^t H_s dM_s \tag{2.1}$$

where M is a martingale and H is, say, left-continuous or continuous. Even though dM does not make sense as a measure (the paths of martingales, just like Brownian paths, have too wild oscillations for that), it is easy to cook up a definition which makes intuitive sense when H is a *simple process*, that is, H takes only finitely many (bounded) values. Indeed, it suffices to require that the integral process in (2.1) varies in the same way as M on the intervals over which H is constant, and has jumps when H does. A natural approach is then to try to extend this definition to more general classes of processes by “taking a limit” of integrals

$$\int_0^t H_s^n dM_s \rightarrow \int_0^t H_s dM_s \tag{2.2}$$

where the integrands in the left-hand side are simple and approximate H .

In implementing this method, one faces several technical difficulties. The strategy is to construct a suitable function space where the sequence on the left-hand side of (2.2) forms a Cauchy sequence. If the function space is complete, the sequence of integrals has a limit, which we may call the integral of H with respect to M . But we must also guarantee that this limit does not depend on the approximating sequence. It remains to find a space which has the desired properties. The key property which we will use (over and over again) is that martingales have a finite quadratic variation:

$$[M]_t := \lim_{n \rightarrow \infty} \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2 \tag{2.3}$$

exists and is finite, and is nondecreasing in t . Furthermore, one can show (Theorem 3.5) that $M_t^2 - [M]_t$ is a martingale. Now, when H is simple, it is not hard to convince yourself that the integral (2.1) must also be a martingale. So what should be the quadratic variation of $\int_0^t H_s dM_s$? Based on the approximation (2.3), the amount of quadratic variation that we add to the integral between t and $t + dt$ is approximately $H_t^2 d[M]_t$. Hence any sensible definition of the stochastic integral must satisfy

$$\left[\int_0^\cdot H_s dM_s \right]_t = \int_0^t H_s^2 d[M]_s \tag{2.4}$$

The key insight of Itô was the realization that this property was sufficient to define the integral. Indeed, using the optional stopping theorem, this is essentially the same as requiring:

$$\mathbb{E} \left(\left(\int_0^\infty H_s dM_s \right)^2 \right) = \mathbb{E} \left(\int_0^\infty H_s^2 d[M]_s \right). \quad (2.5)$$

Interpreting the right-hand side as an L^2 norm on the space of bounded integrands, this statement is saying that the stochastic integral must be a certain isometry between Hilbert spaces. The left-hand side shows that the correct space of martingales is the set of martingales with $\mathbb{E}([M]_\infty^2) < \infty$, or, equivalently (as it turns out), martingales which are bounded in L^2 . This space, endowed with the norm on the left-hand side of (2.5) is indeed complete and simple processes are dense in it. Formula (2.5) is then relatively easy to prove for simple processes. This implies, at once, that the sequence in the left-hand side of (2.2) is Cauchy (and hence has a limit), and the isometry property shows that this limit cannot depend on the approximating sequence.

At this point we have finished the construction of the stochastic integral for martingales which are bounded in L^2 . Stopping at suitable stopping times, it is then easy to extend this definition to general martingales, or indeed to processes known as *local martingales*. Adding a “finite variation” component for which the integral (2.1) is defined as a good old Stieltjes-Lebesgue integral finishes the construction for *semi-martingales*.

Having spoken about the bigger picture in a nutshell, it is now time to rewind the tape and go back to the beginning.

2.2 Finite variation integrals

Finite variation processes are essentially those for which the standard notion of integral (the one you learn about in measure theory courses) is well-defined. Since finite variation is a pathwise property, we will first establish integrals with respect to deterministic integrands and lift it to stochastic processes in the last part of this section.

Recall that a function $f : \mathbb{R} \rightarrow \mathbb{R}$ is *càdlàg* or *rcll* if it is right-continuous and has left limits. For such functions we write $\Delta f(t) := f(t) - f(t-)$ where $f(t-) = \lim_{s \uparrow t} f(s)$. Suppose $a : [0, \infty) \rightarrow \mathbb{R}$ is an increasing càdlàg function. Then there exists a unique Borel measure da on $(0, \infty)$ such that $da((s, t]) = a(t) - a(s)$, the *Lebesgue-Stieltjes measure* with distribution function a . Since da is a proper measure, there is no problem in defining, for any non-negative measurable function h and $t \geq 0$:

$$(h \cdot a)(t) = \int_{(0, t]} h(s) da(s). \quad (2.6)$$

We may extend this definition to a càdlàg function $a = a' - a''$, where a' and a'' are both increasing càdlàg, and to integrable $h : [0, \infty) \rightarrow \mathbb{R}$. Subject to the finiteness of all the terms on the right we define

$$h \cdot a = h^+ \cdot a' - h^+ \cdot a'' - h^- \cdot a' + h^- \cdot a''. \quad (2.7)$$

where $h^\pm := \max\{\pm h, 0\}$ are the positive and negative part of h .

To be able to make this definition we have assumed that a was the difference between two nondecreasing functions. We now ask for an analytic characterization of those functions which have this property. If a is a measurable function and I an interval, we define (with a slight abuse of notation) $da(I) := a(\sup I) - a(\inf I)$, even though da is not really a measure.

Lemma 2.1 *Let $a : [0, \infty) \rightarrow \mathbb{R}$ be càdlàg and define $v^n(0) = 0$, and for all $t > 0$*

$$v^n(t) = \sum_{k=0}^{\lceil 2^n t \rceil - 1} |a((k+1)2^{-n}) - a(k2^{-n})|. \quad (2.8)$$

Then $v(t) := \lim_{n \rightarrow \infty} v^n(t)$ exists for all $t \geq 0$ and is nondecreasing in t .

Proof. Let $t_n^+ = 2^{-n} \lceil 2^n t \rceil$ and $t_n^- = 2^{-n} (\lceil 2^n t \rceil - 1)$ and write

$$v^n(t) = \sum_{\substack{I \in \Delta_n \\ \inf I < t}} |da(I)| = \sum_{\substack{I \in \Delta_n \\ \sup I < t}} |da(I)| + |a(t_n^+) - a(t_n^-)|. \quad (2.9)$$

where $\Delta_n = \{(k2^{-n}, (k+1)2^{-n}) : k \in \mathbb{N}\}$. The first term is nondecreasing in n by the triangle inequality, and so has a limit as $n \rightarrow \infty$. The second converges to $|\Delta a(t)| = |a(t) - a(t-)|$ as a is càdlàg, and so $v(t)$ exists for all $t \geq 0$. Since $v^n(t)$ is nondecreasing in t for all n , the same holds for $v(t)$. \square

Definition 2.1 $v(t)$ is called the *total variation of a over $(0, t]$* and a is said to be of *finite variation* if $v(t) < \infty$ for all $t \geq 0$.

Proposition 2.2 *A càdlàg function $a : [0, \infty) \rightarrow \mathbb{R}$ can be expressed as $a = a' - a''$, with a', a'' increasing and càdlàg, iff a is of finite variation. In this case, $t \mapsto v(t)$ is càdlàg with $\Delta v(t) = |\Delta a(t)|$ and $a^\pm := \frac{1}{2}(v \pm a)$ are the smallest functions a' and a'' with that property.*

Proof. Suppose $v(t) < \infty$ for all $t \geq 0$.

Step 1: let us show that v is cadlag. Fix $T > 0$ and consider

$$u^n(t) = \sum_{\substack{I \in \Delta_n \\ t < \inf I < \sup I < T}} |da(I)| \quad \text{for } t \leq T. \quad (2.10)$$

$u^n(t)$ is clearly non-increasing in t , and it is easy to see that it is also right-continuous. These two properties imply that $\{t \in [0, T] : u^n(t) \leq x\}$ is closed for all $x \geq 0$. Now, just as for $v^n(t)$, the sum defining $u^n(t)$ is nondecreasing in n by the triangular inequality. Thus for all $t \geq 0$, $u^n(t)$ has a limit as $n \rightarrow \infty$ which we may call $u(t)$. We have that

$$\{t \in [0, T] : u(t) \leq x\} = \bigcap_{n \in \mathbb{N}} \{t \in [0, T] : u^n(t) \leq x\} \quad (2.11)$$

is closed as a countable intersection of closed sets. This, together with the fact that $u(t)$ is non-increasing in t , implies that u is right-continuous. Furthermore, observe that for all $t < T$:

$$v^n(T) = v^n(t) + u^n(t) + |a(T_n^+) - a(T_n^-)|, \quad (2.12)$$

The final term on the right converges to $|\Delta a(T)|$ as $n \rightarrow \infty$ because a is right-continuous. Hence for all $t < T$ we have $v(t) = v(T) - u(t) + |\Delta a(T)|$ and since T was arbitrary, v is right-continuous. v has left limits since it is nondecreasing, and taking the limit $n \rightarrow \infty$ in (2.9) we get $v(t) = v(t-) + |\Delta a(t)|$.

Step 2: $v = a^+ - a^-$. Having made these observations, define two functions a^+ and a^- by

$$a^+ = \frac{1}{2}(v + a) \quad \text{and} \quad a^- = \frac{1}{2}(v - a) \quad \text{are càdlàg.} \quad (2.13)$$

Since v is càdlàg, then a^+ and a^- are also càdlàg. It thus suffices to prove that they are non-decreasing. However, note that for each $m \in \mathbb{Z}^+$,

$$\begin{aligned} dv^m(I) &= |da(I)| \quad \text{for all } I \in \Delta_m \\ \text{and } dv^n(I) &\geq |da(I)| \quad \text{for all } I \in \Delta_m \text{ if } n \geq m. \end{aligned} \quad (2.14)$$

Thus $da^\pm(I) = \frac{1}{2}dv(I) \pm \frac{1}{2}da(I) \geq 0$ for all $I \in \bigcup_{m \geq 1} \Delta_m$, and so it follows easy (by right-continuity) that a^+ and a^- are nondecreasing.

Step 3: minimality. Suppose now $a = a' - a''$ where a', a'' are nondecreasing with $a(0) = a'(0) = a''(0) = 0$ without loss of generality. Then for any $I \in \Delta_n, n \geq 0$,

$$|da(I)| \leq da'(I) + da''(I). \quad (2.15)$$

Summing over $I \in \Delta_n$ with $\sup I < t$ in (2.9), the terms in the sum telescope and we obtain

$$v^n(t) \leq a'(t_n^+) + a''(t_n^+). \quad (2.16)$$

Letting $n \rightarrow \infty$, the left-hand side converges to $v(t)$ by definition, and the right-hand side converges to $a'(t) + a''(t)$ since a' and a'' are right-continuous. Note that we can also write $v(t) = a^+(t) + a^-(t)$ and hence the last inequality shows

$$a^+(t) + a^-(t) \leq a'(t) + a''(t)$$

for all $t \geq 0$. Adding and subtracting $a = a^+ - a^- = a' - a''$ on both sides we get $a^+(t) \leq a'(t)$ and $a^-(t) \leq a''(t)$ for all $t \geq 0$, as required.

Converse: Assume that $a = a' - a''$ for two cadlag nondecreasing functions a', a'' , and let us show that $v(t) < \infty$. This is the easy direction: reasoning as in (2.16) we obtain that $v^n(t) \leq a'(t_n) + a''(t_n)$ and so letting $n \rightarrow \infty$, we get $v(t) \leq a'(t) + a''(t)$, as required. \square

Exercise. Let B be a standard Brownian motion. Show that B has infinite variation almost surely. (*Hint:* otherwise the quadratic variation would be finite.) Thus to construct an integral with respect to B one must create something genuinely new.

Suppose now that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Recall that a process $X : \Omega \times [0, \infty) \rightarrow \mathbb{R}$ is *adapted* to $(\mathcal{F}_t)_{t \geq 0}$ if $X_t = X(\cdot, t)$ is \mathcal{F}_t -measurable for all $t \geq 0$, and X is *càdlàg* if $X(\omega, \cdot)$ is càdlàg for all $\omega \in \Omega$.

Definition 2.2 Let A be a càdlàg adapted process. Its *total variation process* V is defined pathwise (for each $\omega \in \Omega$) as the total variation of $A(\omega, \cdot)$. We say that A is of *finite variation* if $A(\omega, \cdot)$ is of finite variation for all $\omega \in \Omega$.

Lemma 2.3 Let A be a càdlàg adapted process with finite total variation V . Then V is càdlàg adapted and pathwise nondecreasing.

Proof. Using the same partition as in (2.9) we get

$$V_t = \lim_{n \rightarrow \infty} \tilde{V}_t^n + |\Delta A_t| \quad (2.17)$$

where $\tilde{V}_t^n := \sum_{k=0}^{2^n t_n^- - 1} |A_{(k+1)2^{-n}} - A_{k2^{-n}}|$ is adapted for all $n \in \mathbb{N}$ since $t_n^- < t$ and ΔA_t is \mathcal{F}_t -measurable since A is càdlàg adapted. Thus V is adapted and it is càdlàg and increasing because $V(\omega, \cdot)$ is càdlàg and increasing for all $\omega \in \Omega$. \square

In the next section we will introduce a suitable class of integrands H for a pathwise definition of the stochastic integral

$$(H \cdot A)(\omega, t) = \int_{(0,t]} H(\omega, s) dA(\omega, s) . \quad (2.18)$$

2.3 Previsible processes

Definition 2.3 The *previsible σ -algebra* \mathcal{P} on $\Omega \times (0, \infty)$ is the σ -algebra generated by sets of the form $E \times (s, t]$ where $E \in \mathcal{F}_s$ and $s < t$. A *previsible process* H is a \mathcal{P} -measurable map $H : \Omega \times (0, \infty) \rightarrow \mathbb{R}$.

Proposition 2.4 Let X be càdlàg adapted and $H_t = X_{t-}$, $t > 0$. Then H is previsible.

Proof. $H : \Omega \times (0, \infty) \rightarrow \mathbb{R}$ is left-continuous and adapted.

Set $t_n^- = k2^{-n}$ when $k2^{-n} < t \leq (k+1)2^{-n}$ and

$$H_t^n = H_{t_n^-} = \sum_{k=0}^{\infty} H_{k2^{-n}} \mathbf{1}_{\{(k2^{-n}, (k+1)2^{-n}]\}}(t) . \quad (2.19)$$

So H^n is previsible for all $n \in \mathbb{N}$ since $H_{t_n^-}$ is $\mathcal{F}_{t_n^-}$ -measurable as H is adapted and $t_n^- < t$. But $t_n^- \nearrow t$ and so $H_t^n \rightarrow H_t$ as $n \rightarrow \infty$ by left-continuity and H is also previsible. \square

Proposition 2.5 Let H be a previsible process. Then H_t is \mathcal{F}_{t-} -measurable for all $t > 0$, where $\mathcal{F}_{t-} = \sigma(\mathcal{F}_s : s < t)$.

Proof. Optional problem 4 in Example sheet 1 asks for a proof.

Remark. \mathcal{P} is the smallest σ -algebra such that all adapted left-continuous processes are measurable.

Examples. (i) Brownian motion is previsible by Proposition 2.4, since it is continuous.

- (ii) A Poisson process $(N_t)_{t \geq 0}$ or, indeed, any other continuous-time Markov chain with discrete state space is not previsible, since N_t is not \mathcal{F}_{t-} -measurable.

Proposition 2.6 *Let A be a càdlàg adapted finite variation process with total variation V . Let H be previsible such that for all $t \geq 0$ and all $\omega \in \Omega$*

$$\int_{(0,t]} |H(\omega, s)| dV(\omega, s) < \infty . \quad (2.20)$$

Then the process defined pathwise by

$$(H \cdot A)_t = \int_{(0,t]} H_s dA_s \quad (2.21)$$

is well-defined, càdlàg, adapted and of finite variation.

Proof. First note that the integral in (2.21) is well-defined for all t due to the finiteness of the integral in (2.20). (More precisely, (2.20) implies that all four terms defining (2.21) in (2.7) are finite).

We now show that $(H \cdot A)$ is càdlàg for each fixed $\omega \in \Omega$. We have $\mathbf{1}_{\{(0,s]\}} \rightarrow \mathbf{1}_{\{(0,t]\}}$ as $s \searrow t$, $\mathbf{1}_{\{(0,s]\}} \rightarrow \mathbf{1}_{\{(0,t]\}}$ as $s \nearrow t$ and

$$(H \cdot A)_t = \int_{(0,\infty)} H_s \mathbf{1}_{\{(0,t]\}}(s) dA_s . \quad (2.22)$$

Hence, by dominated convergence, the following limits exist

$$(H \cdot A)_{t+} = (H \cdot A)_t \quad \text{and} \quad (H \cdot A)_{t-} = \int_{(0,\infty)} H_s \mathbf{1}_{\{(0,t)\}}(s) dA_s , \quad (2.23)$$

and $H \cdot A$ is càdlàg with $\Delta(H \cdot A)_t = \int_{(0,\infty)} H_s \mathbf{1}_{\{t\}}(s) dA_s = H_t \Delta A_t$.

Next, we show that $H \cdot A$ is adapted via a monotone class argument. Suppose first $H = \mathbf{1}_{\{B \times (s,u]\}}$ where $B \in \mathcal{F}_s$. Then $(H \cdot A)_t = \mathbf{1}_{\{B\}}(A_{t \wedge u} - A_{s \wedge t})$ which is clearly \mathcal{F}_t -measurable. Now let

$$\Pi = \{B \times (s, u] : B \in \mathcal{F}_s, s < u\} \quad \text{and} \quad (2.24)$$

$$\mathcal{A} = \{C \in \mathcal{P} : (\mathbf{1}_{\{C\}} \cdot A)_t \text{ is } \mathcal{F}_t\text{-measurable}\} \quad (2.25)$$

so that Π is a π -system and $\Pi \subseteq \mathcal{A}$. But $\mathcal{A} \subseteq \mathcal{P} = \sigma(\Pi)$ and \mathcal{A} is a λ -system. [$\Omega \times (0, \infty) \in \mathcal{A}$; if $C \subseteq D \in \mathcal{A}$ then $((\mathbf{1}_{\{D\}} - \mathbf{1}_{\{C\}}) \cdot A)_t$ is \mathcal{F}_t -measurable, which gives $D \setminus C \in \mathcal{A}$; if $C_n \in \mathcal{A}$ with $C_n \nearrow C$ then $C \in \mathcal{A}$ since a limit of measurable functions is measurable.]

Hence, by Dynkin's lemma, $\sigma(\Pi) \subseteq \mathcal{A}$. But by definition, $\sigma(\Pi) = \mathcal{P}$ and $\mathcal{A} \subseteq \mathcal{P}$. Thus $\mathcal{A} = \mathcal{P}$. Suppose now that H is non-negative, \mathcal{P} -measurable. For all $n \in \mathbb{N}$ set

$$H^n := 2^{-n} \lfloor 2^n H \rfloor = \sum_{k=1}^{\infty} 2^{-n} k \underbrace{\mathbf{1}_{\{H \in [2^{-n}k, 2^{-n}(k+1))\}}}_{\in \mathcal{P}} , \quad (2.26)$$

so that $(H^n \cdot A)_t$ is \mathcal{F}_t -measurable. We have $(H^n \cdot A)_t \nearrow (H \cdot A)_t$ by monotone convergence (applied for each ω). Hence, $(H \cdot A)_t$ is \mathcal{F}_t -measurable. This extends in the usual way to \mathcal{P} -measurable $H = H^+ - H^-$ such that $|H| \cdot V = (H^+ \cdot V) + (H^- \cdot V) < \infty$.

We show finite variation for each fixed $\omega \in \Omega$. If $H^\pm = \max\{\pm H, 0\}$, $A^\pm = \frac{1}{2}(V \pm A)$ then analogous to (2.7)

$$H \cdot A = (H^+ \cdot A^+ + H^- \cdot A^-) - (H^+ \cdot A^- + H^- \cdot A^+) . \quad (2.27)$$

This is the difference of two increasing functions and thus $H \cdot A$ is of finite variation. \square

Example. Suppose that H is a previsible process, such as Brownian motion, and that

$$\int_{(0,t]} |H_s| ds < \infty \quad \text{for all } \omega \in \Omega \text{ and } t \geq 0 . \quad (2.28)$$

Then $\int_{(0,t]} H_s ds$ is càdlàg, adapted and of finite variation.

2.4 Local martingales

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ where $(\mathcal{F}_t)_{t \geq 0}$ satisfies what is technically known as *the usual conditions*, i.e. \mathcal{F} is \mathbb{P} -complete (equivalently, \mathcal{F}_0 contains all \mathbb{P} -null sets), and $(\mathcal{F}_t)_{t \geq 0}$ is right-continuous in the sense that

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s \quad \text{for all } t \geq 0 . \quad (2.29)$$

Note for instance that the filtration generated by Brownian motion completed by zero-probability events satisfies the usual conditions (this is essentially a consequence of the simple Markov property and Blumenthal's zero-one law).

Recall that an adapted process X is a *martingale* if it is integrable ($\mathbb{E}(|X_t|) < \infty$ for all $t \geq 0$) and if

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s \quad \text{a.s. for all } s \leq t . \quad (2.30)$$

We write \mathcal{M} for the set of all càdlàg martingales. The following result is fundamental and will be used repeatedly in this course.

Theorem 2.7 Optional stopping theorem (OST) *Let X be a cadlag adapted integrable process. Then the following are equivalent:*

- (i) X is a martingale
- (ii) $X^T = (X_{t \wedge T}, t \geq 0)$ is a martingale for all bounded stopping times T .
- (iii) For all bounded stopping times S, T , $\mathbb{E}(X_T | \mathcal{F}_S) = X_{S \wedge T}$ a.s. .
- (iv) $\mathbb{E}(X_T) = \mathbb{E}(X_0)$ for all bounded stopping times T .

Proof. It is well known that (i) $\Rightarrow \dots \Rightarrow$ (iv). We show how (iv) implies (i). Let $s < t$ and fix $u > t$. Let $A \in \mathcal{F}_s$, and define a random time T by saying $T = t$ if A occurs, or $T = u$ otherwise. Similarly, define $S = s$ if A occurs and $S = u$ otherwise. Note that both S and T are stopping times, and are bounded. Thus by (iv):

$$\mathbb{E}(X_T) = \mathbb{E}(X_0) = \mathbb{E}(X_S). \quad (2.31)$$

On the other hand,

$$\mathbb{E}(X_T) = \mathbb{E}(X_t \mathbf{1}_{\{A\}}) + \mathbb{E}(X_u \mathbf{1}_{\{A^c\}})$$

and similarly:

$$\mathbb{E}(X_S) = \mathbb{E}(X_s \mathbf{1}_{\{A\}}) + \mathbb{E}(X_u \mathbf{1}_{\{A^c\}}).$$

Plugging this into (2.31) and cancelling the terms $\mathbb{E}(X_u \mathbf{1}_{\{A^c\}})$, we find:

$$\mathbb{E}(X_t \mathbf{1}_{\{A\}}) = \mathbb{E}(X_s \mathbf{1}_{\{A\}})$$

for all $s < t$ and all $A \in \mathcal{F}_s$. This means (by definition) that

$$\mathbb{E}(X_t | \mathcal{F}_s) = X_s, \quad a.s.$$

as required. Hence, since X is adapted and integrable, X is a martingale.

It is also the case that \mathcal{M} is *stable under stopping*. This observation leads us to define a slightly more general class of processes, called local martingales.

Definition 2.4 A càdlàg adapted process X is a *local martingale*, $X \in \mathcal{M}_{loc}$, if there exists a sequence $(T_n)_{n \in \mathbb{N}}$ of stopping times with $T_n \nearrow \infty$ such that $(X_t^{T_n})_{t \geq 0} \in \mathcal{M}$ for all $n \in \mathbb{N}$. We say that the sequence $(T_n)_{n \in \mathbb{N}}$ *reduces* X .

In particular $\mathcal{M} \subseteq \mathcal{M}_{loc}$ since any sequence $(T_n)_{n \in \mathbb{N}}$ of stopping times reduces X by OST(ii).

Recall that a family $\mathcal{X} = (X_i)_{i \in I}$ of random variables is called *uniformly integrable* (UI) if

$$\sup_{i \in I} \mathbb{E}(|X_i| \mathbf{1}_{\{|X_i| \geq \lambda\}}) \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty. \quad (2.32)$$

Lemma 2.8 *If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ then the set*

$$\mathcal{X} = \{\mathbb{E}(X | \mathcal{G}) : \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\} \quad \text{is UI.} \quad (2.33)$$

Proof. See Advanced Probability course.

We now give necessary and sufficient conditions for a local martingale to be a martingale.

Proposition 2.9 *The following statements are equivalent:*

- (i) X is a martingale

(ii) X is a local martingale and for all $t \geq 0$ the set

$$\mathcal{X}_t = \{X_T : T \text{ is a stopping time, } T \leq t\} \text{ is UI.} \quad (2.34)$$

Proof. Suppose (i) holds. By the Optional Stopping Theorem, if T is a stopping time with $T \leq t$, then $X_T = \mathbb{E}(X_t | \mathcal{F}_T)$ a.s.. Thus by Lemma 2.8 \mathcal{X}_t is uniformly integrable. If (ii) holds, suppose $(T_n)_{n \geq 0}$ reduces X . Let T be any bounded stopping time, $T \leq t$, say. By the Optional Stopping Theorem applied to the martingale X^{T_n} ,

$$\mathbb{E}(X_0) = \mathbb{E}(X_0^{T_n}) = \mathbb{E}(X_T^{T_n}) = \mathbb{E}(X_{T \wedge T_n}). \quad (2.35)$$

Since $\{X_{T \wedge T_n} : n \in \mathbb{N}\}$ is uniformly integrable by assumption, $\mathbb{E}(X_{T \wedge T_n}) \rightarrow \mathbb{E}(X_T)$ as $n \rightarrow \infty$. Therefore,

$$\mathbb{E}(X_T) = \mathbb{E}(X_0).$$

But then by the Optional Stopping Theorem again, X must be a martingale. \square

An extremely useful consequence of the above is the following:

Corollary 2.10 *Let M be a local martingale, and assume that M is bounded. Then M is a true martingale. More generally, if M is a local martingale such that for all $t \geq 0$, $|M_t| \leq Z$ for some $Z \in L^1$, then M is a true martingale.*

Remark. Occasionally, we will need the following stronger version of (iii) in the Optional stopping theorem: if X is a uniformly integrable martingale, then for any stopping times S, T

$$\mathbb{E}(X_T | \mathcal{F}_S) = X_{S \wedge T} \quad (2.36)$$

almost surely.

Proposition 2.11 *A nonnegative local martingale M is a supermartingale.*

Proof. This follows simply from the definition of local martingales and Fatou's lemma for conditional expectations. \square

Remark. A martingale can be interpreted as the fortune of a player in a fair game. A local martingale which is not a true martingale, on the other hand, is the fortune of a player in a game which looks locally fair: unfortunately, this is only because there are going to be times of huge increases of X followed by an eventual ruin. Overall, as the above proposition shows, the expected fortune decreases. A local martingale is thus something akin to a bubble in the market. (Thanks are due to M. Tehranchi for this analogy).

Proposition 2.12 *Let M be a continuous local martingale ($M \in \mathcal{M}_{c,loc}$) starting from 0. Set $S_n = \inf \{t \geq 0 : |M_t| = n\}$. Then $(S_n)_{n \geq 0}$ reduces M .*

Proof. Note that $\{S_n \leq t\} = \bigcap_{k \in \mathbb{N}} \bigcup_{\substack{s \in \mathbb{Q} \\ s \leq t}} \{|M_s| > n - 1/k\} \in \mathcal{F}_t$,

and so S_n is a stopping time. For each $\omega \in \Omega$, $(S_n(\omega), n \geq 0)$ must be nondecreasing by the mean-value theorem since M is continuous, and $\lim_{n \rightarrow \infty} S_n$ can only be infinite by continuity as well. Hence $S_n \nearrow \infty$ a.s.. Let $(T_k)_{k \in \mathbb{N}}$ be a reducing sequence for M , i.e. $M^{T_k} \in \mathcal{M}$. By OST, also $M^{S_n \wedge T_k} \in \mathcal{M}$ and so $M^{S_n} \in \mathcal{M}_{loc}$ for each $n \in \mathbb{N}$. But M^{S_n} is bounded and so also a martingale. \square

Theorem 2.13 *Let M be a continuous local martingale which is also of finite variation, and such that $M_0 = 0$ a.s. Then M is indistinguishable from 0.*

Remarks.

- (i) In particular Brownian motion is not of finite variation.
- (ii) This makes it clear that the theory of finite variation integrals we have developed is useless for integrating with respect to continuous local martingales.
- (iii) It is essential to assume that M is continuous in this theorem.

Proof. Let V denote the total variation process of M . Then V is continuous and adapted with $V_0 = 0$. Set $S_n = \inf\{t \geq 0 : V_t = n\}$. Then S_n is a stopping time for all $n \in \mathbb{N}$ since V is adapted, and $S_n \nearrow \infty$ as $n \rightarrow \infty$ since V_t is nondecreasing and finite for all $t \geq 0$.

It suffices to show $M^{S_n} \equiv 0$ for all $n \in \mathbb{N}$. By OST, $M^{S_n} \in \mathcal{M}_{loc}$. Also

$$|M_t^{S_n}| \leq |V_t^{S_n}| \leq n, \quad (2.37)$$

and so, by Proposition 2.9, $M^{S_n} \in \mathcal{M}$.

Replacing M by M^{S_n} we can reduce to the case where M is a bounded martingale of bounded variation, i.e. V is bounded.

Lemma 2.14 *Let M be a martingale and such that for some given $s < t$, $\mathbb{E}(M_s^2) < \infty$ and $\mathbb{E}(M_t^2) < \infty$. Then*

$$\mathbb{E}(M_t^2 - M_s^2 | \mathcal{F}_s) = \mathbb{E}((M_t - M_s)^2 | \mathcal{F}_s), \quad a.s. \quad (2.38)$$

(This trick will be used over and over again in what follows, so it is a good point to memorize it).

Proof. By expanding the square $(M_t - M_s)^2$, the right-hand side is equal to

$$\begin{aligned} \mathbb{E}((M_t - M_s)^2 | \mathcal{F}_s) &= \mathbb{E}(M_t^2 | \mathcal{F}_s) - 2M_s \mathbb{E}(M_t | \mathcal{F}_s) + M_s^2 \\ &= \mathbb{E}(M_t^2 | \mathcal{F}_s) - 2M_s^2 + M_s^2 \\ &= \mathbb{E}(M_t^2 - M_s^2 | \mathcal{F}_s) \end{aligned}$$

as required.

Coming back to the proof of the theorem, fix $t > 0$ and set $t_k = kt/N$ for $0 \leq k \leq N$. By (2.38),

$$\begin{aligned} \mathbb{E}(M_t^2) &= \mathbb{E}\left(\sum_{k=0}^{N-1} (M_{t_{k+1}}^2 - M_{t_k}^2)\right) = \mathbb{E}\left(\sum_{k=0}^{N-1} (M_{t_{k+1}} - M_{t_k})^2\right) \\ &\leq \mathbb{E}\left(\underbrace{\sup_{k < N} |M_{t_{k+1}} - M_{t_k}|}_{\leq V_t \leq n} \underbrace{\sum_{k=0}^{N-1} |M_{t_{k+1}} - M_{t_k}|}_{\leq V_t \leq n}\right). \end{aligned} \quad (2.39)$$

As M is bounded and continuous

$$\sup_{k < N} |M_{t_{k+1}} - M_{t_k}| \rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad (2.40)$$

and so, by bounded convergence,

$$\mathbb{E} \left(\sup_{k < N} |M_{t_{k+1}} - M_{t_k}| \sum_{k=0}^{N-1} |M_{t_{k+1}} - M_{t_k}| \right) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (2.41)$$

Hence, $\mathbb{E}(M_t^2) = 0$ for all $t \geq 0$. Since M is continuous, M is indistinguishable from 0. \square

Definition 2.5 A continuous *semimartingale* X is an adapted continuous process which may be written as

$$X = X_0 + M + A \quad \text{with} \quad M_0 = A_0 = 0, \quad (2.42)$$

where $M \in \mathcal{M}_{c,loc}$ and A is a finite variation continuous process.

Note that as a consequence of Theorem 2.13, the decomposition is unique up to indistinguishability. This is known as the *Doob-Meyer decomposition*.

Remark. The proof of the last theorem tells us something extremely useful for the following. If t_k is the dyadic subdivision, then the calculation shows that

$$\mathbb{E}(M_t^2) = \mathbb{E} \left(\sum_{k:t_k \leq t} (M_{t_{k+1}} - M_{t_k})^2 \right)$$

so there is good reason to believe that if M is say, bounded in L^2 , then it has finite quadratic variation Q_t and moreover

$$M_t^2 - Q_t$$

has constant expectation 0. In fact, we will see that this is indeed the case and $M_t^2 - Q_t$ is also a martingale.

3 The stochastic integral

In this section we establish the stochastic integral with respect to continuous semimartingales. In places, we develop parts of the theory also for càdlàg semimartingales, where this involves no extra work. However, parts of the construction will use crucially the assumption of continuity. A more general theory exists, but it is beyond the scope of this course.

Recall that we say a process X is *bounded in L^2* if

$$\sup_{t \geq 0} \|X_t\|_2 < \infty \quad (3.1)$$

where here and in the rest of the course, for a random variable X :

$$\|X\|_2 := \mathbb{E}(|X|^2)^{1/2}. \quad (3.2)$$

Write \mathcal{M}^2 for the set of all càdlàg L^2 -bounded martingales, and \mathcal{M}_c^2 for the set of continuous martingales bounded in L^2 . Recall the following two fundamental results from Advanced probability:

Theorem 3.1 *Let $X \in \mathcal{M}^2$. There exists $X_\infty \in L^2$ such that*

$$X_t \rightarrow X_\infty \quad \text{a.s. and in } L^2, \text{ as } t \rightarrow \infty. \quad (3.3)$$

Moreover, $X_t = \mathbb{E}(X_\infty | \mathcal{F}_t)$ a.s. for all $t \geq 0$.

The second result which we will need is **Doob's L^2 inequality**:

Theorem 3.2 *For $X \in \mathcal{M}^2$,*

$$\mathbb{E} \left(\sup_{t \geq 0} |X_t|^2 \right) \leq 4 \mathbb{E} (X_\infty^2). \quad (3.4)$$

Similar to the construction to the Lebesgue integral in measure theory, we start by constructing the stochastic integral when the the integrand is very simple.

3.1 Simple integrands and L^2 properties

Definition 3.1 A *simple process* is any map $H : \Omega \times (0, \infty) \rightarrow \mathbb{R}$ of the form

$$H(\omega, t) = \sum_{k=0}^{n-1} Z_k(\omega) \mathbf{1}_{\{(t_k, t_{k+1}]\}}(t) \quad (3.5)$$

where $n \in \mathbb{N}$, $0 = t_0 < \dots < t_n < \infty$ and Z_k is a *bounded* \mathcal{F}_{t_k} -measurable random variable for all k . We denote the set of simple processes by \mathcal{S} . Given $H \in \mathcal{S}$, we denote $\|H\|_\infty = \text{esssup}|H|$ the essential supremum of H , i.e., the smallest $M > 0$ such that $\sup_{t \geq 0} |H(t, \omega)| \leq M$ almost surely.

Note that \mathcal{S} is a vector space and that (by definition) every simple process is previsible. We now define the stochastic integral for simple processes.

Definition 3.2 For $H = \sum_{k=0}^{n-1} Z_k \mathbf{1}_{\{(t_k, t_{k+1}]\}} \in \mathcal{S}$ and $M \in \mathcal{M}^2$ set

$$(H \cdot M)_t = \sum_{k=0}^{n-1} Z_k (M_{t_{k+1} \wedge t} - M_{t_k \wedge t}) . \quad (3.6)$$

Proposition 3.3 Let $H \in \mathcal{S}$ and $M \in \mathcal{M}^2$. Let T be a stopping time. Then

- (i) $H \cdot M^T = (H \cdot M)^T$.
- (ii) $H \cdot M \in \mathcal{M}^2$.
- (iii) $\mathbb{E}((H \cdot M)_\infty^2) \leq \|H\|_\infty^2 \mathbb{E}((M_\infty - M_0)^2)$.

Proof. (i) For all $t \geq 0$ we have

$$\begin{aligned} (H \cdot M^T)_t &= \sum_{k=0}^{n-1} Z_k (M_{t_{k+1} \wedge t}^T - M_{t_k \wedge t}^T) = \\ &= \sum_{k=0}^{n-1} Z_k (M_{t_{k+1} \wedge t \wedge T} - M_{t_k \wedge t \wedge T}) = (H \cdot M)_{t \wedge T} = (H \cdot M)_t^T . \end{aligned} \quad (3.7)$$

(ii) For $t_k \leq s \leq t < t_{k+1}$, $(H \cdot M)_t - (H \cdot M)_s = Z_k(M_t - M_s)$, so that

$$\mathbb{E}((H \cdot M)_t - (H \cdot M)_s \mid \mathcal{F}_s) = Z_k \mathbb{E}(M_t - M_s \mid \mathcal{F}_s) = 0 . \quad (3.8)$$

This extends easily to general $s \leq t$ and hence $H \cdot M$ is a martingale. To show it is bounded in L^2 , note that if $j < k$ we have the following "orthogonality relation":

$$\begin{aligned} &\mathbb{E}\left(Z_j(M_{t_{j+1}} - M_{t_j}) Z_k(M_{t_{k+1}} - M_{t_k})\right) = \\ &= \mathbb{E}\left(Z_j(M_{t_{j+1}} - M_{t_j}) Z_k \mathbb{E}(M_{t_{k+1}} - M_{t_k} \mid \mathcal{F}_{t_k})\right) = 0 . \end{aligned} \quad (3.9)$$

Thus let $t \geq t_n$ for the moment. To compute $\mathbb{E}((H \cdot M)_t^2)$, we expand the square and use the above orthogonality relation:

$$\begin{aligned} \mathbb{E}((H \cdot M)_t^2) &= \mathbb{E}\left(\left(\sum_{k=0}^{n-1} Z_k(M_{t_{k+1}} - M_{t_k})\right)^2\right) = \sum_{k=0}^{n-1} \mathbb{E}(Z_k^2(M_{t_{k+1}} - M_{t_k})^2) \\ &\leq \|H\|_\infty^2 \sum_{k=0}^{n-1} \mathbb{E}((M_{t_{k+1}} - M_{t_k})^2) = \|H\|_\infty^2 \mathbb{E}((M_t - M_0)^2) . \end{aligned} \quad (3.10)$$

(On two occasions, we used the trick (2.38)). By Doob's L^2 inequality applied to $M - M_0$,

$$\sup_{t \geq 0} \mathbb{E}((H \cdot M)_t^2) \leq 4\|H\|_\infty^2 \mathbb{E}((M_\infty - M_0)^2) . \quad (3.11)$$

Similarly, if $t_j \leq t < t_{j+1}$, then

$$\mathbb{E}((H \cdot M)_t^2) \leq \|H\|_\infty^2 \mathbb{E}((M_t - M_0)^2).$$

Doob's inequality applies equally well in this case and we conclude, for all $t \geq 0$,

$$\sup_{t \geq 0} \mathbb{E}((H \cdot M)_t^2) \leq 4 \|H\|_\infty^2 \mathbb{E}((M_\infty - M_0)^2).$$

Hence, $H \cdot M \in \mathcal{M}^2$.

(iii) Since $H \cdot M \in \mathcal{M}^2$, it converges to a limit in L^2 . We may thus let $t \rightarrow \infty$ in (3.10), which gives the result. \square

To extend the simple integral defined in the last section, we will need some Hilbert space properties of the set of integrators we are considering. As before, we work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ where $(\mathcal{F}_t)_{t \geq 0}$ satisfies the usual conditions.

Definition 3.3 For all càdlàg adapted processes X define the triple norm¹

$$\|X\| := \left\| \sup_{t \geq 0} |X_t| \right\|_2.$$

We write \mathcal{C}^2 for the set of all càdlàg adapted processes X such that $\|X\| < \infty$. On \mathcal{M}^2 , define the norm $\|X\| := \|X_\infty\|_2$.

Remark. Note that the function $\|\cdot\|$ on \mathcal{M}^2 defines indeed a norm. The only point which demands justification is the requirement that if $\|M\| = 0$, then M is indistinguishable from 0. But if $\|M\| = 0$, then $\mathbb{E}(M_\infty^2) = 0$ so $M_\infty = 0$ a.s. By the martingale convergence theorem

$$M_t = \mathbb{E}(M_\infty | \mathcal{F}_t) \text{ a.s.}$$

so $M_t = 0$ a.s. as well. Since M is càdlàg, it is indistinguishable from 0.

We may now state some L^2 properties which show that the space of square-integrable martingales can be seen as a Hilbert space. As we will see later, this underlying Hilbert structure is key to the construction of the stochastic integral (formally, it is defined as an isometry between Hilbert spaces).

Proposition 3.4 *We have*

- (i) $(\mathcal{C}^2, \|\cdot\|)$ is complete
- (ii) $\mathcal{M}^2 = \mathcal{M} \cap \mathcal{C}^2$
- (iii) $(\mathcal{M}^2, \|\cdot\|)$ is a Hilbert space with $\mathcal{M}_c^2 = \mathcal{M}_c \cap \mathcal{M}^2$ as a closed subspace
- (iv) $X \mapsto X_\infty : \mathcal{M}^2 \rightarrow L^2(\mathcal{F}_\infty)$ is an isometry

Proof. (i) Suppose $(X^n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $(\mathcal{C}^2, \|\cdot\|)$. Then we can find a subsequence $(n_k)_{k \in \mathbb{N}}$ such that $\sum_{k=1}^{\infty} \|X^{n_{k+1}} - X^{n_k}\| < \infty$. Then by the triangular inequality,

$$\left\| \sum_{k=1}^{\infty} \sup_{t \geq 0} |X_t^{n_{k+1}} - X_t^{n_k}| \right\|_2 \leq \sum_{k=1}^{\infty} \|X^{n_{k+1}} - X^{n_k}\| < \infty \quad (3.12)$$

¹Usual conventions about versions apply, dealing with equivalence classes analogous to L^p -spaces

and so for almost every $\omega \in \Omega$,

$$\sum_{k=1}^{\infty} \sup_{t \geq 0} |X_t^{n_{k+1}}(\omega) - X_t^{n_k}(\omega)| < \infty.$$

Since the space of càdlàg functions equipped with the $\|\cdot\|_{\infty}$ norm is complete, there exists a càdlàg process X such that $(X_t^{n_k}(\omega))_{k \in \mathbb{N}} \rightarrow X(\omega)$ as $k \rightarrow \infty$ uniformly in $t \geq 0$. Now

$$\|X^n - X\|^2 = \mathbb{E} \left(\sup_{t \geq 0} |X_t^n - X_t|^2 \right) \quad (3.13)$$

$$\begin{aligned} &\leq \liminf_{k \rightarrow \infty} \mathbb{E} \left(\sup_{t \geq 0} |X_t^n - X_t^{n_k}|^2 \right) \quad \text{by Fatou's lemma} \\ &= \liminf_{k \rightarrow \infty} \|X^n - X^{n_k}\|^2 \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned} \quad (3.14)$$

because $(X^n)_{n \in \mathbb{N}}$ is a Cauchy sequence. Hence $(\mathcal{C}^2, \|\cdot\|)$ is complete.

(ii) For $X \in \mathcal{C}^2 \cap \mathcal{M}$ we have

$$\sup_{t \geq 0} \|X_t\|_2 \leq \left\| \sup_{t \geq 0} |X_t| \right\|_2 = \|X\| < \infty \quad (3.15)$$

and so $X \in \mathcal{M}^2$. On the other hand, if $X \in \mathcal{M}^2$, by Doob's inequality,

$$\|X\| \leq 2\|X\| < \infty, \quad \text{and so } X \in \mathcal{C}^2 \cap \mathcal{M}. \quad (3.16)$$

(iii) $(X, Y) \mapsto \mathbb{E}(X_{\infty}Y_{\infty})$ defines an inner product on \mathcal{M}^2 whose associated norm is precisely the double norm $\|\cdot\|$. Moreover, for $X \in \mathcal{M}^2$, we have shown in (ii) that

$$\|X\| \leq \|X\| \leq 2\|X\|, \quad (3.17)$$

that is, $\|\cdot\|$ and $\|\cdot\|$ are equivalent on \mathcal{M}^2 . Thus \mathcal{M}^2 is complete for $\|\cdot\|$ if and only if it is complete for $\|\cdot\|$, and by (i) it is thus sufficient to show that \mathcal{M}^2 is closed in $(\mathcal{C}^2, \|\cdot\|)$. If $X^n \in \mathcal{M}^2$ and $\|X^n - X\| \rightarrow 0$ as $n \rightarrow \infty$ for some X , then X is certainly càdlàg adapted and L^2 -bounded. Furthermore, by Jensen's inequality for conditional expectations,

$$\begin{aligned} \|\mathbb{E}(X_t | \mathcal{F}_s) - X_s\|_2 &\leq \|\mathbb{E}(X_t - X_t^n | \mathcal{F}_s)\|_2 + \|X_s^n - X_s\|_2 \\ &\leq \|X_t - X_t^n\|_2 - \|X_s^n - X_s\|_2 \end{aligned} \quad (3.18)$$

$$\leq 2\|X^n - X\| \rightarrow 0 \quad (3.19)$$

as $n \rightarrow \infty$ and so $X \in \mathcal{M}^2$. By the same argument \mathcal{M}_c^2 is closed in $(\mathcal{M}^2, \|\cdot\|)$, where continuity of $t \mapsto X_t(\omega)$ follows by uniform convergence in t .

(iv) For $X, Y \in \mathcal{M}^2$, $\|X - Y\| = \|X_{\infty} - Y_{\infty}\|_2$ by definition. \square

3.2 Quadratic variation

Definition 3.4 For a sequence $(X^n)_{n \in \mathbb{N}}$ we say that $X^n \rightarrow X$ *uniformly on compacts in probability (u.c.p.)* if

$$\forall \epsilon > 0 \forall t \geq 0 : \mathbb{P} \left(\sup_{s \leq t} |X_s^n - X_s| > \epsilon \right) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (3.20)$$

Theorem 3.5 Quadratic variation

For each $M \in \mathcal{M}_{c,loc}$ there exists a unique (up to indistinguishability) continuous adapted nondecreasing process $[M]$ such that $M^2 - [M] \in \mathcal{M}_{c,loc}$ and such that $[M]_0 = 0$ a.s. Moreover, for

$$[M]_t^n := \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2 \quad (3.21)$$

we have $[M]_t^n \rightarrow [M]$ u.c.p. as $n \rightarrow \infty$. We call $[M]$ the quadratic variation process of M .

Example. Let B be a standard Brownian motion. Then we know that $(B_t^2 - t)_{t \geq 0} \in \mathcal{M}_c$. But then by Theorem 3.5, $t = [B]_t$.

Lemma. Let $M \in \mathcal{M}$ be bounded. Suppose that $l \in \mathbb{N}$ and $0 = t_0 < t_1 < \dots < t_l < \infty$. Then $\mathbb{E} \left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right)$ is bounded.

Proof of the Lemma. First note that

$$\begin{aligned} \mathbb{E} \left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right) &= \sum_{k=0}^{l-1} \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^4 \right) + \\ &+ 2 \sum_{k=0}^{l-1} \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 \sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2 \right). \end{aligned} \quad (3.22)$$

For each fixed k we have

$$\begin{aligned} \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 \sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2 \right) &= \\ &= \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 \mathbb{E} \left(\sum_{j=k+1}^{l-1} (M_{t_{j+1}} - M_{t_j})^2 \middle| \mathcal{F}_{t_{k+1}} \right) \right) = \\ &= \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 \mathbb{E} \left(\sum_{j=k+1}^{l-1} (M_{t_{j+1}}^2 - M_{t_j}^2) \middle| \mathcal{F}_{t_{k+1}} \right) \right) = \\ &= \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 \mathbb{E} (M_{t_l}^2 - M_{t_{k+1}}^2 | \mathcal{F}_{t_{k+1}}) \right) = \\ &= \mathbb{E} \left((M_{t_{k+1}} - M_{t_k})^2 (M_{t_l}^2 - M_{t_{k+1}}^2) \right). \end{aligned} \quad (3.23)$$

After inserting this in (3.22) we get the estimate

$$\begin{aligned} \mathbb{E} \left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right) &\leq \\ &\leq \mathbb{E} \left(\left(\sup_j |M_{t_{j+1}} - M_{t_j}|^2 + 2 \sup_j |M_{t_l} - M_{t_j}|^2 \right) \sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right). \end{aligned} \quad (3.24)$$

Now, M is uniformly bounded by C , say. So using the inequality $(x - y)^2 \leq 2(x^2 + y^2)$, we obtain

$$\begin{aligned} \mathbb{E} \left(\left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right)^2 \right) &\leq 12C^2 \mathbb{E} \left(\sum_{k=0}^{l-1} (M_{t_{k+1}} - M_{t_k})^2 \right) = \\ &= 12C^2 \mathbb{E}((M_{t_l} - M_{t_0})^2) \leq 48C^4. \end{aligned} \quad (3.25)$$

Proof of Theorem 3.5. Wlog we will consider the case $M_0 = 0$.

Uniqueness: If A and A' are two increasing processes satisfying the conditions for $[M]$ then

$$A_t - A'_t = (M_t^2 - A'_t) - (M_t^2 - A_t) \in \mathcal{M}_{c,loc} \quad (3.26)$$

is of finite variation and thus $A \equiv A'$ *a.s.* by Theorem 2.13.

Existence: First we assume that M is bounded, which implies $M \in \mathcal{M}_c^2$. Fix $T > 0$ deterministic. Let

$$H_t^n = M_{2^{-n}\lfloor 2^n t \rfloor}^T = \sum_{k=0}^{\lfloor 2^n T \rfloor - 1} M_{k2^{-n}} \mathbf{1}_{\{(k2^{-n}, (k+1)2^{-n}]\}}(t). \quad (3.27)$$

Then $H^n \in \mathcal{S}$ for all $n \in \mathbb{N}$. Hence X^n defined by

$$X_t^n = (H^n \cdot M)_t = \sum_{k=0}^{\lfloor 2^n T \rfloor - 1} M_{k2^{-n}} (M_{(k+1)2^{-n} \wedge t} - M_{k2^{-n} \wedge t}), \quad (3.28)$$

is in \mathcal{M}_c^2 by Proposition 3.3 and by continuity of M . Recall that $\|X^n\| = \|X_\infty^n\|_2 = \|X_T^n\|_2$ since X_t^n is constant for $t \geq T$. For $n \geq m$ we have

$$\begin{aligned} \|X^n - X^m\|^2 &= \mathbb{E}(\|(H^n - H^m) \cdot M\|_T^2) \\ &\leq \mathbb{E} \left(\sup_{0 \leq t \leq T} |H_t^n - H_t^m|^2 \sum_{k=0}^{\lfloor 2^n T \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2 \right) \\ &\leq \mathbb{E} \left(\sup_{0 \leq t \leq T} |H_t^n - H_t^m|^4 \right)^{1/2} \mathbb{E} \left(\left(\sum_{k=0}^{\lfloor 2^n T \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})^2 \right)^2 \right)^{1/2} \end{aligned} \quad (3.29)$$

by Hölder's inequality. Since M is bounded, the second term is bounded by the lemma, and

$$\mathbb{E} \left(\sup_{0 \leq t \leq T} |H_t^n - H_t^m|^4 \right) = \mathbb{E} \left(\sup_{0 \leq t \leq T} |M_{2^{-n}\lfloor 2^n t \rfloor} - M_{2^{-m}\lfloor 2^m t \rfloor}|^4 \right) \rightarrow 0 \quad (3.30)$$

as $n, m \rightarrow \infty$ by uniform continuity of M on $[0, T]$ and bounded convergence. Hence X^n is a Cauchy sequence in $(\mathcal{M}_c^2, \|\cdot\|)$ and so, by Proposition 3.4, converges to a limit $Y = (Y_t, 0 \leq t \leq T) \in \mathcal{M}_c^2$. Now for any n and $1 \leq k \leq \lfloor 2^n T \rfloor$,

$$M_{k2^{-n}}^2 - 2X_{k2^{-n}}^n = \sum_{j=0}^{k-1} (M_{(j+1)2^{-n}} - M_{j2^{-n}})^2 = [M]_{k2^{-n}}^n. \quad (3.31)$$

Hence, $M_t^2 - 2X_t^n$ is increasing along the sequence of times $(k2^{-n}, 1 \leq k \leq \lfloor 2^n T \rfloor)$. Passing to the limit $n \rightarrow \infty$, $M_t^2 - 2X_t^n$ must be *a.s.* increasing. Set

$$[M]_t := M_t^2 - 2Y_t, \quad t \in [0, T]$$

on the set where $M^2 - 2Y$ is increasing and $[M] \equiv 0$ otherwise. Hence, $[M]$ is a continuous increasing process and $M^2 - [M] = 2Y$ is a martingale on $[0, T]$.

We extend the definition of $[M]_t$ to $t \in [0, \infty)$ by applying the foregoing for all $T \in \mathbb{N}$. Note that the process $[M]$ obtained with T is the restriction to $[0, T]$ of $[M]$ defined with $T + 1$. Now, note that $M_{2^{-n}\lfloor t2^n \rfloor}^2$ converges to M^2 u.c.p. by uniform continuity, and convergence of X^n towards X also holds in the u.c.p. sense since it holds in the stronger $(\mathcal{M}_c^2, \|\cdot\|)$ sense. Thus the theorem is proved when M is bounded.

Now we turn to the general case $M \in \mathcal{M}_{c,loc}$. Define

$$T_n := \inf \{t \geq 0 : |M_t| \geq n\}.$$

Then $(T_n)_{n \in \mathbb{N}}$ reduces M and we can apply the bounded case to M^{T_n} , writing $A^n = [M^{T_n}]$. By uniqueness, $A_{t \wedge T_n}^{n+1}$ and A_t^n are indistinguishable. Thus there exists an increasing process A such that for all $n \in \mathbb{N}$, $A_{t \wedge T_n}$ and A_t^n are indistinguishable. Define $[M]_t = A_t$. By construction, $(M_{t \wedge T_n}^2 - A_{t \wedge T_n})_{t \geq 0} \in \mathcal{M}_c$ and so $(M_t^2 - A_t)_{t \geq 0} \in \mathcal{M}_{c,loc}$, as required. It remains to show that $[M]$ is the u.c.p. limit of its dyadic approximations. Let $[M]^{(m)}$ be the dyadic approximation at stage m . Note that for fixed $n \geq 1$, we have $[M^{T_n}]^m \rightarrow [M^{T_n}]$ u.c.p. as $m \rightarrow \infty$ by the bounded case. Since for all fixed $t \geq 0$, $\mathbb{P}(T_n \geq 1) \rightarrow 1$ as $n \rightarrow \infty$, we obtain that $[M]^m \rightarrow [M]$ u.c.p. as $m \rightarrow \infty$. \square

Remark. Note that $[M]$ is nondecreasing and thus of finite variation, and that if T is any possibly random time,

$$[M^T] = [M]^T.$$

Theorem 3.6 *If $M \in \mathcal{M}_c^2$, $M^2 - [M]$ is a uniformly integrable martingale. Moreover, $\mathbb{E}([M]_\infty) = \mathbb{E}(M_\infty^2)$.*

Proof. Let $S_n = \inf \{t \geq 0 : |M_t| \geq n\}$. S_n is a stopping time and $[M]_{t \wedge S_n} \leq n$. Thus, the stopped local martingale satisfies

$$|M_{t \wedge S_n}^2 - [M]_{t \wedge S_n}| \leq n + \sup_{t \geq 0} M_t^2 \quad (3.32)$$

is bounded by an integrable random variable and thus a true martingale (see remark after Proposition 2.9). Thus

$$\mathbb{E}([M]_{t \wedge S_n}) = \mathbb{E}(M_{t \wedge S_n}^2) \quad \text{for all } t \geq 0. \quad (3.33)$$

We take the limit $t \rightarrow \infty$, using monotone convergence on the left and dominated convergence on the right, and then $n \rightarrow \infty$ by the same arguments to get

$$\mathbb{E}([M]_\infty) = \mathbb{E}(M_\infty^2) < \infty. \quad (3.34)$$

Hence, $|M_t^2 - [M]_t|$ is dominated by $\sup_{t \geq 0} M_t^2 + [M]_\infty$ which is integrable. Thus $M^2 - [M]$ is a true martingale and is uniformly integrable since:

$$\mathbb{E}\left(\sup_{t \geq 0} |M_t^2 - [M]_t|\right) \leq \mathbb{E}\left(\left(\sup_{t \geq 0} M_t\right)^2 + [M]_\infty\right) \leq 5\mathbb{E}(M_\infty^2) < \infty. \quad \square \quad (3.35)$$

Remark. Some textbooks use the notation $\langle M \rangle$ rather than $[M]$ for the quadratic variation. In general, $\langle M \rangle$ should be previsible and means something slightly different (beyond the scope of this course), but it coincides with $[M]$ when M is continuous.

3.3 Itô integrals

Proposition 3.7 *Let μ be a finite measure on the previsible σ -algebra \mathcal{P} . Then \mathcal{S} is a dense subspace of $L^2(\mathcal{P}, \mu)$.*

Proof. As the Z_k 's are bounded, it is certainly true that $\mathcal{S} \subseteq L^2(\mathcal{P}, \mu)$. Denote by $\bar{\mathcal{S}}$ the closure of \mathcal{S} in $L^2(\mathcal{P}, \mu)$. Set

$$\mathcal{A} = \{A \in \mathcal{P} : \mathbf{1}_{\{A\}} \in \bar{\mathcal{S}}\} . \quad (3.36)$$

Then \mathcal{A} is a λ -system. [Check: $\mathbf{1}_{\{\Omega \times (0, n)\}} \in \mathcal{S}$ so $\mathbf{1}_{\{\Omega \times (0, \infty)\}} \in \bar{\mathcal{S}}$ and $\Omega \times (0, \infty) \in \mathcal{A}$; if $C \subseteq D \in \mathcal{A}$ then $D \setminus C \in \mathcal{A}$; if $C_n \in \mathcal{A}$ and $C_n \nearrow C$ then $C \in \mathcal{A}$ since $\bar{\mathcal{S}}$ is the closure of \mathcal{S} in $L^2(\mathcal{P}, \mu)$]. \mathcal{A} contains the π -system $\{B \times (s, t] : B \in \mathcal{F}_s, s \leq t\}$, which generates \mathcal{P} . Hence, by Dynkin's lemma, $\mathcal{A} = \mathcal{P}$. The result now follows since linear combinations of measurable indicator functions are dense in L^2 . \square

Given $M \in \mathcal{M}_c^2$, define a measure μ on \mathcal{P} by

$$\mu(A \times (s, t]) = \mathbb{E}(\mathbf{1}_{\{A\}}([M]_t - [M]_s)) \quad \text{for all } s < t, A \in \mathcal{F}_s . \quad (3.37)$$

Since \mathcal{P} is generated by the π -system of events of this form, this uniquely specifies μ . Alternatively, write

$$\mu(d\omega \otimes dt) = \lambda(\omega, dt) \mathbb{P}(d\omega) , \quad (3.38)$$

where $\lambda(\omega, \cdot)$ is the Lebesgue-Stieltjes measure associated to $[M](\omega)$. Thus, for a previsible process $H \geq 0$,

$$\int_{\Omega \times (0, \infty)} H d\mu = \mathbb{E} \left(\int_0^\infty H_s d[M]_s \right) . \quad (3.39)$$

Definition 3.5 Set $L^2(M) = L^2(\Omega \times (0, \infty), \mathcal{P}, \mu)$ and write

$$\|H\|_M^2 = \|H\|_{L^2(M)}^2 = \mathbb{E} \left(\int_0^\infty H_s^2 d[M]_s \right) , \quad (3.40)$$

so that $L^2(M)$ is the space of previsible processes H such that $\|H\|_M^2 < \infty$.

Note that the simple processes $\mathcal{S} \subseteq L^2(M)$ for all $M \in \mathcal{M}_c^2$.

Theorem 3.8 Itô isometry

For every $M \in \mathcal{M}_c^2$ there exists a unique isometry $I : L^2(M) \rightarrow \mathcal{M}_c^2$ such that $I(H) = H \cdot M$ for all $H \in \mathcal{S}$.

Proof. Let $H = \sum_{k=0}^{n-1} Z_k \mathbf{1}_{\{(t_k, t_{k+1}]\}} \in \mathcal{S}$. By Proposition 3.3, $H \cdot M \in \mathcal{M}_c^2$ with

$$\|H \cdot M\|^2 = \sum_{k=0}^{n-1} \mathbb{E}(Z_k^2 (M_{t_{k+1}} - M_{t_k})^2). \quad (3.41)$$

But $M^2 - [M]$ is a martingale so that

$$\begin{aligned} \mathbb{E}(Z_k^2 (M_{t_{k+1}} - M_{t_k})^2) &= \mathbb{E}\left(Z_k^2 \mathbb{E}((M_{t_{k+1}} - M_{t_k})^2 \mid \mathcal{F}_{t_k})\right) = \\ &= \mathbb{E}\left(Z_k^2 \mathbb{E}(M_{t_{k+1}}^2 - M_{t_k}^2 \mid \mathcal{F}_{t_k})\right) = \mathbb{E}(Z_k^2 ([M]_{t_{k+1}} - [M]_{t_k})), \end{aligned} \quad (3.42)$$

and so $\|H \cdot M\|^2 = \mathbb{E}\left(\int_0^\infty H_s^2 d[M]_s\right) = \|H\|_M^2$.

Now let $H \in L^2(M)$. We have thus defined a function I from \mathcal{S} to \mathcal{M}_c^2 , which is an isometry. However, \mathcal{S} is dense in $L^2(M) = L^2(\mathcal{P}, \mu)$ by Proposition 3.7. This implies that there is a unique way to extend I to $L^2(M)$ which makes I into an isometry. Indeed, let $H \in L^2(M)$. Then there exists H^n a sequence of simple processes such that $H^n \rightarrow H$ in $L^2(M)$. Then by linearity:

$$\|I(H^n) - I(H^m)\| = \|I(H^n - H^m)\| = \|H^n - H^m\|_M$$

so $I(H^n)$ is a Cauchy sequence in $(\mathcal{M}_c^2, \|\cdot\|)$, which is complete. Therefore, $I(H^n)$ converges to some limit which we may denote by $I(H)$. It is easy to check that $I(H)$ does not depend on the sequence H^n chosen to approximate H : if $H^n \rightarrow H$ and $K^n \rightarrow H$ in $L^2(M)$, then $\|I(H^n) - I(K^n)\| = \|H^n - K^n\|_M \rightarrow 0$ as $n \rightarrow \infty$, so the limits of $I(H^n)$ and $I(K^n)$ must be indistinguishable. $I(H)$ is then, indeed, an isometry on $L^2(M)$. For $H \in \mathcal{S}$ we have consistently $I(H) = H \cdot M$ by choosing $H^n \equiv H$. \square

Definition 3.6 We write

$$I(H)_t = (H \cdot M)_t = \int_0^t H_s dM_s$$

for all $H \in L^2(M)$. The process $H \cdot M$ is *Itô's stochastic integral* of H with respect to M .

Remark. By Theorem 3.8, this is consistent with our previous definition of $H \cdot M$ for $H \in \mathcal{S}$.

Proposition 3.9 Let $M \in \mathcal{M}_c^2$ and $H \in L^2(M)$. Let T be a stopping time. Then $H \mathbf{1}_{(0, T]} \in L^2(M)$ and $H \in L^2(M^T)$, and we have:

$$(H \cdot M)^T = (H \mathbf{1}_{\{(0, T]\}}) \cdot M = H \cdot (M^T). \quad (3.43)$$

Proof. Let $H \in L^2(M)$. It is trivial to check that $H \mathbf{1}_{(0, T]} \in L^2(M)$ (to see that it is previsible, note that $\mathbf{1}_{(0, T]}(t)$ is left-continuous and hence previsible). To see that $H \in L^2(M^T)$, note that $[M^T] = [M]^T$ by the discrete approximation in the definition of quadratic variation, and thus

$$\mathbb{E} \int_0^\infty H_s^2 d[M^T]_s = \mathbb{E} \int_0^T H_s^2 d[M]_s \leq \mathbb{E} \int_0^\infty H_s^2 d[M]_s < \infty.$$

Step 1. Take $M \in \mathcal{M}_c^2$ and suppose first that $H \in \mathcal{S}$. If T takes only finitely many values, $H \mathbf{1}_{\{(0,T]\}} \in \mathcal{S}$ and $(H \cdot M)^T = (H \mathbf{1}_{\{(0,T]\}}) \cdot M$ is easily checked. For general T , set $T_n = (2^{-n} \lceil 2^n T \rceil) \wedge n$ which is a stopping time that takes only finitely many values. Then $T_n \nearrow T$ as $n \rightarrow \infty$ and so

$$\|H \mathbf{1}_{\{(0,T_n]\}} - H \mathbf{1}_{\{(0,T]\}}\|_M^2 = \mathbb{E} \left(\int_0^\infty H_t^2 (\mathbf{1}_{\{(0,T_n]\}} - \mathbf{1}_{\{(0,T]\}})^2(t) d[M]_t \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.44)$$

by dominated convergence, and so $H \mathbf{1}_{\{(0,T_n]\}} \cdot M \rightarrow H \mathbf{1}_{\{(0,T]\}} \cdot M$ in \mathcal{M}_c^2 by Theorem 3.8. But $(H \cdot M)_t^{T_n} \rightarrow (H \cdot M)_t^T$ a.s. by continuity and hence, $(H \cdot M)^T = (H \mathbf{1}_{\{(0,T]\}}) \cdot M$ since $(H \cdot M)^{T_n} = (H \mathbf{1}_{\{(0,T_n]\}}) \cdot M$ for all $n \in \mathbb{N}$ by the first part. On the other hand we already know $(H \cdot M)^T = H \cdot (M^T)$ by Proposition 3.3.

Step 2. Now for $H \in L^2(M)$ choose $H^n \in \mathcal{S}$ such that $H^n \rightarrow H$ in $L^2(M)$. Then $H^n \cdot M \rightarrow H \cdot M$ in \mathcal{M}_c^2 , so $(H^n \cdot M)^T \rightarrow (H \cdot M)^T$ in \mathcal{M}_c^2 by Doob's inequality. Also,

$$\|H^n \mathbf{1}_{\{(0,T]\}} - H \mathbf{1}_{\{(0,T]\}}\|_M^2 = \mathbb{E} \left(\int_0^T (H^n - H)_s^2 d[M]_s \right) \leq \|H^n - H\|_M^2 \rightarrow 0 \quad (3.45)$$

as $n \rightarrow \infty$, so $(H^n \mathbf{1}_{\{(0,T]\}}) \cdot M \rightarrow (H \mathbf{1}_{\{(0,T]\}}) \cdot M$ in \mathcal{M}_c^2 by the isometry property of Theorem 3.8. Again, by equating the limits of both sequences we get $(H \cdot M)^T = (H \mathbf{1}_{\{(0,T]\}}) \cdot M$. Moreover,

$$\begin{aligned} \|H^n - H\|_{M^T}^2 &= \mathbb{E} \left(\int_0^\infty (H^n - H)_s^2 d[M^T]_s \right) = \\ &= \mathbb{E} \left(\int_0^T (H^n - H)_s^2 d[M]_s \right) \leq \|H^n - H\|_M^2 \rightarrow 0, \end{aligned} \quad (3.46)$$

so $H^n \cdot (M^T) \rightarrow H \cdot (M^T)$ in \mathcal{M}_c^2 . Hence, $(H \cdot M)^T = H \cdot (M^T)$. \square

Proposition 3.9 allows us to make a final extension of Itô's integral to locally bounded, previsible integrands.

Definition 3.7 Let H be previsible. Say that H is *locally bounded* if there exist stopping times $S_n \nearrow \infty$ a.s. such that $H \mathbf{1}_{(0,S_n]}$ is bounded for all $n \in \mathbb{N}$, i.e. there exists $C_n < \infty$ nonrandom such that $\sup_{t \geq 0} |H_t \mathbf{1}_{(0,S_n]}(t)| \leq C_n$ a.s.. Note that a left-continuous process is always previsible and locally bounded.

Definition 3.8 Let H be a previsible locally bounded process and let $M \in \mathcal{M}_{c,loc}$. Choose stopping times $S'_n = \inf \{t \geq 0 : |M_t| \geq n\} \nearrow \infty$ a.s., and note that $M^{S'_n} \in \mathcal{M}_c^2$ for all $n \in \mathbb{N}$. Set $T_n = S_n \wedge S'_n$ and define

$$(H \cdot M)_t := (H \mathbf{1}_{(0,T_n]} \cdot M^{T_n})_t \quad \text{for all } t \leq T_n. \quad (3.47)$$

Remarks.

- (i) The stochastic integral in the right-hand side of (3.47) is well-defined: indeed, every bounded previsible process is in $L^2(M)$ whenever $M \in \mathcal{M}_c^2$. Moreover, with $H \mathbf{1}_{(0,S_n]}$ also $H \mathbf{1}_{(0,T_n]}$ is bounded and $M^{T_n} = (M^{S'_n})^{T_n} \in \mathcal{M}_c^2$, so $H \mathbf{1}_{(0,T_n]} \cdot M^{T_n}$ makes sense (it falls within the category of processes covered by Theorem 3.8).

- (ii) Proposition 3.9 ensures that the right-hand side does not depend on n for all n large enough that $T_n \geq t$.
- (iii) Note also that the definition does not depend on the sequence of stopping times $(T_n)_{n \geq 0}$ used to reduce M and H , so long as H^{T_n} is bounded and $M^{T_n} \in \mathcal{M}_c^2$ for all $n \geq 0$. Such a sequence of stopping times is said to *localize* H and M .
- (iv) It is furthermore consistent with our previous definitions of stochastic integral when $M \in \mathcal{M}_c^2$ and $H \in L^2(M)$.

Proposition 3.10 *Let $M \in \mathcal{M}_{c,loc}$ and H, K be locally bounded previsible processes. Let T be a stopping time. Then*

- (i) $(H \cdot M)^T = (H \mathbf{1}_{(0,T]}) \cdot M = H \cdot (M^T)$,
- (ii) $H \cdot M$ is a continuous local martingale,
- (iii) The quadratic variation of $\int_0^\cdot H_s dM_s$ is $\int_0^\cdot H_s^2 d[M]_s$,
- (iv) $H \cdot (K \cdot M) = (H K) \cdot M$.

Remark. As we will see, martingales are completely characterized by their quadratic variation, so (iii) is a fundamental property. (In addition, as already discussed informally at the very beginning of the construction, this property is in some sense what motivates the entire theory of stochastic integration). (iv) Should be seen as a sort of stochastic version of the fundamental theorem of calculus, since it is telling us that:

$$d \left(\int_0^t K_s dM_s \right) = K_t dM_t.$$

Proof. (i) Let us start by checking the first of these equalities. By Proposition 3.9, we know that

$$(H \mathbf{1}_{(0,T_n]} \cdot M^{T_n})^T = H \mathbf{1}_{(0,T]} \mathbf{1}_{(0,T_n]} \cdot M^{T_n}$$

As $n \rightarrow \infty$, the left-hand side converges pointwise *a.s.* to $(H \cdot M)^T$ by definition, while the right-hand side also converges pointwise *a.s.* to $H \mathbf{1}_{(0,T]} \cdot M$ since the sequence (T_n) also “reduce” $H \mathbf{1}_{(0,T]}$ and M in the sense of Definition 3.8. The second equality follows the same argument.

(ii) By (i),

$$(H \cdot M)^{T_n} = (H \mathbf{1}_{(0,T_n]}) \cdot M^{T_n} \in \mathcal{M}_c^2 \tag{3.48}$$

which implies (ii).

(iii) Assume first that $M \in \mathcal{M}_c^2$ and that H is uniformly bounded in time. For any stopping time T , we have by the isometry property of Theorem 3.8:

$$\begin{aligned} \mathbb{E}((H \cdot M)_T^2) &= \mathbb{E}((H \mathbf{1}_{(0,T]} \cdot M)_\infty^2) = \\ &= \mathbb{E}((H^2 \mathbf{1}_{(0,T]} \cdot [M])_\infty) = \mathbb{E}((H^2 \cdot [M])_T) \end{aligned} \tag{3.49}$$

By the optional stopping theorem, we conclude that $(H \cdot M)^2 - H^2 \cdot [M]$ is a martingale. Moreover, since H is locally bounded and $[M]$ continuous one also shows that $H^2 \cdot [M]$ is continuous with probability 1. Therefore, by Theorem 3.5, we have $[H \cdot M] \equiv H^2 \cdot [M]$. In the general case, note that as a consequence of (i) and of the fact that $[M^T] = [M]^T$, we may write

$$\begin{aligned}
[H \cdot M] &= \lim_{n \rightarrow \infty} [H \cdot M]^{T_n} \\
&= \lim_{n \rightarrow \infty} [(H \cdot M)^{T_n}] \\
&= \lim_{n \rightarrow \infty} [H 1_{(0, T_n]} \cdot M^{T_n}] \\
&= \lim_{n \rightarrow \infty} H^2 1_{(0, T_n]} \cdot [M]^{T_n} \quad (\text{by the above}) \\
&= H^2 \cdot [M] \quad (\text{by continuity})
\end{aligned}$$

where the limits in these equalities are a.s. pointwise limits.

(iv) The case $H, K \in \mathcal{S}$ is tedious but elementary. For H, K uniformly bounded and $M \in \mathcal{M}_c^2$, there exist $H^n, K^n \in \mathcal{S}$, $n \in \mathbb{N}$ such that $H^n \rightarrow H$ and $K^n \rightarrow K$ in $L^2(M)$. Furthermore, we may also assume that $\|H^n\|_\infty$ and $\|K^n\|_\infty$ are uniformly bounded in n (indeed, truncating K^n at $\|K\|_\infty + 1$ can only improve the L^2 difference between K^n and K). We first prove an upper bound on $\|H\|_{L^2(K \cdot M)}$:

$$\begin{aligned}
\|H\|_{L^2(K \cdot M)}^2 &= \mathbb{E}((H^2 \cdot [K \cdot M])_\infty) \\
&\stackrel{\text{by (iii)}}{=} \mathbb{E}((H^2 \cdot (K^2 \cdot [M]))_\infty) \\
&\stackrel{*}{=} \mathbb{E}(((H K)^2 \cdot [M])_\infty) \\
&= \|H K\|_{L^2(M)}^2 \\
&\leq \min \{ \|H\|_\infty^2 \|K\|_{L^2(M)}^2, \|H\|_{L^2(M)}^2 \|K\|_\infty^2 \}, \quad (3.50)
\end{aligned}$$

where $*$ holds by Optional problem 5 on example sheet 1, since $[M]$ is nondecreasing and thus of finite variation. We have $H^n \cdot (K^n \cdot M) = (H^n K^n) \cdot M$ and using (3.50)

$$\begin{aligned}
\|H^n \cdot (K^n \cdot M) - H \cdot (K \cdot M)\| &\leq \|(H^n - H) \cdot (K^n \cdot M)\| + \|H \cdot ((K^n - K) \cdot M)\| \\
&= \|H^n - H\|_{L^2(K^n \cdot M)} + \|H\|_{L^2((K^n - K) \cdot M)} \\
&\leq \|H^n - H\|_{L^2(M)} \|K^n\|_\infty + \|H\|_\infty \|K^n - K\|_{L^2(M)} \\
&\rightarrow 0 \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

So $H^n \cdot (K^n \cdot M) \rightarrow H \cdot (K \cdot M)$ in \mathcal{M}_c^2 . Similarly, $(H^n K^n) \cdot M \rightarrow (H K) \cdot M$ in \mathcal{M}_c^2 , which implies the result. \square

Definition 3.9 Let X be a continuous semimartingale $X = X_0 + M + A$ with $M \in \mathcal{M}_{c,loc}$, A a finite variation process and $M_0 = A_0 = 0$. We set the quadratic variation of X to be that of its martingale part, $[X] := [M]$, independently of A .

This definition finds its justification in the fact that

$$\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (X_{(k+1)2^{-n}} - X_{k2^{-n}})^2 \rightarrow [X]_t \quad u.c.p. \quad (3.51)$$

as $n \rightarrow \infty$, as is not hard to show.

Definition 3.10 For a continuous semimartingale X and H locally bounded and pre-visible, we define the *stochastic integral*

$$H \cdot X = H \cdot M + H \cdot A, \quad \text{writing also} \quad (H \cdot X)_t = \int_0^t H_s dX_s, \quad (3.52)$$

where $H \cdot M$ is Itô's integral from Definition 3.8 and $H \cdot A$ is the finite variation integral defined in Proposition 2.6. We agree that

$$dZ_t = H_t dX_t \quad \text{means} \quad Z_t - Z_0 = \int_0^t H_s dX_s \quad (3.53)$$

Note that $H \cdot X$ is already given in Doob-Meyer decomposition and is thus obviously a continuous semimartingale. Under the additional assumption that H is left-continuous, one can show that the Riemann sum approximation to the integral converges.

Proposition 3.11 *Let X be a continuous semimartingale and H be a left-continuous adapted process. Then*

$$\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} H_{k2^{-n}} (X_{(k+1)2^{-n}} - X_{k2^{-n}}) \rightarrow \int_0^t H_s dX_s \quad \text{u.c.p. as } n \rightarrow \infty. \quad (3.54)$$

Proof. We can treat the finite variation part $X_0 + A$ and the local martingale part M separately. The first is proved in problem 6 on example sheet 2 (in fact, uniformly on compacts for all ω). So it suffices to show that

$$\sum_{k=0}^{\lfloor 2^n t \rfloor - 1} H_{k2^{-n}} (M_{(k+1)2^{-n}} - M_{k2^{-n}}) \rightarrow (H \cdot M)_t \quad \text{u.c.p. as } n \rightarrow \infty \quad (3.55)$$

when $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. By localization, we can reduce to the case where $M \in \mathcal{M}_c^2$ and H_t is bounded uniformly for $t > 0$. Let $H_t^n = H_{2^{-n} \lfloor 2^n t \rfloor}$. Then $H_t^n \rightarrow H_t$ as $n \rightarrow \infty$ by left continuity. Now,

$$(H^n \cdot M)_t = \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} H_{k2^{-n}} (M_{(k+1)2^{-n}} - M_{k2^{-n}}) + H_{2^{-n} \lfloor 2^n t \rfloor} (M_t - M_{2^{-n} \lfloor 2^n t \rfloor}) \quad (3.56)$$

where, since M is continuous (and therefore almost surely uniformly continuous on any compact interval), $M_t - M_{2^{-n} \lfloor 2^n t \rfloor} \rightarrow 0$ u.c.p. as $n \rightarrow \infty$. We can thus ignore the second term on the right. Now

$$\|H^n - H\|_M = \mathbb{E} \left(\int_0^\infty (H_t^n - H_t)^2 d[M]_t \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.57)$$

by bounded convergence and the fact that $H_t^n \rightarrow H_t$ for every t as $n \rightarrow \infty$. By the isometry property, $H^n \cdot M \rightarrow H \cdot M$ in \mathcal{M}_c^2 . Using Doob's inequality, it is easy to see that this implies u.c.p. convergence. \square

To step away from the theory for a moment and look at a concrete example, you should try your hands at proving the following result. This will be generalized in a moment in Theorem 3.16 so you can go look for some inspiration there if you are stuck.

Proposition 3.12 *Let $(M_t, t \geq 0)$ be a continuous local martingale. Then for all $t \geq 0$,*

$$M_t^2 = M_0^2 + 2 \int_0^t M_s dM_s + [M]_t.$$

In particular if $(B_t, t \geq 0)$ is a one-dimensional standard Brownian motion, then

$$B_t^2 = 2 \int_0^t B_s dB_s + t$$

is a semi-martingale.

3.4 Covariation

In practice we do not calculate integrals from first principles, but rather use tools of calculus such as integration by parts or the chain rule. In this section we derive these tools for stochastic integrals, which differ from ordinary calculus in certain correction terms. A useful tool for deriving these rules will be the covariation of two local martingales.

Definition 3.11 *Let $M, N \in \mathcal{M}_{c,loc}$, and set*

$$[M, N] = \frac{1}{4}([M + N] - [M - N]) \quad (\text{polarization identity}) \quad (3.58)$$

$[M, N]$ is called the covariation of M and N .

Theorem 3.13 *Let $M, N \in \mathcal{M}_{c,loc}$. Then we have:*

(i) $[M, N]$ is the unique (up to indistinguishability) continuous adapted process with finite variation such that $MN - [M, N]$ is a continuous local martingale started from 0.

(ii) For $n \geq 1$ and for all $t \geq 0$, let

$$[M, N]_t^n := \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (M_{(k+1)2^{-n}} - M_{k2^{-n}})(N_{(k+1)2^{-n}} - N_{k2^{-n}}). \quad (3.59)$$

Then $[M, N]_t^n \rightarrow [M, N]$ u.c.p. as $n \rightarrow \infty$,

(iii) for $M, N \in \mathcal{M}_c^2$, $MN - [M, N]$ is a UI martingale

(iv) $[M, N]$ is a symmetric bilinear form.

Proof. (i) Note that $MN = \frac{1}{4}((M + N)^2 - (M - N)^2)$ It is thus obvious that $MN - [M, N]$ is a continuous local martingale. Moreover, finite variation is an obvious consequence of the definition and uniqueness follows easily from Theorem 2.13.

(ii) and (iii) follow from polarizing the sum (3.59) just as in (3.58) and applying Theorems 3.5 and 3.6.

For (iv), the symmetry comes from the uniqueness in (i), while the bilinearity also follows from (i).

Remark. Of course, $[M, M] = [M]$.

Proposition 3.14 Kunita-Watanabe Identity

Let $M, N \in \mathcal{M}_{c,loc}$ and H be a locally bounded previsible process. Then

$$[H \cdot M, N] = H \cdot [M, N]. \quad (3.60)$$

Proof. We may assume by localization that $M, N \in \mathcal{M}_c^2$ and that H is uniformly bounded in time. Note that $H \cdot [M, N]$ is of finite variation, and thus by the uniqueness of Theorem 3.13 (i), it suffices to prove that

$$(H \cdot M)N - H \cdot [M, N] \in \mathcal{M}_{c,loc}.$$

By the optional stopping theorem, it suffices to prove that for all bounded stopping times T ,

$$\mathbb{E}((H \cdot M)_T N_T) = \mathbb{E}((H \cdot [M, N])_T) \quad (3.61)$$

and by considering the stopped processes H^T, M^T and N^T it suffices to prove that $\mathbb{E}((H \cdot M)_\infty N_\infty) = \mathbb{E}((H \cdot [M, N])_\infty)$. If H is of the form $Z \mathbf{1}_{\{(s,t)\}}$ with Z bounded \mathcal{F}_s measurable, then this identity becomes

$$\mathbb{E}\{Z(M_t - M_s)N_\infty\} = \mathbb{E}\{Z([M, N]_t - [M, N]_s)\}.$$

However, note that since $MN - [M, N]$ is a martingale, we have:

$$\begin{aligned} \mathbb{E}\{Z(M_t - M_s)N_\infty\} &= \mathbb{E}\{ZM_t\mathbb{E}(N_\infty|\mathcal{F}_t)\} - \mathbb{E}\{ZM_s\mathbb{E}(N_\infty|\mathcal{F}_s)\} \\ &= \mathbb{E}\{Z\mathbb{E}(M_t N_t - M_s N_s|\mathcal{F}_s)\} \\ &= \mathbb{E}\{Z\mathbb{E}([M, N]_t - [M, N]_s|\mathcal{F}_s)\} \\ &= \mathbb{E}\{Z([M, N]_t - [M, N]_s)\}, \end{aligned}$$

as required. (3.61) then extends by linearity to all $H \in \mathcal{S}$. If H is bounded, we may find a sequence $H^n \rightarrow H$ in $L^2(M)$ such that $H^n \in \mathcal{S}$ and is uniformly bounded. The Lebesgue convergence theorem then shows that (3.61) holds. This proves the result. \square

Remark. Note that a consequence of this identity is that $[H \cdot M, H \cdot N] = H^2 \cdot [M, N]$. As an exercise, try to derive directly the latter identity by polarization arguments.

Definition 3.12 Let X, Y be continuous semi-martingales. We define their covariation $[X, Y]$ to be the covariation of their respective martingale parts in the Doob-Meyer decomposition.

It is not hard to see that $\lim_{n \rightarrow \infty} [X, Y]^n = [X, Y]$ u.c.p where

$$[X, Y]_t^n := \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (X_{(k+1)2^{-n}} - X_{k2^{-n}})(Y_{(k+1)2^{-n}} - Y_{k2^{-n}}).$$

An important property of the covariation is that two independent semi-martingales have zero covariation. However, just as there exist many pairs of random variables with zero correlation which are not independent, the converse is false. A notable exception is the Lévy characterization of Brownian motion.

Proposition 3.15 *Let X, Y be independent continuous semi-martingales. Then $[X, Y] = 0$.*

Proof. The case where $X, Y \in \mathcal{M}_{c,loc}$ is treated in exercise 13 of Example Sheet 2. You should think about the case where X, Y are general continuous semi-martingales. \square

Remark. Note that the Kunita-Watanabe identity $[H \cdot X, Y] = H \cdot [X, Y] = [X, H \cdot Y]$ also holds for continuous semi-martingales.

3.5 Itô's formula

Theorem 3.16 Integration by parts. *Let X, Y be continuous semimartingales. Then*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t . \quad (3.62)$$

Proof. Since both sides are continuous in t , it suffices to consider $t = M 2^{-N}$ for $M, N \geq 1$. Note that

$$X_t Y_t - X_s Y_s = X_s(Y_t - Y_s) + Y_s(X_t - X_s) + (X_t - X_s)(Y_t - Y_s) \quad (3.63)$$

so for $n \geq N$

$$\begin{aligned} X_t Y_t - X_0 Y_0 &= \sum_{k=0}^{M 2^{n-N} - 1} \left(X_{k 2^{-n}} (Y_{(k+1) 2^{-n}} - Y_{k 2^{-n}}) + Y_{k 2^{-n}} (X_{(k+1) 2^{-n}} - X_{k 2^{-n}}) + \right. \\ &\quad \left. + (X_{(k+1) 2^{-n}} - X_{k 2^{-n}}) (Y_{(k+1) 2^{-n}} - Y_{k 2^{-n}}) \right) \\ &\xrightarrow{u.c.p.} (X \cdot Y)_t + (Y \cdot X)_t + [X, Y]_t \quad \text{as } n \rightarrow \infty \end{aligned} \quad (3.64)$$

by Proposition 3.11 and Theorem 3.59. \square

Note the extra covariation term which we do not get in the deterministic case. The next result, Itô's formula, tells us that a smooth function of a continuous semimartingale is again a continuous semimartingale and gives us its precise decomposition in a sort of chain rule.

Theorem 3.17 Itô's formula

Let X^1, X^2, \dots, X^d be continuous semimartingales and set $X = (X^1, \dots, X^d)$.

Let $f \in C^2(\mathbb{R}^d, \mathbb{R})$. Then

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d[X^i, X^j]_s . \quad (3.65)$$

Remarks. (i) In particular, $f(X)$ is a continuous semimartingale with decomposition

$$\begin{aligned}
f(X_t) &= f(X_0) + \underbrace{\sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s) dM_s^i}_{\in \mathcal{M}_{c,loc}} + \\
&\quad + \underbrace{\sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s) dA_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d[M^i, M^j]_s}_{\text{finite variation}}. \quad (3.66)
\end{aligned}$$

where the covariation of the \mathbb{R}^d -valued semimartingale $X = X_0 + A + M$ is $[X^i, X^j] = [M^i, M^j]$, due to quadratic variation and the polarization identity (3.58).

(ii) Intuitive proof by Taylor expansion for $d = 1$:

$$\begin{aligned}
f(X_t) &= f(X_0) + \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (f(X_{(k+1)2^{-n}}) - f(X_{k2^{-n}})) + (f(X_t) - f(X_{\lfloor 2^n t \rfloor 2^{-n}})) = \\
&= f(X_0) + \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} f'(X_{k2^{-n}})(X_{(k+1)2^{-n}} - X_{k2^{-n}}) + \\
&\quad + \frac{1}{2} \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} f''(X_{k2^{-n}})(X_{(k+1)2^{-n}} - X_{k2^{-n}})^2 + \text{error terms} \\
&\xrightarrow{u.c.p.} f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d[X]_s. \quad (3.67)
\end{aligned}$$

We will not follow this method of proof, because the error terms are hard to deal with.

Proof. (for $d = 1$)

Write $X = X_0 + M + A$, where A has total variation process V . Let

$$T_r = \inf \{t \geq 0 : |X_t| + V_t + [M]_t > r\}. \quad (3.68)$$

Then $(T_r)_{r \geq 0}$ is a family of stopping times with $T_r \nearrow \infty$. It is sufficient to prove (3.65) on the time intervals $[0, T_r]$. Let $\mathcal{A} \subseteq C^2(\mathbb{R}, \mathbb{R})$ denote the subset of functions f for which the formula holds. Then

- (i) \mathcal{A} contains the functions $f(x) \equiv 1$ and $f(x) = x$. (ii) \mathcal{A} is a vector space.

Below we will show that \mathcal{A} is, in fact, an algebra, i.e. in addition

$$(iii) f, g \in \mathcal{A} \Rightarrow fg \in \mathcal{A}.$$

Finally we will show that

- (iv) if $f_n \in \mathcal{A}$ and $f_n \rightarrow f$ in $C^2(B_r, \mathbb{R})$ for all $r > 0$ then $f \in \mathcal{A}$, where $f_n \rightarrow f$ in $C^2(B_r, \mathbb{R})$ means that $\Delta_{n,r} \rightarrow 0$ as $n \rightarrow \infty$ with $B_r = \{x : |x| \leq r\}$ and

$$\Delta_{n,r} := \max \left\{ \sup_{x \in B_r} |f_n(x) - f(x)|, \sup_{x \in B_r} |f'_n(x) - f'(x)|, \sup_{x \in B_r} |f''_n(x) - f''(x)| \right\}. \quad (3.69)$$

(i) - (iii) imply that \mathcal{A} contains all polynomials. By Weierstrass' approximation theorem, these are dense in $C^2(B_r, \mathbb{R})$ and so (iv) implies $\mathcal{A} = C^2(B_r, \mathbb{R})$.

Proof of (iii): Suppose $f, g \in \mathcal{A}$ and set $F_t = f(X_t)$, $G_t = g(X_t)$. Since the formula holds for f and g , F and G are continuous semimartingales. Integration by parts (Theorem ??) yields

$$F_t G_t - F_0 G_0 = \int_0^t F_s dG_s + \int_0^t G_s dF_s + [F, G]_t. \quad (3.70)$$

By Proposition 3.10(iv) we have $F \cdot G = F \cdot (1 \cdot G)$, and using Itô's formula for $(1 \cdot G)_s = g(X_s) - g(X_0)$ we get again by Proposition 3.10(iv)

$$\int_0^t F_s dG_s = \int_0^t f(X_s) g'(X_s) dX_s + \frac{1}{2} \int_0^t f(X_s) g''(X_s) d[X]_s. \quad (3.71)$$

By the Kunita-Watanabe identity (Proposition 3.14) we have $[f' \cdot X, G] = f' \cdot [X, G]$. Applying this a second time for G leads to

$$[F, G]_t = [f'(X) \cdot X, g'(X) \cdot X]_t = \int_0^t f'(X_s) g'(X_s) d[X]_s. \quad (3.72)$$

Substituting these into (3.70), we obtain Itô's formula for $f g$.

Proof of (iv): Let $f_n \in \mathcal{A}$ such that $f_n \rightarrow f$ in $C^2(B_r, \mathbb{R})$. Then

$$\begin{aligned} \int_0^{t \wedge T_r} |f'_n(X_s) - f'(X_s)| dV_s + \frac{1}{2} \int_0^{t \wedge T_r} |f''_n(X_s) - f''(X_s)| d[M]_s &\leq \\ &\leq \Delta_{n,r} (V_{t \wedge T_r} + \frac{1}{2} [M]_{t \wedge T_r}) \leq r \Delta_{n,r} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (3.73)$$

and so

$$\int_0^{t \wedge T_r} f'_n(X_s) dA_s + \frac{1}{2} \int_0^{t \wedge T_r} f''_n(X_s) d[M]_s \rightarrow \int_0^{t \wedge T_r} f'(X_s) dA_s + \frac{1}{2} \int_0^{t \wedge T_r} f''(X_s) d[M]_s.$$

Moreover, $M^{T_r} \in \mathcal{M}_c^2$ and so

$$\begin{aligned} \left\| (f'_n(X) \cdot M)^{T_r} - (f'(X) \cdot M)^{T_r} \right\|^2 &= \mathbb{E} \left(\int_0^{T_r} (f'_n(X_s) - f'(X_s))^2 d[M]_s \right) \\ &\leq \Delta_{n,r}^2 \mathbb{E}([M]_{T_r}) \leq r \Delta_{n,r}^2 \rightarrow 0, \end{aligned} \quad (3.74)$$

as $n \rightarrow \infty$ and so $(f'_n(X) \cdot M)^{T_r} \rightarrow (f'(X) \cdot M)^{T_r}$ in \mathcal{M}_c^2 . For any fixed r , $X^{T_r} \in B_r$ and taking the limit $n \rightarrow \infty$ in Itô's formula for f_n we obtain

$$f(X_{t \wedge T_r}) = f(X_0) + \int_0^{t \wedge T_r} f'(X_s) dX_s + \frac{1}{2} \int_0^{t \wedge T_r} f''(X_s) d[X]_s. \quad (3.75)$$

□

Remark. For $d > 1$, (i) becomes ' \mathcal{A} contains the constant 1 and the coordinate functions $f_1(x) = x^1, \dots, f_d(x) = x^d$.' Check that you can then follow the same argument, dealing with all the different components $X^i, M^i, [M^i, M^j]$ etc.

Corollary 3.18 *Let X^1, X^2, \dots, X^d be continuous semimartingales and set $X = (X^1, \dots, X^d)$. Let $f \in C^2(\mathbb{R}_+ \times \mathbb{R}^d, \mathbb{R})$. Then*

$$f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s) ds + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s) dX_s^i \\ + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) d[X^i, X^j]_s .$$

Proof. This is an immediate consequence of (3.65). Indeed, the process $t \mapsto t$ is nondecreasing and so of finite variation, so (t, X_t^1, \dots, X_t^d) is a $(d+1)$ -dimensional semi-martingale. The result follows by applying Itô's formula to this $d+1$ -dimensional process, and observing that since $t \mapsto t$ is of finite variation, it does not contribute to any of the covariation terms.

4 Applications to Brownian motion and Martingales

4.1 Brownian martingales

As we will see in a few moments, martingales are very useful to understand (and ultimately prove results about) the behaviour of random processes. We start our study of Brownian motion by finding martingales associated with it. If $x, y \in \mathbb{C}^d$, we note $\langle x, y \rangle = \sum_{i=1}^d x_i \bar{y}_i$ their complex scalar product.

Theorem 4.1 *Let $(B_t, t \geq 0)$ be an (\mathcal{F}_t) -Brownian motion.*

(i) *If $d = 1$ and $B_0 \in L^1$, the process $(B_t, t \geq 0)$ is a (\mathcal{F}_t) -martingale.*

(ii) *If $d = 1$ and $B_0 \in L^2$, the process $(B_t^2 - t, t \geq 0)$ is a (\mathcal{F}_t) -martingale.*

(iii) *In any dimension, let $u = (u_1, \dots, u_d) \in \mathbb{C}^d$. If $E[|\exp(\langle u, B_0 \rangle)|] < \infty$, the process defined by*

$$M_t = \exp(\langle u, B_t \rangle - tu^2/2)$$

is also a (\mathcal{F}_t) -martingale for every $u \in \mathbb{C}^d$, where u^2 is a notation for $\sum_{i=1}^d u_i^2$.

Notice that in (iii), we are dealing with \mathbb{C} -valued processes. The definition of $E[X|\mathcal{G}]$ the conditional expectation for a random variable $X \in L^1(\mathbb{C})$ is $E[\Re X|\mathcal{G}] + iE[\Im X|\mathcal{G}]$, and we say that an integrable process $(X_t, t \geq 0)$ with values in \mathbb{C} , and adapted to a filtration (\mathcal{F}_t) , is a martingale if its real and imaginary parts are. Notice that the hypothesis on B_0 in (iii) is automatically satisfied whenever $u = iv$ is purely imaginary, i.e., $v \in \mathbb{R}$.

Proof. (i) If $s \leq t$, $\mathbb{E}[B_t - B_s | \mathcal{F}_s] = \mathbb{E}[B_{t-s}^{(s)}] = 0$, where $B_u^{(s)} = B_{u+s} - B_s$ has mean 0 and is independent of \mathcal{F}_s , by the simple Markov property. The integrability of the process is obvious by hypothesis on B_0 .

(ii) Integrability is an easy exercise using that $B_t - B_0$ is independent of B_0 . We have, for $s \leq t$, $B_t^2 = (B_t - B_s)^2 + 2B_s(B_t - B_s) + B_s^2$. Taking conditional expectation given \mathcal{F}_s and using the simple Markov property gives that $\mathbb{E}[B_t^2] = (t - s) + B_s^2$, hence the result. A proof using Itô's formula is to say that B is an \mathcal{F} -local martingale and hence

$$B_t^2 = 2 \int_0^t B_s dB_s + t.$$

Thus $M_t = \int_0^t B_s dB_s$ is an \mathcal{F} -local martingale. It thus suffices to show that it is a true martingale, which can be proved for instance by observing that the quadratic variation is

$$[M]_t = \int_0^t B_s^2 ds$$

which has finite expectation for all $t > 0$ by Fubini's theorem. By problem 11 in Example sheet 2, $(M_{s \wedge t}, s \geq 0)$ is a martingale bounded in L^2 and hence M is a true martingale.

(iii) To check integrability, note that $E[\exp(\lambda B_t)] = \exp(t\lambda^2/2)$ whenever B is a standard Brownian motion, and since $|e^z| = e^{\Re z}$, then we have

$$\begin{aligned}\mathbb{E}(|\exp\langle u, B_t \rangle|) &= \mathbb{E}(|\exp\langle u, (B_t - B_0 + B_0) \rangle|) \\ &= \mathbb{E}(|\exp\langle u, B_t - B_0 \rangle|)\mathbb{E}(|\exp\langle u, B_0 \rangle|) \\ &= \exp\left(\sum_{i=1}^d t(\Re u_i)^2/2\right)\mathbb{E}|\exp\langle u, B_0 \rangle| < \infty.\end{aligned}$$

To show that M is a martingale, consider $X_{d+1} = t$ which is a continuous semi-martingale. Let $f(x_1, \dots, x_d, x_{d+1}) = \exp(\sum_{i=1}^d u_i x_i - (1/2)u^2 x_{d+1})$. $f \in C^2(\mathbb{R}^{d+1}, \mathbb{C})$ so we may apply Itô's formula and obtain:

$$M_t = M_0 + \int_0^t \sum_{i=1}^d u_i \exp(\langle u, B_s \rangle - su^2/2) dB_s^i$$

since $d[B^i, B^j]_t = \delta_{i,j} dt$ and $[B^i, t] = 0$ for all $1 \leq i, j \leq d$, so that the finite variations term cancel. It thus suffices to show that: $\int_0^t u_i \exp(\langle u, B_s \rangle - su^2/2) dB_s^i$ is a true martingale. We take the quadratic variation of the real and imaginary parts, and it suffices by Fubini's theorem to show that

$$\int_0^t \mathbb{E} \left(\exp\left(\sum_{i=1}^d 2r_i B_s^i - s\langle u, \bar{u} \rangle^2\right) \right) ds < \infty \quad (4.1)$$

where r_i is the complex modulus of u_i and $\langle u, \bar{u} \rangle^2 = \sum_{i=1}^d r_i^2$. (4.1) follows instantly from the independence of the coordinates and the fact that $E[\exp(rB_t)] = \exp(tr^2/2)$. \square

A classical application of these martingales is to show the following result, often referred to as the *gambler's ruin estimates*.

Theorem 4.2 *Let $(B_t, t \geq 0)$ be a standard Brownian motion and $T_x = \inf\{t \geq 0 : B_t = x\}$. Then for $x, y > 0$, one has*

$$P(T_{-y} < T_x) = \frac{x}{x+y}, \quad E[T_x \wedge T_{-y}] = xy.$$

Proof. Let $T = T_{-y} \wedge T_x$, which is a stopping time. Moreover, B^T is bounded (by $\max(x, y)$) so we may apply the optional stopping theorem to find that $\mathbb{E}(B_T) = \mathbb{E}(B_0) = 0$. On the other hand, $\mathbb{E}(B_T) = -yp + x(1-p)$, where $p = P(T_{-y} < T_x)$ is the probability of interest to us. Thus $py = (1-p)x$ and the first statement follows easily. For the second statement, observe that $B_{t \wedge T}^2 - (t \wedge T)$ is a martingale (since martingales are stable by stopping) and thus

$$\mathbb{E}(B_{t \wedge T}^2) = \mathbb{E}(t \wedge T).$$

We may let $t \rightarrow \infty$ since the left-hand side is bounded and the right-hand side is monotone, and deduce that

$$\mathbb{E}(B_T^2) = \mathbb{E}(T)$$

Using the first statement,

$$\mathbb{E}(B_T^2) = \frac{x}{x+y}y^2 + \frac{y}{y+x}x^2 = xy$$

and the claim follows. \square

Similarly,

Theorem 4.3 *The Laplace transform of T_x for $x \in \mathbb{R}$ is given by $\mathbb{E}(e^{-qT_x}) = e^{-|x|\sqrt{2q}}$. Moreover, the random variable $T = T_x \wedge T_{-y}$ has a Laplace transform given by*

$$\mathbb{E}(e^{-qT}) = \frac{\sinh(\sqrt{2q}x) + \sinh(\sqrt{2q}y)}{\sinh(\sqrt{2q}(x+y))}$$

and when $y = x$, T is independent from the event $\{T_{-x} < T_x\}$.

Proof. The first statement follows directly from the optional stopping theorem and the fact that $e^{\lambda B_t - (\lambda^2/2)t}$ is a martingale. The second statement is a bit more involved. Let

$$M_t = e^{-\lambda^2 t/2} \sinh(\lambda(B_t + y))$$

is also a martingale since it can be written as

$$\frac{1}{2}e^{-\lambda^2 t/2} e^{\lambda(B_t + y)} - \frac{1}{2}e^{-\lambda^2 t/2} e^{-\lambda(B_t + y)}$$

which is the sum of two martingales. Now, stopping at $T = T_{-x} \wedge T_y$, M is bounded so we can use the optional stopping theorem to obtain:

$$\begin{aligned} \sinh(\lambda y) &= \mathbb{E}(\sinh(\lambda(B_T + y))e^{-T\lambda^2/2}) \\ &= \mathbb{E}(\sinh(\lambda((x+y))e^{-T\lambda^2/2} \mathbf{1}_{\{T_x > T_{-y}\}}) \end{aligned}$$

Thus:

$$\mathbb{E}(e^{-T\lambda^2/2} \mathbf{1}_{\{T_x > T_{-y}\}}) = \frac{\sinh(\lambda y)}{\sinh(\lambda(x+y))}.$$

By symmetry,

$$\mathbb{E}(e^{-T\lambda^2/2} \mathbf{1}_{\{T_{-y} > T_x\}}) = \frac{\sinh(\lambda x)}{\sinh(\lambda(x+y))}.$$

Adding up the two terms,

$$\mathbb{E}(e^{-T\lambda^2/2}) = \frac{\sinh(\lambda y) + \sinh(\lambda x)}{\sinh(\lambda(x+y))}$$

When $x = y$, it suffices to check that

$$\mathbb{E}(e^{-T\lambda^2/2} \mathbf{1}_{\{T_{-y} > T_x\}}) = \mathbb{E}(e^{-T\lambda^2/2})P(T_{-y} < T_x) = \frac{1}{2}\mathbb{E}(e^{-T\lambda^2/2})$$

which is easy to check. \square

Another family of martingales is provided by the result below. This is the first hint of a deep connection between Brownian motion and second-order elliptic partial differential operators, a theme which we will explore in greater detail later on in the course. (This also connects to the theory of *martingale problems* developed by Stroock and Varadhan, which has proved to be one of the most successful tools in probability theory).

Theorem 4.4 Let $(B_t, t \geq 0)$ be a (\mathcal{F}_t) -Brownian motion. Let $f(t, x) : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{C}$ be continuously differentiable in the variable t and twice continuously differentiable in x . Then,

$$M_t^f = f(t, B_t) - f(0, B_0) - \int_0^t ds \left(\frac{\partial}{\partial t} + \frac{1}{2} \Delta \right) f(s, B_s), \quad t \geq 0$$

is a (\mathcal{F}_t) -local martingale, where $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ is the Laplacian operator acting on the spatial coordinate of f . If moreover, the first derivatives are uniformly bounded on every compact interval (that is, for all $T > 0$,

$$\sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^d} \left| \frac{\partial f}{\partial x_i}(t, x) \right| < \infty$$

for all $1 \leq i \leq d$), then M^f is a true martingale.

Proof. By Itô's formula,

$$M_t^f = \int_0^t \sum_{i=1}^d \frac{\partial f}{\partial x_i}(s, B_s) dB_s^i$$

is indeed a local martingale. The fact it is a true martingale when the first partial derivatives are uniformly bounded on every compact time interval, follows from the fact that the quadratic variation of M^f is bounded on every compact time interval, and hence it is a true martingale (even bounded in L^2) on every compact time interval.

4.2 Martingales characterized by their quadratic variation

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ where $(\mathcal{F}_t)_{t \geq 0}$ satisfies the usual conditions.

Theorem 4.5 Lévy's characterization of Brownian motion

Let $X^1, \dots, X^d \in \mathcal{M}_{c,loc}$. The two following statements are equivalent.

(i) For all $t \geq 0$, $[X^i, X^j]_t = \delta_{ij}t$.

(ii) $X = (X^1, \dots, X^d)$ is a Brownian motion in \mathbb{R}^d .

Proof. It suffices to show that, for $0 \leq s \leq t$, $X_t - X_s \sim N(0, (t-s)I)$ and the increment is independent of \mathcal{F}_s . By uniqueness of characteristic functions, this is equivalent to showing that for all $s \leq t$ and for all $\theta \in \mathbb{R}^d$,

$$\mathbb{E}(\exp(i\langle \theta, X_t - X_s \rangle) | \mathcal{F}_s) = \exp(-\frac{1}{2}\|\theta\|^2(t-s)). \quad (4.2)$$

(Here $\langle \cdot, \cdot \rangle$ is the usual scalar product on \mathbb{R}^d and $\|\theta\|$ is the Euclidean norm). Fix $\theta \in \mathbb{R}^d$ and set

$$Y_t = \langle \theta, X_t \rangle = \theta_1 X_t^1 + \dots + \theta_d X_t^d$$

. Then Y is a local martingale, and by the assumptions and the bilinearity of the covariation, $[Y] = \sum_{i=1}^d \theta_i^2 t = t\|\theta\|^2$. Define also and

$$Z_t = \exp(iY_t + \frac{1}{2}[Y]_t) = \exp(i\langle \theta, X_t \rangle + \frac{1}{2}t\|\theta\|^2).$$

By Itô's formula, it is easy to check that Z is a local martingale. Moreover, Z is bounded on $[0, t]$ for all $t \geq 0$ (since $[Y]_t = \|\theta\|^2 t$) and so is a true martingale by Proposition 2.9. Hence, $\mathbb{E}(Z_t | \mathcal{F}_s) = Z_s$, or equivalently:

$$\mathbb{E} \left(\frac{Z_t}{Z_s} \middle| \mathcal{F}_s \right) = 1, \quad a.s.$$

(4.2) follows directly. \square

Proposition 4.6 *Let B be a Brownian motion and let h be a deterministic function in $L^2(\mathbb{R}^+, \mathcal{B}, \lambda)$ with Lebesgue measure λ . Set $X = \int_0^\infty h_s dB_s$. Then $X \sim N(0, \|h\|_2^2)$.*

Proof. Let $M_t = \int_0^t h_s dB_s$. Then $M \in \mathcal{M}_{c,loc}$ and (since $[h \cdot B] = h^2 \cdot [B]$) we have $[M]_t = \int_0^t h_s^2 ds$. Now let $Z_t = \exp(iuM_t + \frac{1}{2}u^2[M]_t)$. This is in \mathcal{M}_{loc} as in the proof of Lévy's characterization and, as it is uniformly bounded by $\exp(\frac{1}{2}u^2\|h\|_2^2)$, is in fact in \mathcal{M}^2 . Hence,

$$1 = \mathbb{E}(Z_0) = \mathbb{E}(Z_\infty) = \mathbb{E}(\exp(iuX)) \exp(\frac{1}{2}u^2\|h\|_2^2). \quad (4.3)$$

\square

Theorem 4.7 Dubins-Schwarz Theorem

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$ and $[M]_\infty = \infty$ a.s.. Set $\tau_s = \inf \{t \geq 0 : [M]_t > s\}$, $B_s = M_{\tau_s}$. Then τ_s is an $(\mathcal{F}_t)_{t \geq 0}$ -stopping time. If $\mathcal{G}_s = \mathcal{F}_{\tau_s}$ then $(\mathcal{G}_s)_{s \geq 0}$ is a filtration and B is a $(\mathcal{G}_t)_{t \geq 0}$ -Brownian motion. Moreover $M_t = B_{[M]_t}$.

Remark. So any continuous local martingale is a (stochastic) time-change of Brownian motion. In this sense, Brownian motion is the most general continuous local martingale.

Proof. Since $[M]$ is continuous and adapted, τ_s is a stopping time, and since $[M]_\infty = \infty$ it must be that $\tau_s < \infty$ a.s. for all $s \geq 0$. We start the proof by the following lemma.

Lemma 4.8 *B is a.s. continuous*

Proof. Note that $s \mapsto \tau_s$ is càdlàg and nondecreasing and thus B is càdlàg. So it remains to show $B_{s-} = B_s$ for all $s > 0$, or equivalently $M_{\tau_{s-}} = M_{\tau_s}$, where

$$\tau_{s-} = \inf \{t \geq 0 : [M]_t = s\} \quad (4.4)$$

and note that τ_{s-} is also a stopping time. Let $s > 0$. We need to show that M is constant between τ_{s-} and τ_s whenever $\tau_{s-} < \tau_s$, i.e. whenever $[M]$ is constant. Note that $(M^2 - [M])^{\tau_s}$ is uniformly integrable since $\mathbb{E}([M]^{\tau_s}) < \infty$. Hence, by the optional stopping theorem (the uniformly integrable version 2.36), we get:

$$\mathbb{E}(M_{\tau_s}^2 - [M]_{\tau_s} \mid \mathcal{F}_{\tau_{s-}}) = M_{\tau_{s-}}^2 - [M]_{\tau_{s-}}.$$

But by assumption, $[M]_{\tau_s} = [M]_{\tau_{s-}}$ and M is a martingale, we obtain

$$\mathbb{E}(M_{\tau_s}^2 - M_{\tau_{s-}}^2 \mid \mathcal{F}_{\tau_{s-}}) = \mathbb{E}((M_{\tau_s} - M_{\tau_{s-}})^2 \mid \mathcal{F}_{\tau_{s-}}) = 0$$

and so M is *a.s.* constant between τ_{s-} and τ_s . This proves that B is almost surely continuous at time s . To prove that B is *a.s.* continuous simultaneously for all $s \geq 0$, note that if $T_r = \inf\{t > 0 : M_t \neq M_r\}$ and $S_r = \inf\{t > 0 : [M]_t \neq [M]_r\}$ then the previous argument says that for all fixed $r > 0$ (and hence for all $r \in \mathbb{Q}_+$), $T_r = S_r$ *a.s.* But observe that T_r and S_r are both càdlàg. Thus equality holds almost surely for all $r \geq 0$ and hence almost surely, M and $[M]$ are constant on the same intervals. This implies the result. \square

We also need the following lemma.

Lemma 4.9 *B is adapted to $(\mathcal{G}_t)_{t \geq 0}$.*

Proof. It is trivial to check that $(\mathcal{G}_s)_{s \geq 0}$ is a filtration. Indeed, if $S \leq T$ *a.s.* are two stopping times for the complete filtration (\mathcal{F}_t) , and if $A \in \mathcal{F}_S$, then for all $t \geq 0$,

$$A \cap \{T \leq t\} = (A \cap \{S \leq t\}) \cap \{T \leq t\}$$

up to zero-probability events. The first event in the right-hand side is in \mathcal{F}_t by assumption, and the second is also in \mathcal{F}_t since T is a stopping time. Since (\mathcal{F}_t) is complete, we conclude that $A \in \mathcal{F}_T$ as well, and hence $\mathcal{F}_S \subseteq \mathcal{F}_T$. From this, since $\tau_r \leq \tau_s$ almost surely if $r \leq s$, (\mathcal{G}_s) is a filtration. It thus suffices to show that if X is a càdlàg adapted process and T is a stopping time, then $X_T \mathbf{1}_{\{T < \infty\}}$ is \mathcal{F}_T -measurable. Note that a random variable Z is \mathcal{F}_T -measurable if $Z \mathbf{1}_{\{T \leq t\}} \in \mathcal{F}_t$ for every $t \geq 0$. If T only takes countably many values $\{t_k\}_{k=1}^\infty$, then

$$X_T \mathbf{1}_{\{T < \infty\}} = \sum_{k=1}^{\infty} X_{t_k} \mathbf{1}_{\{T=t_k\}}$$

so it is trivial to check that $X_T \mathbf{1}_{\{T < \infty\}}$ is \mathcal{F}_T -measurable, since every term in the above sum is. In the general case, let $T_n = 2^{-n} \lceil 2^n T \rceil$ where $\lceil x \rceil$ denotes smallest $n \in \mathbb{Z}_+$ with $n \geq x$. Then T_n is also a stopping time, finite whenever $T < \infty$, and such that $T_n \geq T$ while $T_n \rightarrow T$ almost surely. Thus for all $u \geq 0$, and for all $n \geq 1$, $X_{T_n} \mathbf{1}_{\{T_n \leq u\}}$ is \mathcal{F}_u -measurable. Furthermore, by right-continuity of X , $\lim_{n \rightarrow \infty} X_{T_n} \mathbf{1}_{\{T_n \leq u\}} = X_T \mathbf{1}_{\{T < u\}}$. Thus $X_T \mathbf{1}_{\{T < u\}}$ is \mathcal{F}_u -measurable. Naturally, $X_T \mathbf{1}_{\{T=u\}} = X_u \mathbf{1}_{\{T=u\}}$ is also \mathcal{F}_u -measurable, so we deduce that $X_T \mathbf{1}_{\{T \leq u\}}$ is \mathcal{F}_u -measurable. \square

Having proved this lemma, we can now finish the proof of the Dubins-Schwarz theorem. Fix $s > 0$. Then $[M^{\tau_s}]_\infty = [M]_{\tau_s} = s$, by continuity of $[M]$. Thus by example sheet 2, problem 11, $M^{\tau_s} \in \mathcal{M}_c^2$ since $\mathbb{E}([M^{\tau_s}]_\infty) < \infty$. In particular, $(M_{t \wedge \tau_s}, s \geq 0)$ is uniformly integrable by Doob's inequality. Applying the uniformly integrable version of the optional stopping theorem (2.36) a first time, we obtain

$$\mathbb{E}(M_{\tau_s} | \mathcal{F}_{\tau_r}) = M_{\tau_r}$$

a.s. and thus B is a \mathcal{G} -martingale. Furthermore, since $M^{\tau_s} \in \mathcal{M}_c^2$, by Theorem 3.6, $(M^2 - [M])^{\tau_s}$ is also a uniformly integrable martingale. By (2.36) again, for $r \leq s$,

$$\begin{aligned} \mathbb{E}(B_s^2 - s | \mathcal{G}_r) &= \mathbb{E}((M^2 - [M])_{\tau_s} | \mathcal{F}_{\tau_r}) \\ &= M_{\tau_r}^2 - [M]_{\tau_r} = B_r^2 - r. \end{aligned}$$

Hence, $B \in \mathcal{M}_c$ with $[B]_s = s$ and so, by Lévy's characterization, B is a $(\mathcal{G}_t)_{t \geq 0}$ -Brownian motion. \square

Before we head on to our next topic, here are a few complements to this theorem, given without proof. The first result is a strengthening of the Dubins-Schwarz theorem.

Theorem 4.10 (*Dubins-Schwarz*) *Let M be a continuous local martingale with $M_0 = 0$ a.s. Then we may enlarge the probability space and define a Brownian motion B on it in such a way that*

$$M = B_{[M]_t} \text{ a.s. for all } t \geq 0.$$

More precisely, taking an independent Brownian motion β , if

$$B_s = \begin{cases} M_{\tau_s} & \text{for } s \leq [M]_\infty \\ M_\infty + \beta_{s-[M]_\infty} & \text{for all } s \geq [M]_\infty \end{cases}$$

then B is a Brownian motion and for all $t \geq 0$, $M_t = B_{[M]_t}$.

See Revuz-Yor (Chapter V, Theorem (1.10)) for a proof.

Remark. One informal (but very informative!) conclusion that one can draw from this theorem is the fact that *the quadratic variation should be regarded as a natural clock for the martingale*. This is demonstrated for instance in the following theorem.

Theorem 4.11 *Let M be a continuous local martingale. Then*

- (i) $P(\lim_{t \rightarrow \infty} |M|_t = \infty) = 0$
- (ii) $\{\omega : \lim_{t \rightarrow \infty} M_t \text{ exists and is finite}\} = \{\omega : [M]_\infty < \infty\}$ up to null events.
- (iii) $\{[M]_\infty = \infty\} = \{\limsup_{t \rightarrow \infty} M_t = +\infty \text{ and } \liminf_{t \rightarrow \infty} M_t = -\infty\}$ up to null events.

Another fundamental application of the Lévy characterization is the following result, which tells us that Brownian motion is *conformally invariant*. That is, given a conformal mapping f between two simply-connected planar domains D and D' , the trajectories of a Brownian motion in D' (that is, stopped upon reaching the boundary $\partial D'$) are the image by f of Brownian motion in D , up to reparametrization.

Theorem 4.12 *Let $d = 2$ and identify \mathbb{R}^2 with the complex field \mathbb{C} . Let $f : D \rightarrow D'$ be a conformal map (i.e., complex analytic). Let $z \in D$ and let $\tau_D = \inf\{t > 0 : B_t \notin D\}$. Then there exists a nondecreasing random function $(\sigma(t), t \geq 0)$ such that $(f(B_{\sigma(t)}), t \geq 0)$ has the same distribution as $(B'_{t \wedge \tau_{D'}}, t \geq 0)$ where B' is a Brownian motion started from $f(z)$.*

The proof is one of the exercises of Example Sheet 3, where you will be asked to explicit the stochastic time-change $\sigma(t)$. It will be useful to recall the *Cauchy-Riemann equations* for complex analytic functions: if $f = u + iv$ is a complex-differentiable function with real and imaginary parts u and v , then

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \end{cases}$$

from which it follows by further differentiation that both u and v are harmonic functions (i.e., $\Delta u \equiv \Delta v \equiv 0$) all over D). Applying Itô's formula and the Cauchy-Riemann equations shows that the real and imaginary parts of $f(B_t)$ have zero covariation and that they have identical quadratic variation. Applying the ideas of the Dubins-Schwartz theorem yields the result.

In principle, Theorem 4.12 (in combination with the famous Riemann mapping theorem) can be used to extract all the information we need about the behaviour of Brownian motion. For instance, the exit distribution from a domain D is simply the conformal image of the uniform measure of the circle by a map from the disc to this domain.

Remark. The ramifications of this result are huge. On the one hand, it serves as a bridge between probability and complex analysis, and in the example sheet you will prove a result of complex analysis just using Brownian motion. This connection is one aspect of the deep connection between random processes and harmonic analysis (which will be further developed later on). On the other hand, conformal invariance of Brownian motion, already observed by Paul Lévy in the 1940's, can be seen as the starting point of one of the most fascinating recent theories developed in probability, which is that of SLE (for Schramm-Loewner Evolution) and conformally invariant processes in the plane.

4.3 Recurrence and Transience of Brownian motion

As explained in the introduction to these notes, Brownian motion is the *scaling limit* of d -dimensional random walks (this theorem will actually be proved in its strong form in the next subsection). One of the most striking results about random walks is Polya's theorem which says that simple random walk is recurrent in dimension 1 and 2, while it is transient in dimension 3. What is the situation for Brownian motion? Being the scaling limit of simple random walk, one might expect the answer to be the same for Brownian motion. It turns out that this is almost the case: there is however something subtle happening in dimension 2. In the planar case, Brownian motion is neighbourhood-recurrent (it visits any neighbourhood of any point "infinitely often") but almost surely does not hit any point chosen in advance.

We work with the Wiener measure \mathbb{W} on the space of continuous functions, and recall that \mathbb{W}_x denote the law of a Brownian motion started at x . Let \mathbb{E}_x denote the expectation under this probability measure. In the sequel, $B(x, r)$ and $\bar{B}(x, r)$ denote the Euclidean ball of radius r about $x \in \mathbb{R}^d$.

Theorem 4.13 (i) *If $d = 1$, Brownian motion is point-recurrent in the sense that:*

$$\mathbb{W}_0 - a.s. \text{ for all } x \in \mathbb{R}, \{t \geq 0 : B_t = x\} \text{ is unbounded}$$

(ii) *If $d \geq 3$, Brownian motion is transient, in the sense that $\|B_t\| \rightarrow \infty$ almost surely as $t \rightarrow \infty$.*

(iii) *If $d = 2$, Brownian motion is neighbourhood-recurrent, in the sense that for every $x \in \mathbb{R}^d$, every open set is visited infinitely often \mathbb{W}_x -almost surely. Equivalently, for any $\varepsilon > 0$,*

$$\{t \geq 0 : \|B_t\| < \varepsilon\} \text{ is unbounded}$$

\mathbb{W}_x -almost surely for every $x \in \mathbb{R}^2$. However, points are polar in the sense that for every $x \in \mathbb{R}^2$,

$$\mathbb{W}_0(B_t = x \text{ for some } t > 0) = 0.$$

Proof. (i) is a consequence of (ii) in Proposition ??.

(ii) Let $B = (B^1, \dots, B^d)$ be a d -dimensional Brownian motion with $d \geq 3$. Clearly it suffices to prove the result for $d = 3$ since

$$\|B_t\|^2 \geq R_t^2 := \sum_{i=1}^3 \|B_t^i\|^2$$

and we are precisely claiming that the right-hand side tends to infinity as $t \rightarrow \infty$. Now, for $d = 3$, a simple calculation shows that if $f(x) = 1/\|x\|$, then $\Delta f = 0$. Thus by Itô's formula,

$$1/R_t \text{ is a local martingale.}$$

Since it is nonnegative, it follows from Proposition 2.14 that it is a supermartingale. Being nonnegative, the martingale convergence theorem tells us that it has an almost sure limit M as $t \rightarrow \infty$, and it suffices to prove that $M = 0$ almost surely. Now on the event $\{M > 0\}$, R_t must be bounded, and thus so is $|B_t^1|$. This has probability 0 by (i) and hence $M = 0$ a.s.

(iii) Let $d = 2$ and let B be a planar Brownian motion. Assume without loss of generality that $B_0 = 1$. We are going to establish that starting from there, B will never hit 0 but will come close to it “infinitely often” (or rather, “unboundedly often”). For $k \in \mathbb{Z}$, let $R_k = e^k$ and let

$$\tau_k = \inf\{t \geq 0 : \|B_t\| = R_k\}$$

and let

$$\tau = \tau_{-\infty} = \inf\{t \geq 0 : \|B_t\| = 0\}.$$

Our first goal will be to show that $\tau = \infty$, almost surely. Define a sequence of stopping times T_n as follows: $T_0 = 0$, and by induction if $Z_n = \|B_{T_n}\|$ then

$$T_{n+1} = \inf\{t \geq T_n, \|B_t\| \in \{e^{-1}Z_n, eZ_n\}\}.$$

Notice that if $k, m \geq 1$ are two large integers, the probability that $\tau_{-k} < \tau_m$ is the probability that Z_n visits e^{-k} before e^m . Put it another way, it is also the probability that $(\log Z_n, n \geq 0)$ visits $-k$ before m .

On the other hand, by Itô's formula, $M_t = \log \|B_{t \wedge \tau}\|$ is a local martingale since

$$(x, y) \mapsto \log(x^2 + y^2) \text{ is harmonic on } \mathbb{R}^2$$

Since M_t is bounded on $[T_n; T_{n+1}]$, it follows from the Optional Stopping Theorem that given $\log Z_n = k \in \mathbb{Z}$,

$$P(\log Z_{n+1} = k + 1 | \log Z_n = k) = P(\log Z_{n+1} = k - 1 | \log Z_n = k) = 1/2.$$

Moreover, the strong Markov property of Brownian motion implies that $(\log Z_n, n \geq 0)$ is a Markov chain. In other words, $(\log Z_n, n \geq 0)$ is nothing but *simple random walk* on \mathbb{Z} . In particular, it is recurrent. Therefore, for any $m \geq 0$,

$$P(\tau_{-k} < \tau_m) \rightarrow 0$$

as $k \rightarrow \infty$. Therefore,

$$P(\tau < \tau_m) = 0$$

for all $m \geq 0$. This implies that $\tau = \infty$ almost surely since $\tau_m \rightarrow \infty$ as $m \rightarrow \infty$. On the other hand, this argument shows that $\tau_k < \infty$ for all $k \in \mathbb{Z}$, and there are infinitely many times that B visits this ball after returning to a radius greater than 1. Thus the set of times such that $B_t \in B(0, R_k)$ is unbounded a.s. \square

Remark. Notice that (iii) implies the fact that $\{t \geq 0 : B_t \in B(x; \varepsilon)\}$ is unbounded for every $x \in \mathbb{R}^2$ and every $\varepsilon > 0$, almost surely. Indeed, one can cover \mathbb{R}^2 by a countable union of balls of a fixed radius). In particular, the trajectory of a 2-dimensional Brownian motion is everywhere dense. On the other hand, it will a.s. never hit a fixed countable family of points (except maybe at time 0), like the points with rational coordinates!

4.4 Donsker's invariance principle

The following theorem completes the description of Brownian motion as a 'limit' of centered random walks as depicted in the beginning of the chapter, and strengthens the convergence of finite-dimensional marginals to that convergence in distribution.

We endow $\mathcal{C}([0, 1], \mathbb{R})$ with the supremum norm, and recall (see the exercises on continuous-time processes) that the product σ -algebra associated with it coincides with the Borel σ -algebra associated with this norm. We say that a function $F : \mathcal{C}([0, 1]) \rightarrow \mathbb{R}$ is continuous if it is continuous with respect to this norm. Often, functions F defined on \mathcal{C} will be called functionals. For instance, the supremum of a path on the interval $[0, 1]$ is a (continuous) functional.

Theorem 4.14 (Donsker's invariance principle) *Let $(X_n, n \geq 1)$ be a sequence of \mathbb{R} -valued integrable independent random variables with common law μ , such that*

$$\int x\mu(dx) = 0 \text{ and } \int x^2\mu(dx) = \sigma^2 \in (0, \infty).$$

Let $S_0 = 0$ and $S_n = X_1 + \dots + X_n$, and define a continuous process that interpolates linearly between values of S , namely

$$S_t = (1 - \{t\})S_{[t]} + \{t\}S_{[t]+1} \quad t \geq 0,$$

where $[t]$ denotes the integer part of t and $\{t\} = t - [t]$. Then

$$S^{[N]} := \left(\frac{S_{Nt}}{\sqrt{\sigma^2 N}}, 0 \leq t \leq 1 \right)$$

converges in distribution to a standard Brownian motion between times 0 and 1, i.e. for every bounded continuous function $F : \mathcal{C}([0, 1]) \rightarrow \mathbb{R}$,

$$\mathbb{E} \left[F(S^{[N]}) \right] \xrightarrow{n \rightarrow \infty} \mathbb{E}_0[F(B)].$$

Notice that this is much stronger than what Proposition 1.1 says. Despite the slight difference of framework between these two results (one uses càdlàg continuous-time version of the random walk, and the other uses an interpolated continuous version), Donsker's invariance principle is stronger. For instance, one can infer from this theorem that the random variable $N^{-1/2} \sup_{0 \leq n \leq N} S_n$ converges to $\sup_{0 \leq t \leq 1} B_t$ in distribution, because $f \mapsto \sup f$ is a continuous operation on $\mathcal{C}([0, 1], \mathbb{R})$. Proposition 1.1 would be powerless to address this issue.

The proof we give here is an elegant demonstration that makes use of a coupling of the random walk with the Brownian motion, called the Skorokhod embedding theorem. It is however specific to dimension $d = 1$. Suppose we are given a Brownian motion $(B_t, t \geq 0)$ on some probability space (Ω, \mathcal{F}, P) .

Let $\mu_+(dx) = P(X_1 \in dx)\mathbf{1}_{\{x \geq 0\}}$, $\mu_-(dy) = P(-X_1 \in dy)\mathbf{1}_{\{y > 0\}}$ define two non-negative measures. Assume that (Ω, \mathcal{F}, P) is a rich enough probability space so that we can further define on it, independently of $(B_t, t \geq 0)$, a sequence of independent identically distributed \mathbb{R}^2 -valued random variables $((Y_n, Z_n), n \geq 1)$ with distribution

$$\mathbb{P}((Y_n, Z_n) \in dx dy) = \frac{1}{C}(x + y)\mu_+(dx)\mu_-(dy),$$

where $C > 0$ is the appropriate normalizing constant that makes this expression a probability measure (this is possible because X has a well-defined expectation).

We define a sequence of random times, by $T_0 = 0, T_1 = \inf\{t \geq 0 : B_t \in \{Y_1, -Z_1\}\}$, and recursively,

$$T_n = \inf\{t \geq T_{n-1} : B_t - B_{T_{n-1}} \in \{Y_n, -Z_n\}\}.$$

By (ii) in Proposition ??, these times are a.s. finite, and they are stopping times with respect to the filtration (\mathcal{F}_t) . We claim that

Lemma 4.15 (Skorokhod's embedding) *The sequence $(B_{T_n}, n \geq 0)$ has the same law as $(S_n, n \geq 0)$. Moreover, the intertimes $(T_n - T_{n-1}, n \geq 1)$ form a sequence of independent random variables with same distribution, and expectation $E[T_1] = \sigma^2$.*

Proof. Let \mathcal{F}^B be the filtration of the Brownian motion, and for each $n \geq 0$, introduce the filtration $\mathcal{G}^n = (\mathcal{G}_t^n, t \geq 0)$ defined by

$$\mathcal{G}_t^n = \mathcal{F}_t^B \vee \sigma(Y_1, Z_1, \dots, Y_n, Z_n).$$

Since (Y_i, Z_i) are independent from \mathcal{F}^B , $(B_t, t \geq 0)$ is a \mathcal{G}^n -Brownian motion for every $n \geq 0$. Moreover, T_n is a stopping time for \mathcal{G}^n . It follows that if $\tilde{B}_t = (B_{T_n+t} - B_{T_n}, t \geq 0)$ then \tilde{B} is independent from $\mathcal{G}_{T_n}^n$. Moreover, (Y_{n+1}, Z_{n+1}) is independent both from $\mathcal{G}_{T_n}^n$ and from \tilde{B} , therefore $(T_{n+1} - T_n)$, which depends only on \tilde{B} and (Y_{n+1}, Z_{n+1}) is independent from $\mathcal{G}_{T_n}^n$. In particular, $(T_{n+1} - T_n)$ is independent from $\sigma(T_0, T_1, \dots, T_n)$. More generally, we obtain that the processes $(B_{t+T_{n-1}} - B_{T_{n-1}}, 0 \leq t \leq T_n - T_{n-1})$ are independent with the same distribution.

It therefore remains to check that B_{T_1} has the same law as X_1 and $\mathbb{E}[T_1] = \sigma^2$. Remember from Proposition 4.2 that given Y_1, Z_1 , the probability that $B_{T_1} = Y_1$ is

$Z_1/(Y_1 + Z_1)$, as follows from the optional stopping theorem. Therefore, for every non-negative measurable function f , by first conditioning on (Y_1, Z_1) , we get

$$\begin{aligned}
\mathbb{E}[f(B_{T_1})] &= \mathbb{E} \left[f(Y_1) \frac{Z_1}{Y_1 + Z_1} + f(-Z_1) \frac{Y_1}{Y_1 + Z_1} \right] \\
&= \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} \frac{1}{C} (x+y) \mu_+(dx) \mu_-(dy) \left(f(x) \frac{y}{x+y} + f(-y) \frac{x}{x+y} \right) \\
&= \frac{1}{C} \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} \mu_+(dx) \mu_-(dy) (yf(x) + xf(-y)) \\
&= \frac{1}{C} \int_{\mathbb{R}_+} \mu_+(dx) f(x) \int_{\mathbb{R}_+^*} y \mu_-(dy) + \frac{1}{C} \int_{\mathbb{R}_+^*} \mu_-(dy) f(-y) \int_{\mathbb{R}_+} x \mu_+(dx)
\end{aligned}$$

Now observe that since $\mathbb{E}(X_1) = 0$, it must be the case that

$$\int_{\mathbb{R}_+} x \mu_+(dx) = \int_{\mathbb{R}_+^*} y \mu_-(dy) = C',$$

say, and thus, the left hand side is equal to

$$\begin{aligned}
\mathbb{E}[f(B_{T_1})] &= \frac{C'}{C} \int_{\mathbb{R}_+} (f(x) \mu_+(dx) + f(-x) \mu_-(dx)) \\
&= \frac{C'}{C} \int_{\mathbb{R}} f(x) \mu(dx) \\
&= \frac{C'}{C} \mathbb{E}[f(X_1)].
\end{aligned}$$

By taking $f \equiv 1$, it must be that $C' = C$, and hence B_{T_1} has the same law as X_1 . For $\mathbb{E}[T_1]$, recall from Proposition 4.2 that $\mathbb{E}[\inf\{t \geq 0 : B_t \in \{x, -y\}\}] = xy$, so by a similar conditioning argument as above,

$$\begin{aligned}
\mathbb{E}[T_1] &= \int_{\mathbb{R}_+ \times \mathbb{R}_+^*} \frac{1}{C} (x+y) xy \mu_+(dx) \mu_-(dy) \\
&= \frac{1}{C} \int_{\mathbb{R}_+} x^2 \mu_+(dx) \int_{\mathbb{R}_+^*} y \mu_-(dy) + \frac{1}{C} \int_{\mathbb{R}_+^*} y^2 \mu_-(dy) \int_{\mathbb{R}_+} x \mu_+(dx) \\
&= \frac{C'}{C} \int_{\mathbb{R}} x^2 \mu(dx) \\
&= \frac{C'}{C} \sigma^2
\end{aligned}$$

but since we already know that $C' = C$, this shows that $\mathbb{E}(T_1) = \sigma^2$, as claimed. \square

We will need another lemma, which tells us that the times T_m are in a fairly strong sense localized around their mean $m\sigma^2$.

Lemma 4.16 *We have the following convergence as $N \rightarrow \infty$:*

$$N^{-1} \sup_{0 \leq n \leq N} |T_n - \sigma^2 n| \rightarrow 0 \quad a.s. \tag{4.5}$$

Proof. By the strong law of large numbers, note that $T_n/n \rightarrow \sigma^2$ almost surely. Thus, fix $\varepsilon > 0$. Then there exists $N_0 = N_0(\varepsilon, \omega)$ such that if $n \geq N_0$, $|n^{-1}T_n - \sigma^2| \leq \varepsilon$. Thus if $N_0 \leq n \leq N$, then

$$N^{-1}|T_n - n\sigma^2| \leq \frac{n}{N}\varepsilon \leq \varepsilon$$

Moreover, $N^{-1} \sup_{0 \leq n \leq N_0} |T_n - n\sigma^2|$ tends to 0 almost surely as $N \rightarrow \infty$, so this implies (4.5). \square

Proof of Donsker's invariance principle. We suppose given a Brownian motion B . For $N \geq 1$, define $B_t^{(N)} = N^{1/2}B_{N^{-1}t}$, $t \geq 0$, which is a Brownian motion by scaling invariance. Perform the Skorokhod embedding construction on $B^{(N)}$ to obtain variables $(T_n^{(N)}, n \geq 0)$. Then, let $S_n^{(N)} = B_{T_n^{(N)}}^{(N)}$. Then by Lemma 4.15, $(S_n^{(N)}, n \geq 0)$ is a random walk with same law as $(S_n, n \geq 0)$. We interpolate linearly between integers to obtain a continuous process $(S_t^{(N)}, 0 \leq t \leq 1)$ which thus has the distribution as $(S_t, 0 \leq t \leq 1)$. Finally, let

$$\tilde{S}_t^{(N)} = \frac{S_{Nt}^{(N)}}{\sqrt{\sigma^2 N}}, t \geq 0$$

and $\tilde{T}_n^{(N)} = N^{-1}T_n^{(N)}$. Finally, let $B'_t = B_{\sigma^2 t}/\sqrt{\sigma^2}$, which is also a Brownian motion. We are going to show that the supremum norm

$$\|B'_t - \tilde{S}_t^{(N)}\|_\infty \rightarrow_p 0$$

as $N \rightarrow \infty$, where \rightarrow_p denotes convergence in probability.

First recall what we have proved in (4.5), and note that this implies convergence in probability. Since $(T_n^{(N)}, n \geq 0)$ has the same distribution as $(T_n, n \geq 0)$ we infer from this that for every $\delta > 0$, letting $\delta' = \delta\sigma^2 > 0$, we have:

$$\mathbb{P} \left(N^{-1} \sup_{0 \leq n \leq N} |T_n^{(N)} - n\sigma^2| \geq \delta' \right) \xrightarrow{N \rightarrow \infty} 0.$$

Therefore dividing by σ^2 :

$$\mathbb{P} \left(\sup_{0 \leq n \leq N} |\tilde{T}_n^{(N)}/\sigma^2 - n/N| \geq \delta \right) \xrightarrow{N \rightarrow \infty} 0.$$

Now, note that if $t = n/N$, then

$$\tilde{S}_t^{(N)} = \frac{S_n^{(N)}}{\sqrt{N\sigma^2}} = \frac{B_{\tilde{T}_n^{(N)}}^{(N)}}{\sqrt{\sigma^2}} = B'_{\tilde{T}_n^{(N)}/\sigma^2}.$$

Thus, by continuity, if $t \in [n/N, (n+1)/N]$, there exists $u \in [T_n^{(N)}/\sigma^2, T_{n+1}^{(N)}/\sigma^2]$ such that $S_t^{(N)} = B'_u$. Therefore, for all $\varepsilon > 0$ and all $\delta > 0$, the event

$$\left\{ \sup_{0 \leq t \leq 1} |\tilde{S}_t^{(N)} - B'_t| > \varepsilon \right\} \subseteq K_\delta^N \cup L_{\delta, \varepsilon}^N,$$

where

$$K_\delta^N = \left\{ \sup_{0 \leq n \leq N} |\tilde{T}_n^{(N)}/\sigma^2 - n/N| > \delta \right\}$$

and

$$L_{\delta,\varepsilon}^N = \{\exists t \in [0, 1], \exists u \in [t - \delta, t + \delta + 1/N] : |B'_t - B'_u| > \varepsilon\}.$$

We already know that $P(K_\delta^N) \rightarrow 0$ as $N \rightarrow \infty$. For $L_{\delta,\varepsilon}^N$, since B' is a.s. uniformly continuous on $[0, 1]$, by taking δ small enough and then N large enough, we can make $P(L^N)$ as small as wanted. More precisely, if

$$L_{2\delta,\varepsilon} = \{\exists t \in [0, 1], \exists u \in [t - 2\delta, t + 2\delta] : |B'_t - B'_u| > \varepsilon\}.$$

then for $N \geq 1/\delta$, $L_{\delta,\varepsilon}^N \subseteq L_{2\delta,\varepsilon}$, and thus for all $\delta > 0$:

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\|\tilde{S}^{(N)} - B'\|_\infty > \varepsilon \right) \leq \mathbb{P}(L_{2\delta,\varepsilon})$$

However, as $\delta \rightarrow 0$, $P(L_{2\delta,\varepsilon}) \rightarrow 0$ by almost sure continuity of B' on $(0, 1)$ and the fact that these events are decreasing. Hence it must be that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left(\|\tilde{S}^{(N)} - B'\|_\infty > \varepsilon \right) = 0.$$

Therefore, $(\tilde{S}^{(N)}, 0 \leq t \leq 1)$ converges in probability for the uniform norm to $(B_t, 0 \leq t \leq 1)$, which entails convergence in distribution. This concludes the proof. \square

4.5 Brownian motion and the Dirichlet problem

Let D be a connected open subset of \mathbb{R}^d for some $d \geq 1$ (though the story is interesting only for $d \geq 2$). We will say that D is a *domain*. Let ∂D be the boundary of D . We denote by Δ the Laplacian on \mathbb{R}^d .

Definition 4.1 *Let $g : \partial D \rightarrow \mathbb{R}$ be a continuous function. A solution of the Dirichlet problem with boundary condition g on D is a function $u : \bar{D} \rightarrow \mathbb{R}$ of class $\mathcal{C}^2(D) \cap \mathcal{C}(\bar{D})$, such that*

$$\begin{cases} \Delta u = 0 & \text{on } D \\ u|_{\partial D} = g. \end{cases} \quad (4.6)$$

A solution of the Dirichlet problem is the mathematical counterpart of the following physical problem: given an object made of homogeneous material, such that the temperature $g(y)$ is imposed at point y of its boundary, the solution $u(x)$ of the Dirichlet problem gives the temperature at the point x in the object when equilibrium is attained.

As we will see, it is possible to give a probabilistic resolution of the Dirichlet problem with the help of Brownian motion. This is essentially due to Kakutani. We let \mathbb{E}_x be the law of the Brownian motion in \mathbb{R}^d started at x . In the remaining of the section, let $T = \inf\{t \geq 0 : B_t \notin D\}$ be the first exit time from D . It is a stopping time, as it is the first entrance time in the closed set D^c . We will assume that the domain D is such that $P(T < \infty) = 1$ to avoid complications. Hence B_T is a well-defined random variable.

We will need the following generalization of Itô's formula:

Lemma 4.17 *Let D be a domain (open and connected subset of \mathbb{R}^d) which is a proper subset. Let $f : D \rightarrow \mathbb{R}$ be a \mathcal{C}^2 function on D . Then if X is a semimartingale such that $X_0 \in D$ almost surely, and if $T = \inf\{t \geq 0 : X_t \notin D\}$ then we have:*

$$f(X_t) = \int_0^t \sum_{i=1}^d \frac{\partial f}{\partial x_i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s$$

almost surely for all $t < T$.

Proof. We may assume without loss of generality that $X^i \in \mathcal{M}_c^2$ for each $1 \leq i \leq d$. Let $n \geq 1$ and define $T_n = \inf\{t \geq 0 : d(X_t, D^c) \leq 1/n\}$. Then $T_n \leq T$ almost surely and T_n is nondecreasing, hence $L = \lim_{n \rightarrow \infty} T_n$ exists. We have $L \leq T$ by passing to the limit in $T_n \leq T$, and we also claim that $L \geq T$. Indeed, since the distance is a continuous function, $d(X_L, D^c) = 0$. Note that $d(X_L, D^c) = \inf_{y \in D^c} d(X_L, y) = \inf_{y \in D^c \cap \bar{B}(X_L, 1)} d(X_L, y)$. Since D is open, $D^c \cap \bar{B}(X_L, 1)$ is compact and thus this distance is attained. This means that $X_L \in D^c$ which implies $L \geq T$. Thus $L = T$.

Now, let us introduce a sequence of functions $(\varphi_m)_{m \geq 1}$ which are \mathcal{C}^∞ -approximations of the identity (such as the Gaussian kernel with variance $1/m$.) Consider the function

$$f_{n,m} = (f \mathbf{1}_{\{D_n\}}) \star \varphi_m$$

where D_n is the subdomain $D_n = \{x \in D : d(x, D^c) > 1/n\}$ and \star denotes the convolution of two functions, i.e., $f \star g(x) = \int_{\mathbb{R}^d} f(y)g(x-y)dy$. Since φ_m is \mathcal{C}^∞ , and since convolution is a regularizing operation, the function $f_{n,m}$ is \mathcal{C}^∞ for all n, m . Thus we can apply Itô's formula to $f_{n,m}$. Stopping at time T_n , we get:

$$f_{n,m}(X_{t \wedge T_n}) = \int_0^{t \wedge T_n} \sum_{i=1}^d \frac{\partial f_{n,m}}{\partial x_i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^{t \wedge T_n} \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s. \quad (4.7)$$

However, since f is \mathcal{C}^2 inside D , we have for all $x \in D_n$:

$$\frac{\partial f_{n,m}}{\partial x_i}(x) = \varphi_m \star \frac{\partial f}{\partial x_i}$$

and

$$\frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(x) = \varphi_m \star \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Since φ_m is an approximation of the identity, this means that as $m \rightarrow \infty$,

$$\frac{\partial f_{n,m}}{\partial x_i}(x) \rightarrow \frac{\partial f}{\partial x_i}(x), \quad \text{and} \quad \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(x) \rightarrow \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

pointwise in D . This implies that one can take the limit $m \rightarrow \infty$ in (4.7). Indeed, the second term

$$\int_0^{t \wedge T_n} \frac{\partial^2 f_{n,m}}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s \rightarrow \int_0^{t \wedge T_n} \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s$$

converges because of the Lebesgue convergence theorem since each $X_i \in \mathcal{M}_c^2$. To see that the first term also converges, apply the isometry property of the stochastic integral and the Lebesgue convergence theorem:

$$\left\| \left(\frac{\partial f_{n,m}}{\partial x_i} - \frac{\partial f}{\partial x_i} \right) (X) \cdot X^i \right\|_X^2 = \mathbb{E} \left(\int_0^{T_n} \left(\frac{\partial f_{n,m}}{\partial x_i} - \frac{\partial f}{\partial x_i} \right)^2 (X_s^i) d[X]_s \right) \rightarrow 0.$$

Thus we get the desired formula for all $t \leq T_n$ almost surely. Letting $n \rightarrow \infty$ finishes the proof. \square

In the sequel, $|\cdot|$ is the Euclidean norm on \mathbb{R}^d . The goal of this section is to prove the following result:

Theorem 4.18 *Suppose that $g \in \mathcal{C}(\partial D, \mathbb{R})$ is bounded, and assume that D satisfies a local exterior cone condition (l.e.c.c.), i.e. for every $y \in \partial D$, there exists a nonempty open convex cone with origin at y such that $C \cap B(y, r) \subseteq D^c$ for some $r > 0$. Then the function*

$$u : x \mapsto \mathbb{E}_x [g(B_T)]$$

is the unique bounded solution of the Dirichlet problem (4.6). In particular, if D is bounded and satisfies the l.e.c.c., then u is the unique solution of the Dirichlet problem.

We start with a uniqueness statement.

Proposition 4.19 *Let g be a bounded function in $\mathcal{C}(\partial D, \mathbb{R})$. Set*

$$u(x) = \mathbb{E}_x [g(B_T)].$$

If v is a bounded solution of the Dirichlet problem, then $v = u$.

In particular, we obtain uniqueness when D is bounded. Notice that we do not make any assumption on the regularity of D here besides the fact that $T < \infty$ a.s.

Proof. Let v be a bounded solution of the Dirichlet problem. Let $T_n = \inf\{t \geq 0 : d(X_t, D^c) < 1/n\}$. Since $\Delta v = 0$ inside D , we know by Lemma 4.17 that $M_t = v(B_{t \wedge T_n}) - v(B_0)$ is a local martingale started at 0 (here, $B_0 = x$ almost surely). Moreover, since v is bounded, M is a true martingale which is uniformly integrable. Applying the optional stopping theorem (2.36) at the stopping time T_n ,

$$\mathbb{E}(M_{T_n}) = \mathbb{E}_x(v(B_{T_n}) - v(x)) = \mathbb{E}_x(g(B_{T_n})) - v(x) = 0.$$

Since $T_n \rightarrow T$ almost surely as $n \rightarrow \infty$, and since v is continuous on $\mathcal{C}(\bar{D})$, we get:

$$v(x) = \mathbb{E}_x(g(B_T))$$

as claimed. \square

For every $x \in \mathbb{R}^d$ and $r > 0$, let $\sigma_{x,r}$ be the uniform probability measure on the sphere $\mathbb{S}_{x,r} = \{y \in \mathbb{R}^d : |y - x| = r\}$. It is the unique probability measure on $\mathbb{S}_{x,r}$ that is invariant under isometries of $\mathbb{S}_{x,r}$. We say that a locally bounded measurable function

$h : D \rightarrow \mathbb{R}$ is *harmonic* on D if for every $x \in D$ and every $r > 0$ such that the closed ball $\overline{B}(x, r)$ with center x and radius r is contained in D ,

$$h(x) = \int_{\mathbb{S}_{x,r}} h(y) \sigma_{x,r}(dy).$$

Proposition 4.20 *Let h be harmonic on a domain D . Then $h \in \mathcal{C}^\infty(D, \mathbb{R})$, and $\Delta h = 0$ on D .*

Proof. Let $x \in D$ and $\varepsilon > 0$ such that $\overline{B}(x, \varepsilon) \subseteq D$. Then let $\varphi \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ be non-negative with non-empty compact support in $[0, \varepsilon[$. We have, for $0 < r < \varepsilon$,

$$h(x) = \int_{\mathbb{S}(0,r)} h(x+y) \sigma_{0,r}(dy).$$

Multiplying by $\varphi(r)r^{d-1}$ and integrating over $r \in (0, \varepsilon)$ gives

$$ch(x) = \int_{B(0,\varepsilon)} \varphi(|z|) h(x+z) dz,$$

where $c > 0$ is some constant, where we have used the fact that

$$\int_{\mathbb{R}^d} f(x) dx = C \int_{\mathbb{R}_+} r^{d-1} dr \int_{\mathbb{S}(0,r)} f(ry) \sigma_{0,r}(dy)$$

for some $C > 0$. Therefore, $ch(x) = \int_{B(x,\varepsilon)} \varphi(|z-x|) h(z) dz = \int_{\mathbb{R}^d} \varphi(|z-x|) h(z) dz$ since φ is supported on $B(0, \varepsilon)$. By derivation under the \int sign, we easily get that h is \mathcal{C}^∞ . (Indeed, we may assume that $r \mapsto \varphi(r^{1/2})$ is \mathcal{C}^∞). Another way to say this is to say that $ch(x) = \tilde{\varphi} \star h$ where $\tilde{\varphi}(z) = \varphi(|z|)$. If $r \mapsto \varphi(r^{1/2})$ is \mathcal{C}^∞ , then $\tilde{\varphi} \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ and thus, the convolution being a regularizing operation, this implies $\tilde{\varphi} \star h \in \mathcal{C}^\infty(D, \mathbb{R})$. Thus $h \in \mathcal{C}^\infty(D, \mathbb{R})$.

Next, by translation we may suppose that $0 \in D$ and show only that $\Delta h(0) = 0$. we may apply Taylor's formula to h , obtaining, as $x \rightarrow 0$,

$$h(x) = h(0) + \langle \nabla h(0), x \rangle + \sum_{i=1}^d x_i^2 \frac{\partial^2 h}{\partial x_i^2}(0) + \sum_{i \neq j} x_i x_j \frac{\partial^2 h}{\partial x_i \partial x_j}(0) + o(|x|^2).$$

Now, integration over $\mathbb{S}_{0,r}$ for r small enough yields

$$\int_{\mathbb{S}_{x,r}} h(x) \sigma_{0,r}(dx) = h(0) + C_r \Delta h(0) + o(r^2),$$

where $C_r = \int_{\mathbb{S}_{0,r}} x_1^2 \sigma_{0,r}(dx)$, as the reader may check that all the other integrals up to the second order are 0, by symmetry. Now, it is easy to see that there exists $c > 0$ such that $C_r \geq cr^2$ for all $0 \leq r \leq 1$. Since the left-hand side is $h(0)$ and the error term on the right-hand side is $o(r^2) = o(C_r)$, it follows that $\Delta h(0) = 0$. \square

Therefore, harmonic functions are solutions of certain Dirichlet problems.

Proposition 4.21 *Let g be a bounded measurable function on ∂D , and let $T = \inf\{t \geq 0 : B_t \notin D\}$. Then the function $h : x \in D \mapsto E_x[g(B_T)]$ is harmonic on D , and hence $\Delta h = 0$ on D .*

Proof. For every Borel subsets A_1, \dots, A_k of \mathbb{R}^d and times $t_1 < \dots < t_k$, the map

$$x \mapsto P_x(B_{t_1} \in A_1, \dots, B_{t_n} \in A_n)$$

is measurable by Fubini's theorem, once one has written the explicit formula for this probability. Therefore, by the monotone class theorem, $x \mapsto E_x[F]$ is measurable for every integrable random variable F , which is measurable with respect to the product σ -algebra on $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$. Moreover, h is bounded by assumption.

Now, let $S = \inf\{t \geq 0 : |B_t - x| \geq r\}$ the first exit time of B from the ball of center x and radius r . Then by (ii), Proposition 1.12, $S < \infty$ a.s. By the strong Markov property, $\tilde{B} = (B_{S+t}, t \geq 0)$ is an (\mathcal{F}_{S+t}) Brownian motion started at B_S . Moreover, the first hitting time of ∂D for \tilde{B} is $\tilde{T} = T - S$. Moreover, $\tilde{B}_{\tilde{T}} = B_T$, so that

$$E_x[g(B_T)] = E_x[g(\tilde{B}_{\tilde{T}})] = \int_{\mathbb{R}^d} P_x(B_S \in dy) E_y[g(B_T) \mathbf{1}_{\{T < \infty\}}],$$

and we recognize $\int P_x(B_S \in dy)h(y)$ in the last expression.

Since B starts from x under P_x , the rotation invariance of Brownian motion shows that $B_S - x$ has a distribution on the sphere of center 0 and radius r which is invariant under the orthogonal group, so we conclude that the distribution of B_S is the uniform measure on the sphere of center x and radius r , and therefore that h is harmonic on D . \square

It remains to understand whether the function u of Theorem 4.18 is actually a solution of the Dirichlet problem. Indeed, is not the case in general that $u(x)$ has limit $g(y)$ as $x \in D, x \rightarrow y$, and the reason is that some points of ∂D may be 'invisible' to Brownian motion. The reader can convince himself, for example, that if $D = B(0, 1) \setminus \{0\}$ is the open ball of \mathbb{R}^2 with center 0 and radius 1, whose origin has been removed, and if $g = \mathbf{1}_{\{0\}}$, then no solution of the Dirichlet problem with boundary constraint g exists. The probabilistic reason for that is that Brownian motion does not see the boundary point 0. This is the reason why we have to make regularity assumptions on D in the following theorem.

Proof of Theorem 4.18.

It remains to prove that under the l.e.c.c., u is continuous on \overline{D} , i.e. $u(x)$ converges to $g(y)$ as $x \in D$ converges to $y \in \partial D$. In order to do that, we need a preliminary lemma. Recall that T is the first exit time of D for the Brownian path.

Lemma 4.22 *Let D be a domain satisfying the l.e.c.c., and let $y \in \partial D$. Then for every $\eta > 0$, $P_x(T < \eta) \rightarrow 1$ as $x \in D \rightarrow y$.*

Proof. Let $C_y = y + C$ be a nonempty open convex cone with origin at y such that $C_y \subseteq D^c$ (we leave as an exercise the case when only a neighborhood of this cone around y is contained in D^c). Then it is an elementary geometrical fact that there exists a nonempty open convex cone C' with origin at 0 such that for every $\delta > 0$ small enough,

we can find an $\varepsilon = \varepsilon(\delta) > 0$ such that if $C'_x = x + C'$, then $(C'_x \setminus \overline{B}(x, \delta)) \subseteq C_y$ for all $x \in B(y, \varepsilon)$.

[Here is a justification. Assume without loss of generality that $y = 0$ to simplify, and fix $\delta > 0$. Let O be an open set in the unit sphere such that $C = \{\lambda z, z \in O, \lambda > 0\}$. There exists $\alpha > 0$ and another open set O' in the unit sphere such that $O' \subseteq O$ and if $z \in \mathbb{S}_{0,1}$ with $d(z, O') \leq \alpha$ then $z \in O$. (For instance consider the intersection of the sphere with two concentric open balls centered at some $z_0 \in O$, and take O' the smaller of the two balls intersected with \mathbb{S}). Now, choose $\varepsilon = \delta\alpha/4$. Let $x \in B(0, \varepsilon)$, and let us show that $(x + C') \setminus \overline{B}(x, \delta) \subseteq C$ where C' is the cone generated by O' originating at 0 (which does not depend on δ). For $y \in O'$, let $z = (x + \lambda y)/r$ where $r = \|x + \lambda y\|$, then $z \in \mathbb{S}_{0,1}$. Moreover, note that by the triangular inequality $|r - \lambda| \leq \varepsilon$. Thus if $r \geq \delta/2$,

$$\begin{aligned} \|y - z\| &= \left\| y - \frac{1}{r}(\lambda y + x) \right\| \\ &= \frac{1}{r} \|(r - \lambda)y - x\| \\ &\leq \frac{2}{\delta}((r - \lambda) + \varepsilon) \\ &\leq \frac{4\varepsilon}{\delta} \leq \alpha \end{aligned}$$

by definition of ε . Hence $z \in O$ and hence $x + \lambda y = rz \in C$. Now, if ε is further chosen such that $\varepsilon \leq \delta/2$, then for all $x \in B(0, \varepsilon)$ and for all $u \in (x + C') \setminus \overline{B}(x, \delta)$, $r = \|u\| \leq \delta/2$ automatically by the triangular inequality, and thus the previous conclusion $u \in C$ holds. We have shown that $(x + C') \setminus \overline{B}(x, \delta) \subseteq C$ as desired.]

Now by (iii) in Proposition 1.12, if

$$H_{C'}^\delta = \inf\{t > 0 : B_t \in C' \setminus \overline{B}(0, \delta)\},$$

then

$$\mathbb{P}_0(H_{C'}^\delta < \eta) \rightarrow \mathbb{P}_0(H_{C'} < \eta) = 1 \text{ as } \delta \downarrow 0.$$

Therefore, for all $\alpha > 0$ there exists $\delta > 0$ such that $\mathbb{P}(H_{C'}^\delta \leq \eta) \geq 1 - \alpha$. Choosing $\varepsilon = \varepsilon(\delta)$ associated with this δ , we find that for every $x \in B(y, \varepsilon)$, we have by translation invariance, and letting T_K be the hitting time of a set K ,

$$\begin{aligned} \mathbb{P}_x(T > \eta) &\leq \mathbb{P}_x(T_{C'_x \setminus \overline{B}(x, \delta)} > \eta) \\ &= \mathbb{P}_0(H_{C'}^\delta > \eta) \\ &\leq \alpha \quad (\text{by our choice of } \delta). \end{aligned}$$

This means that $\mathbb{P}_x(T > \eta) \rightarrow 0$ as $x \rightarrow y$, which proves the lemma. \square

We can now finish the proof of Theorem 4.18. Let $y \in \partial D$. We want to estimate the quantity $\mathbb{E}_x[g(B_T)] - g(y)$ for some $x \in D$. For $\eta, \delta > 0$, let

$$A_{\eta, \delta} = \left\{ \sup_{0 \leq t \leq \eta} |B_t - x| \geq \delta/2 \right\}.$$

This event decreases to \emptyset as $\eta \downarrow 0$ because B has continuous paths. Now, for any $\delta, \eta > 0$,

$$\begin{aligned} \mathbb{E}_x[|g(B_T) - g(y)|] &= \mathbb{E}_x[|g(B_T) - g(y)|; \{T \leq \eta\} \cap A_{\delta, \eta}^c] \\ &\quad + \mathbb{E}_x[|g(B_T) - g(y)|; \{T \leq \eta\} \cap A_{\delta, \eta}] \\ &\quad + \mathbb{E}_x[|g(B_T) - g(y)|; \{T \geq \eta\}] \end{aligned}$$

Fix $\varepsilon > 0$. We are going to show that each of these three quantities can be made $< \varepsilon/3$ for x close enough to y . Since g is continuous at y , for some $\delta > 0$, $|y - z| < \delta$ with $y, z \in \partial D$ implies $|g(y) - g(z)| < \varepsilon/3$. Moreover, on the event $\{T \leq \eta\} \cap A_{\delta, \eta}^c$, we know that $|B_T - x| < \delta/2$, and thus $|B_T - y| \leq \delta$ as soon as $|x - y| \leq \delta/2$. Therefore, for every $\eta > 0$, the first quantity is less than $\varepsilon/3$ for $x \in \overline{B}(y, \delta/2)$.

Next, if M is an upper bound for $|g|$, the second quantity is bounded by $2MP(A_{\delta, \eta})$. Hence, by now choosing η small enough, this is $< \varepsilon/3$.

Finally, with δ, η fixed as above, the third quantity is bounded by $2MP_x(T \geq \eta)$. By the previous lemma, this is $< \varepsilon/3$ as soon as $x \in B(y, \alpha) \cap D$ for some $\alpha > 0$. Therefore, for any $x \in B(y, \alpha \wedge \delta/2) \cap D$, $|u(x) - g(y)| < \varepsilon$. This entails the result. \square

Corollary 4.23 *A function $u : D \rightarrow \mathbb{R}$ is harmonic in D if and only if it is in $\mathcal{C}^2(D, \mathbb{R})$, and satisfies $\Delta u = 0$.*

Proof. Let u be of class $\mathcal{C}^2(D, \mathbb{R})$ and be of zero Laplacian, and let $x \in D$. Let ε be such that $B(x, \varepsilon) \subseteq D$, and notice that $u|_{\overline{B}(x, \varepsilon)}$ is a solution of the Dirichlet problem on $\overline{B}(x, \varepsilon)$ with boundary values $u|_{\partial B(x, \varepsilon)}$. Then $B(x, \varepsilon)$ satisfies the l.e.c.c., so that $u|_{B(x, \varepsilon)}$ is the unique such solution, which is also given by the harmonic function of Theorem 4.18. Therefore, u is harmonic on D . \square

4.6 Girsanov's theorem

Given a local martingale M , one may define an exponential martingale associated with it, which will play a crucial role in what follows.

Definition 4.2 Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Set $Z_t = \exp(M_t - \frac{1}{2}[M]_t)$. By Itô's formula

$$dZ_t = Z_t(dM_t - \frac{1}{2}d[M]_t) + \frac{1}{2}Z_t d[M]_t = Z_t dM_t, \quad (4.8)$$

and so $Z \in \mathcal{M}_{c,loc}$. We call Z the *exponential martingale* of M (sometimes also called the *stochastic exponential* of M) and write $Z = \mathcal{E}(M)$.

We start by the following inequality which will be useful in the proof of Girsanov's theorem, but is also interesting in its own right.

Proposition 4.24 Exponential martingale inequality

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Then for all $x > 0, u > 0$,

$$\mathbb{P}\left(\sup_{t \geq 0} M_t > x, [M]_\infty \leq u\right) \leq e^{-x^2/(2u)}. \quad (4.9)$$

Proof.

Fix $x > 0$ and set $T = \inf\{t \geq 0 : M_t \geq x\}$. Fix $\theta \in \mathbb{R}$ and set

$$Z_t = \exp\left(\theta M_t^T - \frac{1}{2}\theta^2[M]_t^T\right). \quad (4.10)$$

Then $Z \in \mathcal{M}_{c,loc}$ and $|Z| \leq e^{\theta x}$. Hence, $Z \in \mathcal{M}_c^2$ and, by OST, $\mathbb{E}(Z_\infty) = \mathbb{E}(Z_0) = 1$. For $u > 0$ we get by Markov's inequality

$$\mathbb{P}\left(\sup_{t \geq 0} M_t > x, [M]_\infty \leq u\right) \leq \mathbb{P}(Z_\infty \geq e^{\theta x - \frac{1}{2}\theta^2 u}) \leq e^{-\theta x + \frac{1}{2}\theta^2 u}. \quad (4.11)$$

Optimizing in θ gives $\theta = x/u$ and the result follows. (It is also possible to use the Dubins-Schwarz theorem if we allow a multiplicative constant on the right-hand side of the inequality). \square

Proposition 4.25 *Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$ and suppose that $[M]$ is a.s. uniformly bounded. Then $\mathcal{E}(M)$ is a UI martingale.*

Proof. Let C be such that $[M]_\infty \leq C$ a.s. By the exponential martingale inequality, for all $x > 0$

$$\mathbb{P}\left(\sup_{t \geq 0} M_t \geq x\right) = \mathbb{P}\left(\sup_{t \geq 0} M_t \geq x, [M]_\infty \leq C\right) \leq e^{-x^2/(2C)}. \quad (4.12)$$

Now, $\sup_{t \geq 0} \mathcal{E}(M)_t \leq \exp\left(\sup_{t \geq 0} M_t\right)$ and

$$\begin{aligned} \mathbb{E}\left(\exp\left(\sup_{t \geq 0} M_t\right)\right) &= \int_0^\infty \mathbb{P}\left(\sup_{t \geq 0} M_t \geq \log \lambda\right) d\lambda \\ &\leq 1 + \int_1^\infty e^{-(\log \lambda)^2/(2C)} d\lambda < \infty. \end{aligned} \quad (4.13)$$

Hence, $\mathcal{E}(M)$ is UI and, by Proposition 2.9, $\mathcal{E}(M)$ is a martingale. \square

Girsanov's theorem is a result which relates absolute continuous changes of the underlying probability measure \mathbb{P} to changes in the drift of the process. In order to see where this deep result comes from on a simple example where one can compute everything by hand, consider the following. Let $\sigma > 0$ and $b \neq 0$, and let $X_t = \sigma B_t + bt$. Then we claim that the law of X is absolutely continuous with respect to the law of Brownian motion $Y_t = \sigma B_t$ with speed σ (but without drift), so long as we restrict ourselves to events of \mathbb{F}_t for some fixed $t > 0$. Indeed, if $n \geq 1$ and $0 = t_0 < t_1 < \dots < t_n = t$ and $x_0 = 0, x_1, \dots, x_n \in \mathbb{R}$, then we have:

$$\mathbb{P}(X_{t_1} = x_1, \dots, X_{t_n} = x_n) = C \exp\left(-\sum_{i=0}^{n-1} \frac{(x_{i+1} - x_i - b(t_{i+1} - t_i))^2}{2\sigma^2(t_{i+1} - t_i)}\right) \prod_{i=1}^n dx_i$$

where C is a factor depending on t_1, \dots, t_n and σ , whose value is of no interest to us. It follows that

$$\frac{\mathbb{P}(X_{t_1} = x_1, \dots, X_{t_n} = x_n)}{\mathbb{P}(Y_{t_1} = x_1, \dots, Y_{t_n} = x_n)} = e^{-Z}$$

where

$$\begin{aligned}
Z &= \sum_{i=0}^{n-1} \frac{(x_{i+1} - x_i - b(t_{i+1} - t_i))^2}{2\sigma^2(t_{i+1} - t_i)} - \frac{(x_{i+1} - x_i)^2}{2\sigma^2(t_{i+1} - t_i)} \\
&= \sum_{i=0}^{n-1} -\frac{b}{\sigma^2}(x_{i+1} - x_i) + \frac{1}{2\sigma^2}b^2(t_{i+1} - t_i) \\
&\longrightarrow -\int_0^t \sigma^{-2}b dY_s + \frac{1}{2} \int_0^t b^2 \sigma^{-2} ds \quad \text{as } n \rightarrow \infty.
\end{aligned}$$

(We have written the last bit as a convergence although there is an exact equality, since it should be clear from this calculation that when σ and b depend on the position x (which is precisely what defines the SDE's developed in the next chapter), then a similar convergence holds.) Thus if \mathbb{Q} is the law of X , and \mathbb{P} the law of Y ,

$$\begin{aligned}
\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} &= \exp \left(\int_0^t \frac{b}{\sigma} dB_s - \int_0^t \frac{b^2}{2\sigma^2} ds \right) \\
&= \mathcal{E}(\sigma^{-1}Y)_t
\end{aligned}$$

So we have written the density of X with respect to Y as an exponential martingale.

The point of view of Girsanov's theorem is a slightly different perspective, essentially the other side of the same coin. We will consider changes of measures given by a suitable exponential martingale, and observe the effect on the drift. It is of fundamental importance in mathematical finance (in the context of "risk neutral measures"). Remember that for two probability measures $\mathbb{P}_1, \mathbb{P}_2$ on a measurable space (Ω, \mathcal{F}) , \mathbb{P}_1 is absolutely continuous with respect to \mathbb{P}_2 , $\mathbb{P}_1 \ll \mathbb{P}_2$, if

$$\mathbb{P}_2(A) = 0 \quad \Rightarrow \quad \mathbb{P}_1(A) = 0 \quad \text{for all } A \in \mathcal{F}. \quad (4.14)$$

In this case, by the *Radon-Nikodym theorem*, there exists a density $f : \Omega \rightarrow [0, \infty)$ which is \mathcal{F} -measurable and \mathbb{P}_2 unique almost surely (and hence \mathbb{P}_1 unique almost surely as well), such that $\mathbb{P}_1 = f \mathbb{P}_2$. That is, for all $A \in \mathcal{F}$,

$$\mathbb{P}_1(A) = \int_{\Omega} f(\omega) \mathbf{1}_{\{A\}}(\omega) d\mathbb{P}_2(\omega).$$

f is also called the Radon-Nikodym derivative, and we denote:

$$\left. \frac{d\mathbb{P}_1}{d\mathbb{P}_2} \right|_{\mathcal{F}} = f$$

(Note that in general, the density f depends on the σ -field \mathcal{F}).

Theorem 4.26 Girsanov's theorem

Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Suppose that $Z = \mathcal{E}(M)$ is a UI martingale. We can define a new probability measure $\tilde{\mathbb{P}} \ll \mathbb{P}$ on (Ω, \mathcal{F}) by

$$\tilde{\mathbb{P}}(A) = \mathbb{E}(Z_{\infty} \mathbf{1}_A), \quad A \in \mathcal{F}. \quad (4.15)$$

Then for every $X \in \mathcal{M}_{c,loc}(\mathbb{P})$, $X - [X, M] \in \mathcal{M}_{c,loc}(\tilde{\mathbb{P}})$. Moreover the quadratic variation of X under \mathbb{P} and of $X - [X, M]$ under $\tilde{\mathbb{P}}$ are identical \mathbb{P} and $\tilde{\mathbb{P}}$ almost surely.

Proof. Since Z is UI, the limit $Z_\infty = \lim_{t \rightarrow \infty} Z_t$ exists \mathbb{P} -almost surely, $Z_\infty \geq 0$ and $\mathbb{E}(Z_\infty) = \mathbb{E}(Z_0) = 1$. Thus $\tilde{\mathbb{P}}(\Omega) = 1$, $\tilde{\mathbb{P}}(\emptyset) = 0$ and countable additivity follows by linearity of expectation and the monotone convergence theorem. $\tilde{\mathbb{P}}(A) = \int_A Z_\infty d\mathbb{P} = 0$ if $\mathbb{P}(A) = 0$ so $\tilde{\mathbb{P}} \ll \mathbb{P}$.

Let $X \in \mathcal{M}_{c,loc}$ and set

$$T_n = \inf \{t \geq 0 : |X_t - [X, M]_t| \geq n\} . \quad (4.16)$$

Since $X - [X, M]$ is continuous, $\mathbb{P}(T_n \nearrow \infty) = 1$ which implies $\tilde{\mathbb{P}}(T_n \nearrow \infty) = 1$. So to show that $Y = X - [X, M] \in \mathcal{M}_{c,loc}(\tilde{\mathbb{P}})$, it suffices to show that

$$Y^{T_n} = X^{T_n} - [X^{T_n}, M] \in \mathcal{M}_c(\tilde{\mathbb{P}}) \quad \text{for all } n \in \mathbb{N} . \quad (4.17)$$

Replacing X by X^{T_n} , we reduce to the case where Y is uniformly bounded. By the integration by parts formula and the Kunita-Watanabe identity,

$$\begin{aligned} d(Z_t Y_t) &= Y_t dZ_t + Z_t dY_t + d[Z_t, Y_t] \\ &= Y_t Z_t dM_t + Z_t (dX_t - d[X_t, M_t]) + Z_t d[M_t, X_t] \\ &= Y_t Z_t dM_t + Z_t dX_t \end{aligned} \quad (4.18)$$

and so $ZY \in \mathcal{M}_{c,loc}(\mathbb{P})$. Also $\{Z_T : T \text{ is a stopping time}\}$ is UI. Since Y is bounded, $\{Z_T Y_T : T \text{ is a stopping time}\}$ is UI. Hence, $ZY \in \mathcal{M}_c(\mathbb{P})$. But then for $s \leq t$,

$$\tilde{\mathbb{E}}(Y_t - Y_s | \mathcal{F}_s) = \mathbb{E}(Z_\infty(Y_t - Y_s) | \mathcal{F}_s) \quad (4.19)$$

$$= \mathbb{E}(Z_t Y_t - Z_s Y_s | \mathcal{F}_s) = 0 \quad (4.20)$$

since $ZY \in \mathcal{M}_c(\mathbb{P})$. Therefore, $Y \in \mathcal{M}_c(\tilde{\mathbb{P}})$ as required. The fact that the quadratic variation $[Y]$ is the same under $\tilde{\mathbb{P}}$ as it comes from the discrete approximation under \mathbb{P} :

$$[Y]_t = [X]_t = \lim_{n \rightarrow \infty} \sum_{k=0}^{\lfloor 2^n t \rfloor - 1} (X_{(k+1)2^{-n}} - X_{k2^{-n}})^2$$

\mathbb{P} -u.c.p. Thus there exists a subsequence n_k for which the convergence holds \mathbb{P} -almost surely uniformly on compacts. Since $\tilde{\mathbb{P}}$ is absolutely continuous with respect to \mathbb{P} this limit also holds $\tilde{\mathbb{P}}$ almost surely for this particular subsequence. Since the whole sequence converges in probability to $[Y]$ in the $\tilde{\mathbb{P}}$ -u.c.p. sense (by general theory, since $Y \in \mathcal{M}_{c,loc}(\tilde{\mathbb{P}})$), this uniquely identifies the limit, and hence the quadratic variation $[Y]$ under $\tilde{\mathbb{P}}$ has the same value as under \mathbb{P} . \square

Corollary 4.27 *Let B be a standard Brownian motion under \mathbb{P} and $M \in \mathcal{M}_{c,loc}$ such that $M_0 = 0$. Suppose $Z = \mathcal{E}(M)$ is a UI martingale and $\tilde{\mathbb{P}}(A) = \mathbb{E}(Z_\infty \mathbf{1}_A)$ for all $A \in \mathcal{F}$. Then $\tilde{B} := B - [B, M]$ is a $\tilde{\mathbb{P}}$ -Brownian motion.*

Proof. Since $\tilde{B} \in \mathcal{M}_{c,loc}(\tilde{\mathbb{P}})$ by Theorem 4.26 and has $[\tilde{B}]_t = [B]_t = t$, by Lévy's characterization, it is a Brownian motion. \square

Remark. This corollary should be read backward: if X is a Brownian motion, then changing the measure by the exponential martingale $\mathcal{E}(M)$, $X = \tilde{X} + [X, M]$ where \tilde{X} is a Brownian motion under the new measure. So the old process (which was just Brownian motion) becomes under the new measure a Brownian motion plus a “drift” term given by the covariation $[X, M]$.

Let $(W, \mathcal{W}, \mathbb{W})$ be the Wiener space. (Recall that $W = C([0, \infty), \mathbb{R})$, $\mathcal{W} = \sigma(X_t : t \geq 0)$ where $X_t : W \rightarrow \mathbb{R}$ with $X_t(w) = w(t)$. The Wiener measure \mathbb{W} is the unique probability measure on (W, \mathcal{W}) such that $(X_t)_{t \geq 0}$ is a Brownian motion started from 0.)

Definition 4.3 Define the *Cameron-Martin space*

$$H = \left\{ h \in W : h(t) = \int_0^t \phi(s) ds \text{ for some } \phi \in L^2([0, \infty)) \right\}. \quad (4.21)$$

For $h \in H$, write $\dot{h} = \phi$ the derivative of h .

Theorem 4.28 Girsanov, Cameron-Martin theorem

Fix $h \in H$ and set \mathbb{W}^h to be the law on (W, \mathcal{W}) of $(B_t + h(t), t \geq 0)$ where B_t is a Brownian motion: that is, for all $A \in \mathcal{W}$,

$$\mathbb{W}^h(A) = \mathbb{W}(\{w \in W : w + h \in A\}). \quad (4.22)$$

Then \mathbb{W}^h is a probability measure on (W, \mathcal{W}) and $\mathbb{W}^h \ll \mathbb{W}$ with Radon-Nikodym density

$$\frac{d\mathbb{W}^h}{d\mathbb{W}} \Big|_{\mathcal{W}} = \exp \left(\int_0^\infty \dot{h}(s) dX_s - \frac{1}{2} \int_0^\infty \dot{h}(s)^2 ds \right). \quad (4.23)$$

Remark. So if we take a Brownian motion and shift it by a deterministic function $h \in H$ then the resulting process has a law which is absolutely continuous with respect to that of the original Brownian motion.

Proof. Set $\mathcal{W}_t = \sigma(X_s, s \leq t)$ and $M_t = \int_0^t \phi(s) dX_s$. Then $M \in \mathcal{M}_c^2(W, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, \mathbb{W})$ and

$$[M]_\infty = \int_0^\infty \phi(s)^2 ds =: C < \infty. \quad (4.24)$$

By Proposition 4.25, $\mathcal{E}(M)$ is a UI martingale, so we can define a new probability measure $\mathcal{P} \ll \mathbb{W}$ on (W, \mathcal{W}) by

$$\frac{d\mathcal{P}}{d\mathbb{W}}(\omega) = \exp \left(M_\infty(\omega) - \frac{1}{2} [M]_\infty(\omega) \right) = \exp \left(\int_0^\infty \phi(s) dX_s(\omega) - \frac{1}{2} \int_0^\infty \phi(s)^2 ds \right) \quad (4.25)$$

and $\tilde{X} = X - [X, M] \in \mathcal{M}_{c,loc}(\mathcal{P})$ by Girsanov’s theorem. Since X is a \mathbb{W} -Brownian motion, by Corollary 4.27, \tilde{X} is a \mathcal{P} -Brownian motion. But by the Kunita-Watanabe identity,

$$\begin{aligned} [X, M]_t &= [X, \phi \cdot X]_t \\ &= \phi \cdot [X, X]_t \\ &= \int_0^t \phi(s) ds = h(t) \end{aligned}$$

hence we get that $\tilde{X}(\omega) = X(\omega) - h = \omega - h$. Hence, under \mathcal{P} , $X = \tilde{X} + h$, where \tilde{X} is a \mathcal{P} -Brownian motion. Therefore, $\mathbb{W}^h = \mathcal{P}$ on $\mathcal{F}_\infty = \mathcal{W}$:

$$\begin{aligned} \mathcal{P}(A) &= \mathcal{P}(\{\omega : \omega \in A\}) = \mathcal{P}(\{\omega : \tilde{X}(\omega) + h \in A\}) \\ &= \mathbb{W}(\{\omega : \omega + h \in A\}) = \mathbb{W}^h(A) \end{aligned}$$

as required. □.

One of the most important applications of Girsanov's theorem is to the study of Brownian motion with constant drift. Indeed, applying the previous result with $\phi(s) = \mathbf{1}_{\{s \leq t\}}$ gives us the following corollary (check it!)

Corollary 4.29 *Let $\gamma \neq 0$ and let \mathbb{W}^γ be the law of $(X_t + \gamma t, t \geq 0)$ under \mathbb{W} . Then for all $t > 0$, and for any $A \in \mathcal{F}_t$,*

$$\mathbb{W}^\gamma(A) = \mathbb{E}_{\mathbb{W}} \left(\mathbf{1}_{\{A\}} \exp\left(\gamma X_t - \frac{1}{2}\gamma^2 t\right) \right). \quad (4.26)$$

This allows us to compute functionals of Brownian motion with drift in terms of Brownian motion without drift – a very powerful technique. You will see some applications of this result in Example Sheet 3.

5 Stochastic Differential Equations

Suppose we have a differential equation, say $\frac{dx(t)}{dt} = b(x(t))$, or, in integral form,

$$x(t) = x(0) + \int_0^t b(x(s)) ds, \quad (5.1)$$

which describes a system evolving in time, be it the growth of a population, the trajectory of a moving object or the price of an asset. Taking into account random perturbations, it may be more realistic to add a noise term:

$$X_t = X_0 + \int_0^t b(X_s) ds + \sigma B_t, \quad (5.2)$$

where B is a Brownian motion and σ is a constant controlling the intensity of the noise. It may be that this intensity depends on the state of the system, in which case we have to consider an equation of the form

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s, \quad (5.3)$$

where the last term is, of course, an Itô integral. (5.3) is a *stochastic differential equation* and may also be written

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t. \quad (5.4)$$

5.1 General definitions

Let B be a Brownian motion in \mathbb{R}^m with $m \geq 1$. Let $d \geq 1$ and suppose

$$\sigma(x) = (\sigma_{ij}(x))_{\substack{1 \leq i \leq d \\ 1 \leq j \leq m}} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$$

and

$$b(x) = (b_i(x))_{1 \leq i \leq d} : \mathbb{R}^d \rightarrow \mathbb{R}^d$$

are given Borel functions, bounded on compact sets. Consider the equation in \mathbb{R}^d :

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt, \quad (5.5)$$

which may be written componentwise as

$$dX_t^i = \sum_{j=1}^m \sigma_{ij}(X_t) dB_t^j + b_i(X_t) dt, \quad 1 \leq i \leq d. \quad (5.6)$$

This general SDE will be called $E(\sigma, b)$. A *solution* to $E(\sigma, b)$ in (5.5) consists of

- a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the usual conditions;
- an $(\mathcal{F}_t)_{t \geq 0}$ -Brownian motion $B = (B^1, \dots, B^m)$ taking values in \mathbb{R}^m ;

- an $(\mathcal{F}_t)_{t \geq 0}$ -adapted continuous process $X = (X^1, \dots, X^d) \in \mathbb{R}^d$ such that

$$X_t = X_0 + \int_0^t \sigma(X_s) dB_s + \int_0^t b(X_s) ds . \quad (5.7)$$

When, in addition, $X_0 = x \in \mathbb{R}^d$, we say that X is a *solution started from x* .

There are several different notions of existence and uniqueness for SDE's, and we need to carefully distinguish between the various modes in which solutions can exist and be unique.

Definition 5.1 Let $E(\sigma, b)$ be the SDE in (5.5).

- We say that $E(\sigma, b)$ has a *solution* if for all $x \in \mathbb{R}^d$, there exists a solution to the SDE started from x .
- There is *uniqueness in law* if all solutions to $E(\sigma, b)$ started from x have the same distribution.
- There is *pathwise uniqueness* if, when we fix $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and B then any two solutions X and X' satisfying $X_0 = X'_0$ *a.s.* are indistinguishable from one another.
- We say that a solution X of $E(\sigma, b)$ started from x is a *strong solution* if X is adapted to the natural filtration of B .

Remark. In general, $\sigma(B_s, s \leq t) \subseteq \mathcal{F}_t$ and a solution might not be measurable with respect to the Brownian motion B . A strong solution depends only on $x \in \mathbb{R}^d$ and the Brownian motion B , and is moreover non-anticipative: if the path of B is known up to time t , then so is the path of X up to time t . We will term *weak* any solution that is not strong.

Example. It is possible to have existence of a weak solution and uniqueness in law without pathwise uniqueness. Suppose β is a Brownian motion in \mathbb{R} with $\beta_0 = x$. Set

$$B_t = \int_0^t \text{sgn}(\beta_s) d\beta_s \quad \text{where} \quad \text{sgn}(x) = \begin{cases} -1 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases} . \quad (5.8)$$

In problem 3 on example sheet 2, we showed that $(\text{sgn}(\beta_t))_{t \geq 0}$ is previsible, so that the Itô integral is well defined and $B \in \mathcal{M}_{c,loc}$. By Lévy's characterization, B is a Brownian motion started from 0, since $[B]_t = [\beta]_t = t$. It is also true that

$$\beta_t = x + \int_0^t \text{sgn}(\beta_s) dB_s . \quad (5.9)$$

(Indeed, by definition $dB_s = \text{sgn}(\beta_s) d\beta_s$ so multiplying by $\text{sgn}(\beta_s)$ yields $\text{sgn}(\beta_s) dB_s = d\beta_s$.) Hence, β is a solution to the SDE

$$dX_t = \text{sgn}(X_t) dB_t , \quad X_0 = x . \quad (5.10)$$

Thus (5.10) has a weak solution. Applying Lévy's characterization again, it is clear that any solution must be a Brownian motion and so there is uniqueness in law. On the

other hand, pathwise uniqueness does not hold: Suppose that β is a solution to (5.10) with $\beta_0 = 0$. Then both β and $-\beta$ are solutions to (5.10) started from 0. Indeed, we may write

$$\begin{aligned} -\beta_t &= -\int_0^t \operatorname{sgn}(\beta_s) dB_s \\ &= \int_0^t \operatorname{sgn}(-\beta_s) dB_s + 2 \int_0^t \mathbf{1}_{\{\beta_s=0\}} dB_s. \end{aligned}$$

The second term is a local martingale since it is an integral with respect to B . The quadratic variation of this local martingale is $4 \int_0^t \mathbf{1}_{\{\beta_s=0\}} ds$ which is 0 almost surely by Fubini's theorem (since β must be a Brownian motion by Lévy's characterization). Hence this local martingale is indistinguishable from 0, and $-\beta$ is a solution to (5.10). It also turns out that β is not a strong solution to (5.10).

The following theorem (whose proof is omitted) shows however that there is no counterexample in the opposite direction.

Theorem 5.1 (Yamada-Watanabe) *Let σ, b be measurable functions. If pathwise uniqueness holds for $E(\sigma, b)$ and there exist solutions, then there is also uniqueness in law. In this case, for every filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and every \mathcal{F}_t -Brownian motion B , and for every $x \in \mathbb{R}^d$, there exists a strong solution X to $E(\sigma, b)$.*

In particular pathwise uniqueness is stronger than weak uniqueness, provided that there exist solutions.

5.2 Lipschitz coefficients

For $U \subseteq \mathbb{R}^d$ and $f : U \rightarrow \mathbb{R}^d$, say f is *Lipschitz* with *Lipschitz constant* $K < \infty$ if

$$|f(x) - f(y)| \leq K|x - y| \quad \text{for all } x, y \in U, \quad (5.11)$$

where $|\cdot|$ denotes the Euclidean norm on \mathbb{R}^d . (If $f : U \rightarrow \mathbb{R}^{d \times m}$ then the left-hand side is the Euclidean norm in $\mathbb{R}^{d \times m}$). The key result of this section will be that SDE with Lipschitz coefficients have pathwise unique solutions which are furthermore always strong.

We start preparing for this result by recalling two important results which will be used in the proof.

Theorem 5.2 Contraction Mapping Theorem

Let (X, d) be a complete metric space and $F : X \rightarrow X$. Suppose that the iterated function F^n is a contraction for some $n \in \mathbb{N}$, i.e.

$$\exists r < 1 \forall x, y \in X : d(F^n(x), F^n(y)) \leq r d(x, y). \quad (5.12)$$

Then F has a unique fixed point.

Remark. This theorem is most well-known when F itself is contractive, rather than F^n . However, the theorem for $n \geq 1$ easily follows from the $n = 1$ result. Indeed, if $n \geq 1$ and F^n is contractive, then (by the theorem for $n = 1$) F^n must have a unique fixed point x . We claim that x is also a fixed point of F . Indeed, let $x_1 = F(x), x_2 = F^2(x), \dots, x_{n-1} = F^{n-1}(x)$. Then since $F^n(x) = x$, we have

$$\begin{aligned} F^n(x_1) &= F(F^{n-1}(x_1)) \\ &= F(F^n(x)) \\ &= F(x) = x_1 \end{aligned}$$

so x_1 is a fixed point of F^n as well. But F^n has a unique fixed point, so $x = x_1$. Therefore, $F(x) = x_1 = x$ and x_1 is a fixed point of F .

Lemma 5.3 Gronwall's Lemma

Let $T > 0$ and let f be a non-negative bounded measurable function on $[0, T]$. Suppose that

$$\exists a, b \geq 0 \forall t \in [0, T] : f(t) \leq a + b \int_0^t f(s) ds . \quad (5.13)$$

Then $f(t) \leq a \exp(bt)$ for all $t \in [0, T]$. In particular if $a = 0$ then $f = 0$.

Theorem 5.4 Suppose that $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$ and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are Lipschitz. Then there is pathwise uniqueness for the SDE

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt . \quad (5.14)$$

Moreover, for each filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and each $(\mathcal{F}_t)_{t \geq 0}$ -Brownian motion B , there exists a strong solution to the SDE started from x , for any $x \in \mathbb{R}^d$.

Proof. (for $d = m = 1$). Fix $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and B . Let $(\mathcal{F}_t^B)_{t \geq 0}$ be the natural filtration generated by B so that $\mathcal{F}_t^B \subseteq \mathcal{F}_t$ for all $t \geq 0$. Suppose that K is the Lipschitz constant for σ and b .

Pathwise uniqueness:

Suppose X and X' are two solutions on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ with $X_0 = X'_0$ a.s.. Fix M and let

$$\tau = \inf \{t \geq 0 : |X_t| \vee |X'_t| \geq M\} . \quad (5.15)$$

Then $X_{t \wedge \tau} = X_0 + \int_0^{t \wedge \tau} \sigma(X_s) dB_s + \int_0^{t \wedge \tau} b(X_s) ds$, and similarly for X' .

Let $T > 0$. If $0 \leq t \leq T$ then, since

$$(x + y)^2 \leq 2(x^2 + y^2) \quad (5.16)$$

for all $x, y \in \mathbb{R}$, we have:

$$\begin{aligned}
& \mathbb{E}((X_{t \wedge \tau} - X'_{t \wedge \tau})^2) \\
& \leq 2\mathbb{E}\left(\left(\int_0^{t \wedge \tau} (\sigma(X_s) - \sigma(X'_s)) dB_s\right)^2\right) + 2\mathbb{E}\left(\left(\int_0^{t \wedge \tau} (b(X_s) - b(X'_s)) ds\right)^2\right) \\
& \leq 2\mathbb{E}\left(\int_0^{t \wedge \tau} (\sigma(X_s) - \sigma(X'_s))^2 ds\right) + 2T\mathbb{E}\left(\int_0^{t \wedge \tau} (b(X_s) - b(X'_s))^2 ds\right) \\
& \quad (\text{by the It\^o isometry and the Cauchy-Schwarz inequality}) \\
& \leq 2K^2(1+T)\mathbb{E}\left(\int_0^{t \wedge \tau} (X_s - X'_s)^2 ds\right) \quad (\text{by the Lipschitz property}) \\
& \leq 2K^2(1+T)\int_0^t \mathbb{E}((X_{s \wedge \tau} - X'_{s \wedge \tau})^2) ds. \tag{5.17}
\end{aligned}$$

Let $f(t) = \mathbb{E}((X_{t \wedge \tau} - X'_{t \wedge \tau})^2)$. Then $f(t)$ is bounded by $4M^2$ and

$$f(t) \leq 2K^2(1+T)\int_0^t f(s) ds. \tag{5.18}$$

Hence, by Gronwall's lemma, $f(t) = 0$ for all $t \in [0, T]$. So $X_{t \wedge \tau} = X'_{t \wedge \tau}$ a.s. and, letting $M, T \rightarrow \infty$, we obtain that X and X' are indistinguishable.

Existence of a strong solution:

We start by constructing a weak solution as a fixed point of a certain mapping. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space and let B be a Brownian motion. Write \mathcal{C}_T for the set of continuous processes $X : [0, T] \rightarrow \mathbb{R}$ adapted to (\mathcal{F}_t) such that

$$\|X\|_T := \left\| \sup_{t \leq T} |X_t| \right\|_2 < \infty \tag{5.19}$$

and \mathcal{C} for the set of continuous adapted processes $X : [0, \infty) \rightarrow \mathbb{R}$ such that

$$\|X\|_T < \infty \quad \text{for all } T > 0. \tag{5.20}$$

Recall from Proposition 3.4 that $(\mathcal{C}_T, \|\cdot\|_T)$ is complete. Let $\mathcal{C}'_T = \mathcal{C}_T \cap \{X_0 = x\}$, and let Φ be a mapping defined on \mathcal{C}'_T by;

$$\Phi(X)_t = x + \int_0^t \sigma(X_s) dB_s + \int_0^t b(X_s) ds \quad \text{for all } t \leq T. \tag{5.21}$$

Note that a solution to $E(\sigma, b)$ is a fixed point of Φ . We start by showing that if $X \in \mathcal{C}'_T$, then so is $\Phi(X)$. For all $y \in \mathbb{R}$,

$$|\sigma(y)| \leq |\sigma(0)| + K|y|, \quad |b(y)| \leq |b(0)| + K|y|. \tag{5.22}$$

Suppose $X \in \mathcal{C}_T$ for some T . Let $M_t = \int_0^t \sigma(X_s) dB_s$, $0 \leq t \leq T$. Then $[M]_T = \int_0^T \sigma(X_s)^2 ds$ and so by (5.16)

$$\mathbb{E}([M]_T) \leq 2T(|\sigma(0)|^2 + K^2\|X\|_T^2) < \infty. \tag{5.23}$$

Hence, by example sheet 2, problem 11, $(M_t)_{0 \leq t \leq T}$ is a martingale bounded in L^2 . So by Doob's L^2 inequality and (5.23)

$$\mathbb{E}\left(\sup_{t \leq T} \left| \int_0^t \sigma(X_s) dB_s \right|^2\right) \leq 8T(|\sigma(0)|^2 + K^2\|X\|_T^2) < \infty. \tag{5.24}$$

Therefore $(M_t, t \geq 0)$ belongs to \mathcal{C}_T . Similarly, by (5.22), (5.16) and the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E} \left(\sup_{t \leq T} \left| \int_0^t b(X_s) ds \right|^2 \right) &\leq T \mathbb{E} \left(\int_0^T |b(X_s)|^2 ds \right) \\ &\leq 2T^2 \left(|b(0)|^2 + K^2 \|X\|_T^2 \right) < \infty. \end{aligned} \quad (5.25)$$

Therefore, $(\int_0^t b(X_s) ds, t \geq 0)$ belongs to \mathcal{C}_T as well. By the triangular inequality, it follows that $\Phi(X) \in \mathcal{C}_T$ and thus $\Phi(X) \in \mathcal{C}'_T$ since by definition $\Phi(X)_0 = x$. Now, let $X, Y \in \mathcal{C}'_T$. By Doob's inequality again and (5.17),

$$\|\Phi(X) - \Phi(Y)\|_T^2 = \mathbb{E} \left(\sup_{0 \leq t \leq T} |\Phi(X)_t - \Phi(Y)_t|^2 \right) \quad (5.26)$$

$$\leq 2 \mathbb{E} \left(\sup_{0 \leq t \leq T} \left| \int_0^t \sigma(X_s) dB_s - \int_0^t \sigma(Y_s) dB_s \right|^2 \right) \quad (5.27)$$

$$+ 2 \mathbb{E} \left(\sup_{0 \leq t \leq T} \left| \int_0^t b(X_s) ds - \int_0^t b(Y_s) ds \right|^2 \right) \quad (5.28)$$

$$\begin{aligned} &\leq 2K^2(4+T) \int_0^T \mathbb{E}(|X_t - Y_t|^2) dt \\ &\leq \underbrace{2K^2(4+T)}_{C_T} \int_0^T \|X - Y\|_t^2 dt \end{aligned} \quad (5.29)$$

By induction using (5.29), we have for all $n \geq 0$ that

$$\begin{aligned} \|\Phi^n(X) - \Phi^n(Y)\|_T^2 &\leq C_T^n \|X - Y\|_T^2 \int_0^T \int_0^{t_{n-1}} \dots \int_0^{t_2} dt_1 \dots dt_{n-1} \\ &= \frac{C_T^n T^n}{n!} \|X - Y\|_T^2 \quad \text{by symmetry.} \end{aligned} \quad (5.30)$$

For n sufficiently large, Φ^n is a contraction on the complete metric space $(\mathcal{C}'_T, \|\cdot\|_T)$. Hence, by the Contraction Mapping Theorem, Φ has a unique fixed point which we may call $X^{(T)} \in \mathcal{C}'_T$. By uniqueness of this fixed point, $X_t^{(T)} = X_t^{(T')}$ for all $t \leq T \wedge T'$ *a.s.* and so we may consistently define $X \in \mathcal{C}$ by

$$X_t = X_t^{(N)} \quad \text{for } t \leq N, \quad N \in \mathbb{N}. \quad (5.31)$$

This is the pathwise unique solution to the SDE started from x . It remains to prove that it is $(\mathcal{F}_t^B)_{t \geq 0}$ -adapted. Define a sequence $(Y^n)_{n \geq 0}$ in \mathcal{C}_T by

$$Y^0 \equiv x, \quad Y^n = \Phi(Y^{n-1}) \quad \text{for } n \geq 1. \quad (5.32)$$

Then Y^n is $(\mathcal{F}_t^B)_{t \geq 0}$ -adapted for each $n \geq 0$. Since $X = \Phi(X)$ for all $n \geq 0$ by (5.30) we have

$$\|X - Y^n\|_T^2 \leq \frac{C_T^n T^n}{n!} \|X - x\|_T^2. \quad (5.33)$$

Hence, $Y^n \rightarrow X$ in \mathcal{C}_T and thus $Y_t^n \rightarrow X_t$ in probability for a fixed $t > 0$. Thus there exists a subsequence n_k such that $Y_t^{n_k} \rightarrow X_t$ almost surely. Since $Y_t^{n_k}$ is \mathcal{F}_t^B -measurable, then so is X_t . Therefore X is $(\mathcal{F}_t^B)_{t \geq 0}$ -adapted and the proof of this theorem is finished. \square

Remark. The uniqueness of the fixed point in the contraction mapping theorem can not be invoked directly to prove pathwise uniqueness of the solutions. What this result give is pathwise uniqueness of solutions in \mathcal{C}_T for any $T > 0$. So if we knew a priori that any solution belongs to \mathcal{C}_T , we could invoke this result. (Note that our proof that $\Phi(X) \in \mathcal{C}_T$ relies on the fact that X is already assumed to be in \mathcal{C}_T). Thus a byproduct of our proof is that indeed any solution belongs to \mathcal{C}_T for any $T > 0$.

Corollary 5.5 *Let σ_{ij}, b_i be Lipschitz functions on \mathbb{R}^d for $1 \leq i, j \leq d$. Then every solution to $E_x(\sigma, b)$ is strong, and there is uniqueness in distribution for the solutions to $E(\sigma, b)$.*

Proof. The proof of the theorem constructs a strong solution for every filtered probability space and Brownian motion defined on it. On the other hand there is pathwise uniqueness of solutions so any solution must be strong. By the Yamada-Watanabe theorem, it also follow from existence of solutions and pathwise uniqueness that uniqueness in distribution holds.

Example. Ornstein-Uhlenbeck process

Fix $\lambda \in \mathbb{R}$ and consider the SDE in \mathbb{R}^2

$$dV_t = dB_t - \lambda V_t dt, \quad V_0 = v_0, \quad dX_t = V_t dt, \quad X_0 = x_0. \quad (5.34)$$

When $\lambda > 0$ this models the motion of a pollen grain on the surface of a liquid, and λ then represents the *viscosity* of that liquid. X represents the x -coordinate of the grain's position and V represents its velocity in the x -direction. $-\lambda V$ is the friction force due to viscosity. Whenever $|V|$ becomes large, the system acts to reduce it. (This is a much more realistic model of random motion from a physical point of view than Brownian motion which oscillates too wildly!) V is called the *Ornstein-Uhlenbeck (velocity) process*. Then there is pathwise uniqueness for this SDE. In fact, this is a rare example of a SDE we can solve explicitly, see example sheet 4.

5.3 Strong Markov property and diffusion processes

In an ordinary differential equation, the future of the trajectory of a particle is entirely determined by its present position. The stochastic analogue for stochastic differential equations is true as well: solutions to SDE's have the strong Markov property, i.e., the distribution of their future depends only on their present position. (In fact, SDE solutions should be viewed as the prototypical example of a strong Markov process.)

Theorem 5.6 (Strong markov property). *Assume that σ and b are two Lipschitz functions. Then for all $x \in \mathbb{R}^d$, if X^x denotes a weak solution started from x to $E(\sigma, b)$, if F is any measurable nonnegative functional on $C([0, \infty), \mathbb{R}^d)$ then almost surely:*

$$\mathbb{E} [F(X_{T+t}^x, t \geq 0) | \mathcal{F}_T] = \mathbb{E} [F(X_t^y, t \geq 0)] \Big|_{y=X_T}. \quad (5.35)$$

Proof. As we will see, the strong Markov property is a relatively straightforward consequence of Corollary 5.5. Let $Y_t = X_{T+t}^x$. Since X is a solution to $E_x(\sigma, b)$, we have

$$X_{T+t}^x - X_T^x = \int_T^{T+t} \sigma(X_s^x) dB_s + \int_T^{T+t} b(X_s^x) ds \quad (5.36)$$

To make the change of variable $u = t + T$, we use the following Lemma:

Lemma 5.7 *Let H be a previsible locally bounded process, and let X be a continuous local martingale. If T is a stopping time and $X^{(T)} = (X_{t+T} - X_T, t \geq 0)$ then*

$$\int_T^{t+T} H_s dX_s = \int_0^t H_{T+u} dX_u^{(T)}$$

Proof. The proof of this lemma is an easy application of the monotone class theorem after observing that the statement is trivial for processes of the form $H = \mathbf{1}_{\{A \times (s, t]\}}$ where $A \in \mathcal{F}_s$. \square

Thus, if $y = X_T$, then making the change of variable in (5.36) we get:

$$Y_t = y + \int_0^t \sigma(Y_u) dB_u^{(T)} + \int_0^t b(Y_u) du$$

where $B_t^{(T)} = B_{T+t} - B_T$ is a Brownian motion independent from \mathcal{F}_T . Y is adapted to the filtration $(\mathcal{F}_{T+u}, u \geq 0)$ which satisfies the usual conditions. Therefore, the previous theorem applies and Y is adapted to $(\mathcal{G}_t)_{t \geq 0}$, where for all $t > 0$, \mathcal{G}_t is the σ -field generated by X_T and $B_s^{(T)}, s \leq t$. Thus, we can write $(Y_t)_{t \geq 0}$ as a certain deterministic and measurable functional Φ of its starting point X_T and the driving Brownian motion, $\Phi(X_T, B^{(T)})$. Furthermore, note that by definition $\Phi(y, B)$ is the unique solution to $E_y(\sigma, b)$ corresponding to the driving Brownian motion B . Hence (by weak uniqueness) $\Phi(y, B)$ has the same law as X^y . Since $B^{(T)}$ is independent from \mathcal{F}_T , it is independent from X_T (because X is adapted to \mathcal{F}). It follows that the left-hand side of (5.35) may be computed as:

$$\begin{aligned} \mathbb{E}[F(Y_t, t \geq 0) | \mathcal{F}_T] &= \mathbb{E}\left[F\left(\Phi\left(X_T, B^{(T)}\right)\right) | \mathcal{F}_T\right] \\ &= \mathbb{E}\left[F\left(\Phi\left(y, B^{(T)}\right)\right)\right]_{y=X_T} \text{ a.s.} \\ &= \mathbb{E}[F(X_t^y, t \geq 0)]_{y=X_T} \text{ a.s.} \end{aligned}$$

which is exactly the content of the strong Markov property. \square

In the remainder of this course we now provide a brief introduction to the theory of diffusion processes, which are Markov processes characterized by martingale properties. We first construct these processes with SDE's and then move on to describe some fundamental connection with PDE's. In the next section we show how diffusions arise as scaling limits of Markov chains.

Define, for $f \in C^2(\mathbb{R}^d)$,

$$Lf(x) = \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x) \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial f}{\partial x_i} \quad (5.37)$$

where $a_{i,j}(x)$ is a bounded measurable function called the diffusivity and $b(x)$, another bounded measurable function, is called the drift. We assume that $(a_{i,j}(x))_{i,j}$ is a symmetric nonnegative matrix for all $x \in \mathbb{R}^d$.

Definition 5.2 Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space satisfying the usual conditions. Say that a process $X = (X_t, t \geq 0)$ is an L -diffusion if for all $f \in C_b^2(\mathbb{R}^d)$, the process M^f is a martingale, where for all $t \geq 0$:

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Lf(X_s) ds. \quad (5.38)$$

For the moment, we don't know whether such processes exist, and we haven't shown any sort of uniqueness. The following result takes care of the existence part.

Theorem 5.8 Let X be a solution (in \mathbb{R}^d) to the SDE

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt$$

where B is a (\mathcal{F}_t) -Brownian motion in \mathbb{R}^d , and where $\sigma = (\sigma_{i,j}(x))_{1 \leq i,j \leq d}$ and $b = (b_i(x))_{1 \leq i \leq d}$ are bounded measurable. Then for all $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$,

$$M_t^f = f(t, X_t) - f(0, X_0) - \int_0^t \left(\frac{\partial f}{\partial t} + Lf \right) (s, X_s) ds \quad (5.39)$$

is a local martingale, where $a = \sigma \sigma^T$. In particular, if the coefficients σ, b are bounded, then X is an L -diffusion.

This result follows simply from an application of Itô's formula.

Remarks.

1. If $a_{i,j}$ is uniformly positive definite (that is, there exists $\varepsilon > 0$ such that

$$\xi_i a_{ij}(x) \xi_j \geq \varepsilon \|\xi\|^2$$

for all $\xi \in \mathbb{R}^d$ and all $x \in \mathbb{R}^d$), then a has a positive-definite square root matrix σ . If a is furthermore Lipschitz, then it can be shown that $\sigma(x)$ is also Lipschitz. It follows that if a, b are bounded Lipschitz functions and a is uniformly positive definite, then L -diffusions exist, by Theorem 5.4, for any given starting point X_0 .

2. Brownian motion in \mathbb{R}^d is an L -diffusion for $L = \frac{1}{2} \Delta$.
3. In the language of Markov processes, we say that L is the *infinitesimal generator* of X . Intuitively, $Lf(x)$ describes the infinitesimal expected change in $f(X)$ given that $X_t = x$. That is,

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left(\frac{f(X_{t+\varepsilon}) - f(X_t)}{\varepsilon} \middle| \mathcal{F}_t, X_t = x \right) = Lf(x)$$

for every $f \in C_b^{1,2}(\mathbb{R}^d)$.

5.4 Some links with PDEs

In this section we state a theorem which we do not prove due to time constraints. (Parts of this result are easy and other less trivial...)

Theorem 5.9 *Let D be an open set in \mathbb{R}^d . Let L be defined by (5.37) for uniformly positive definite Lipschitz bounded coefficients a, b . Let $g \in C(\partial D)$ and let $\phi \in C(\bar{D})$ such that both ϕ and g are bounded. Define:*

$$u(x) = \mathbb{E}_x \left(\int_0^T \phi(X_s) ds + g(X_T) \right), \quad x \in D$$

where X is an L -diffusion and $T = \inf\{t > 0 : X_t \notin D\}$. Then u is the unique continuous function on \bar{D} which is solution to the Dirichlet problem:

$$\begin{cases} Lu + \phi = 0 & \text{in } D \\ u = g & \text{on } \partial D. \end{cases}$$

Another link is provided by the following Cauchy problem – that is, an evolution problem for which the initial condition is prescribed.

Theorem 5.10 *Let $g : \mathbb{R}^d \rightarrow \mathbb{R}$ be a given continuous bounded function, and let X be an L diffusion where L satisfies the same assumptions as in Theorem 5.9. Then if we define:*

$$u(t, x) = \mathbb{E}_x(g(X_t)) \text{ for all } t \geq 0, x \in \mathbb{R}^d$$

then u is the unique solution in $C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ to the problem:

$$\begin{cases} \frac{\partial u}{\partial t} = Lu & \text{on } \mathbb{R}_+ \times \mathbb{R}^d \\ u(0, \cdot) = g & \text{on } \mathbb{R}^d. \end{cases}$$

One word about the proof of the uniqueness part: let v be a solution to this problem, and let u be our candidate. Let us show that $v = u$. Fix $T > 0$ and let $f(t, x) = v(T - t, x)$. Applying Theorem 5.8 to the function f , we see that

$$M_t = v(T - t, X_t), \quad 0 \leq t \leq T$$

is a martingale. Thus

$$M_t = \mathbb{E}(M_T | \mathcal{F}_t) \text{ a.s.}$$

and it follows that $v(T - t, X_t) = \mathbb{E}_x(g(X_T) | \mathcal{F}_t)$, almost surely. However, we know that X is Markovian so the right-hand side of this equation can also be written:

$$v(T - t, X_t) = \mathbb{E}_{X_t}(g(X_{T-t})), \quad \text{a.s.}$$

Writing $y = X_t$ and $s = T - t$ (which is arbitrary since T is arbitrary) we have:

$$v(s, y) = \mathbb{E}_y(g(X_s)).$$

It can be shown that the support of the distribution of X_t is \mathbb{R}^d (this a consequence of uniform positive definiteness). It thus follows that this equation must be true for all $y \in \mathbb{R}^d$. The uniqueness part of the Theorem is proved.

Remark. Note that the Cauchy problem may reformulated as a Dirichlet problem in \mathbb{R}^{d+1} by changing L into

$$\tilde{L} = L - \frac{\partial}{\partial t}$$

Fix a point $(t, x) \in \mathbb{R}^{d+1}$. By Theorem 5.9, the solution $u(t, x)$ is given by $\mathbb{E}(\tilde{X}_T)$ where X is the diffusion with generator \tilde{L} . This corresponds to adding a coordinate X^{d+1} to the diffusion X , such that $X_s^{d+1} = X_0^{d+1} - s$, that is, time is decreasing at speed 1. The time T corresponds to the first time that the “time” coordinate hits 0, i.e., time t if we start from (x, t) . The other d coordinates are then distributed according to $\mathbb{P}_x(X_t \in \cdot)$. This proves Theorem 5.10, given Theorem 5.9.

Example. Let $(B_t, t \geq 0)$ be a 3-dimensional Brownian motion with $B_0 = 0$. Let $\tau = \inf\{t > 0 : \|B_t\| = 1\}$. Compute $\mathbb{E}(\tau)$. Answer: Let $R_t = \|B_t\|$. Then an application of Itô’s formula shows that

$$dR_t = dB_t + \frac{1}{R_t} dt.$$

It follows that $(R_t, t \geq 0)$ is a diffusion process on $(0, \infty)$, with generator:

$$L = \frac{1}{2} \frac{d^2}{dx^2} + \frac{1}{x} \frac{d}{dx}$$

Thus if $\phi \equiv 1$ and $g \equiv 0$ in the previous theorem, $\mathbb{E}_0(\tau) = u(0)$ where $u(x) = \mathbb{E}_x(\tau)$ is a function solving:

$$Lu = -1, \quad \text{for all } x \in (0, 1).$$

Solving this ODE yields that if $f = u'$ then

$$f(x) = \left(-\frac{2}{3}x^3 - c\right)x^{-2}$$

for some constant $c \in \mathbb{R}$, so integrating:

$$u(x) = -\frac{1}{3}x^2 + \frac{c}{x} + c'$$

for a constant $c' \in \mathbb{R}$. But we note that c must be equal to 0. Indeed, otherwise $\mathbb{E}_0(\tau) = \infty$ by the monotone convergence theorem, and if we apply the Markov property at time $t > 0$, we get

$$\begin{aligned} \mathbb{E}_0(\tau) &= \mathbb{E}_0(\tau \mathbf{1}_{\{\tau < t\}}) + \mathbb{E}_0(\tau \mathbf{1}_{\{\tau \geq t\}}) \\ &\leq t + \int_0^1 \mathbb{P}_0(\tau > t; R_t \in dx)(t + \mathbb{E}_x(\tau)) \\ &\leq t + t + \int_0^1 4\pi x^2 dx \frac{e^{-x^2/2t}}{(2\pi t)^{3/2}} \mathbb{E}_x(\tau) \end{aligned}$$

If $c \neq 0$, then $\mathbb{E}_x(\tau) \sim c/x$ as $x \rightarrow 0$. This means that the integral on the right-hand side is finite and hence $\mathbb{E}_0(\tau) < \infty$, which is a contradiction. Thus $c = 0$ and $\mathbb{E}_x(\tau) < \infty$.

Thus

$$u(x) = -\frac{1}{3}x^2 + c'$$

and since $u(1) = 0$ we have $u(x) = \frac{1}{3}(1 - x^2)$. Hence $\mathbb{E}_0(\tau) = 1/3$.

Exercise. The above calculation is not rigorous because of the singularity of $1/x$ at 0 (hence the SDE is not even properly defined at 0, let alone Lipschitzian...) Make it rigorous by considering the diffusion R_t started at level ε and taking $D' = (\varepsilon', 1)$ with $\varepsilon' \ll \varepsilon$. Letting ε' and $\varepsilon \rightarrow 0$, recover the fact that $\mathbb{E}_0(\tau) = 1/3$.

We now present a result called the Feynman-Kac formula, which is a systematic way of solving a class of PDE's by using Brownian motion.

Theorem 5.11 *Let $f \in C_b^2(\mathbb{R}^d)$ and let $V \in L^\infty(\mathbb{R}^d)$, that is, V is uniformly bounded. For all $t, x \geq 0$, let*

$$u(t, x) = \mathbb{E}_x \left(f(B_t) \exp \left(\int_0^t V(B_s) ds \right) \right)$$

Then u is the unique solution $w \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$:

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{2} \Delta u + Vu & \text{on } \mathbb{R}_+ \times \mathbb{R}^d \\ u(0, \cdot) = f & \text{on } \mathbb{R}^d. \end{cases}$$

Proof. Here again, the uniqueness part is an easy application of Itô's formula. Let u be a solution and let $M_t = u(T - t, B_t)E_t$ where $E_t = \exp(\int_0^t V(B_s)ds)$ is of finite variation. By Itô's formula:

$$\begin{aligned} dM_t &= \nabla u(T - t, B_t)E_t dB_t + \left(-\dot{u} + \frac{1}{2} \Delta u + Vu \right) (T - t, B_t)E_t dt \\ &= \nabla u(T - t, B_t)E_t dB_t \end{aligned}$$

since the second term is equal to 0 (because u is a solution to the PDE problem). Thus M is a local martingale, and it is uniformly bounded on $[0, T]$, hence a true martingale. By the Optional Stopping Theorem:

$$u(T, x) = \mathbb{E}_x(M_0) = \mathbb{E}_x(M_T) = \mathbb{E}_x(f(B_T)E_T)$$

which is precisely the claim.

Remark. This formula turns out to be very useful when applied the other way round: in fact, it was originally introduced to compute expectations involving exponential functionals of Brownian motion, which tend to occur frequently in mathematical finance and in statistical mechanics, where V is a potential. (This is presumably why Feynman got interested in this problem). Then we can write:

$$\mathbb{E}_x \left(\exp \left\{ -\beta \int_0^T V(X_s) ds \right\} f(X_T) \right) = u(x, T)$$

where

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u - \beta uV \text{ on } \mathbb{R}_+ \times \mathbb{R}^d$$

and $u(x, 0) = f(x)$ for all $x \in \mathbb{R}^d$. This often makes these computations easier, by bringing in techniques that were developed in analysis (e.g., Fourier analysis). In mathematical finance, the Feynman-Kac formula allows to compute the Black-Scholes formula for the price of a call in terms of a certain PDE. This point of view is in some sense dual to ours, and it is a great advantage to have these two approaches for what is, fundamentally, the same object.

6 Diffusion approximation

6.1 Martingale problems.

Let $\sigma_{i,j}(x)_{1 \leq i,j \leq d}$ and $(b_i(x))_{1 \leq i \leq d}$ be a family of bounded and measurable functions with values in \mathbb{R} . Let $a(x) = \sigma(x)\sigma^T(x)$. (Here again we assume that $x \in \mathbb{R}^d$, so $m = d$).

Definition 6.1 We say that a process $X = (X_t, t \geq 0)$ with values in \mathbb{R}^d , together with a filtered probability space $(\Omega, (\mathcal{F}_t), \mathbb{P})$, solves the martingale problem $\mathbf{M}(a, b)$ if for all $1 \leq i, j \leq d$,

$$\left(X_t^i - \int_0^t b_i(X_s) ds; t \geq 0 \right)$$

and

$$\left(X_t^i X_t^j - \int_0^t a_{i,j}(X_s) ds; t \geq 0 \right)$$

are local martingales.

Of course, the second condition means that $[X^i, X^j]_t = \int_0^t a_{i,j}(X_s) ds$.

For instance, if σ, b are in addition Lipschitz, then there exists $(\Omega, X, (\mathcal{F})_{t \geq 0})$ and an (\mathcal{F}_t) -Brownian motion $(B_t, t \geq 0)$ solution to the stochastic differential equation:

$$dX_t = \sigma(X_t) \cdot dB_t + b(X_t) dt.$$

X then solves the martingale problem $\mathbf{M}(a, b)$. In fact, note that any (weak) solution to $E(\sigma, b)$ gives a solution to the martingale problem $\mathbf{M}(a, b)$. Rather remarkably, these are the only solutions.

Theorem 6.1 Let $a = \sigma\sigma^T$ and let X and $(\Omega, (\mathcal{F}_t), \mathbb{P})$ be a solution to $\mathbf{M}(a, b)$. Then there exists an (\mathcal{F}_t) -Brownian motion $(B_t, t \geq 0)$ in \mathbb{R}^d defined on an enlarged probability space, such that (X, B) solves $E(\sigma, b)$.

Proof. Assume first that σ is invertible for every $x \in \mathbb{R}^d$. Then define $Y_t^i = X_t^i - \int_0^t b_s^i(X_s) ds$, so that $Y^i \in \mathcal{M}_{c,loc}$. Define:

$$B_t^i = \int_0^t \sum_{k=1}^d \sigma_{i,k}^{-1}(X_s) dY_s^k$$

Thus $B^i \in \mathcal{M}_{c,loc}$ and we have

$$[B^i, B^j]_t = \sum_{k,\ell=1}^d \int_0^t \sigma_{i,k}^{-1}(X_s) \sigma_{j,\ell}^{-1}(X_s) a_{k,\ell}(X_s) ds = \delta_{i,j} t$$

so by Lévy's characterization, B is an Brownian motion in \mathbb{R}^d .

Moreover, by the "stochastic version of the fundamental theorem of calculus",

$$\int_0^t \sigma(X_s) dB_s = Y_t - Y_0 = X_t - \int_0^t b(X_t) dt. \quad (6.1)$$

Indeed the i^{th} component of the left-hand side may be written as

$$\sum_{j=1}^d \int_0^t \sigma_{i,j}(X_s) dB_s^j = \int_0^t \sum_{j,k=1}^d \sigma_{i,j} \sigma_{j,k}^{-1} dY_s^k = \int_0^t dY_s^i.$$

But (6.1) is simply the statement that (X, B) solves $E(\sigma, b)$.

When σ is not everywhere invertible, we proceed like in the generalized version of Dubins-Schwartz's (when $[M]_\infty$ is not assumed to be infinity) and let the Brownian motion evolve independently when s is such that $\sigma(X_s)$ is not invertible. See pp.190-191 of Revuz-Yor for the details and Durrett, p.200 for the case $d = 1$. \square

Theorem 6.1 shows that there is a one-to-one correspondence between solutions to the stochastic differential equation $E(\sigma, b)$ and the martingale problem $\mathbf{M}(a, b)$. In particular, there is *uniqueness in distribution* to the solutions of $E(\sigma, b)$, if and only if the solutions to the martingale problem $\mathbf{M}(a, b)$ are unique, where uniqueness means that all solutions to $\mathbf{M}(a, b)$ with identical starting points have the same law.

6.2 Notions of weak convergence of processes

In this section we describe some basic results in the theory of weak convergence of processes, which we do not prove due to the time constraints. We will however use these results in the next section about the convergence of Markov chains towards solutions of certain stochastic differential equations.

The point of view here is similar to the one in Donsker's theorem. We view a process as a random variable with values in the space Ω of trajectories. We thus need to recall a few notions about weak convergence in general metric space. Let (S, d) be a metric space. The distance function $d(x, y)$ satisfies $d(x, y) = 0$ if and only if $x = y$; $d(x, y) \geq 0$; $d(x, z) \leq d(x, y) + d(y, z)$. The open ball $B(x, r)$ is the set $\{y \in S : d(x, y) < r\}$. The Borel σ -field is the field generated by all open sets.

The notion of convergence in distribution is defined in terms of test functions, which are only required to be bounded and continuous (for the topology of S):

Definition 6.2 *Let $(\mu_n)_{n \geq 1}$ be a sequence of probability distributions on S . We say that $\mu_n \rightarrow \mu$ weakly as $n \rightarrow \infty$, if $\int_S f d\mu_n \rightarrow \int_S f d\mu$ as $n \rightarrow \infty$ for all bounded continuous functions f . If μ_n is the law of a random variable X_n and μ that of a random variable X , we say that $X_n \rightarrow X$ in distribution (or in law).*

There are a number of ways one can reformulate the notion of weak convergence in terms of the mass assigned to events that are either closed or open. If $A \subseteq S$, we recall the definition of the frontier of A , which is the set $\partial A := \bar{A} \setminus \text{int}(A)$.

Theorem 6.2 *Let $(X_n)_{n \geq 1}$ be a sequence of random variables with values in S . The following are equivalent.*

- (i) $X_n \rightarrow X$ in distribution.
- (ii) For all closed set K , $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in K) \leq \mathbb{P}(X \in K)$.
- (iii) For all open set O , $\liminf_{n \rightarrow \infty} \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$.

- (iv) For all sets A such that $\mathbb{P}(X \in \partial A) = 0$, $\limsup_{n \rightarrow \infty} \mathbb{P}(X \in A) = \mathbb{P}(X \in A)$.
- (v) For all sets A such that $\mathbb{P}(X \in \partial A) = 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(X \in A) = \mathbb{P}(X \in A)$.
- (vi) For any bounded function f , denote by D_f the set of discontinuities of f . Then for any f such that $\mathbb{P}(X \in D_f) = 0$, $\mathbb{E}(f(X_n)) \rightarrow \mathbb{E}(f(X))$ as $n \rightarrow \infty$.

It is important to note that the random variables X_n need not be related in any particular way. In fact they may even be defined on different probability spaces. However, it turns out that (provided the metric space is sufficiently nice), one can always choose a common probability space for the random variables and define a sequence of random variables Y_n with law identical to X_n , in such a way that convergence occurs almost surely. This is the content of the ‘‘Skorokhod representation theorem’’, which we may occasionally need.

Lemma 6.3 *Suppose S is complete and separable. If $\mu_n \rightarrow \mu$ weakly then there exists random variables Y_n defined on $\Omega = [0, 1]$ equipped with the Lebesgue measure \mathbb{P} , such that $Y_n \stackrel{d}{=} \mu_n$ for all $n \geq 1$, and $\lim_{n \rightarrow \infty} Y_n = Y$, \mathbb{P} -almost surely, where $Y \stackrel{d}{=} \mu$.*

We now specialize to the case where the random variables X_n take values in the space C of continuous trajectories over the compact interval $[0, 1]$. This is precisely the point of view in Donsker’s theorem. We equip C with the distance of the sup-norm:

$$d(f, g) = \|f - g\|_\infty = \sup_{t \in [0, 1]} |f(t) - g(t)|.$$

This turns C into a complete, separable metric space, on which it makes sense to talk about weak convergence.

Example. If $(S_n, n \geq 0)$ is a simple random walk on \mathbb{Z} , then by Donsker’s theorem: $(S_t^{[N]}, 0 \leq t \leq 1)$, converges weakly towards a Brownian motion on $[0, 1]$, where $S_t^{[N]} = N^{-1/2} S_{Nt}$.

Exercise. Use Donsker’s theorem and (vi) in Theorem 6.2 to show that if L_N is the last time before time N that $S_n = 0$:

$$L_N = \max\{n \leq N : S_n = 0\},$$

then $L_N/N \rightarrow L$ in distribution, where L is the last zero of a Brownian motion before time 1. (By an exercise in the first example sheet, this has the arcsine law). The issue here is that the functional which associates to a path $\omega \in C$ its last zero $L(\omega)$ before time 1, is not continuous with respect to the topology of C : for instance, consider the sequence of function $f_\varepsilon(t) = \varepsilon t, 0 \leq t \leq 1$, as $\varepsilon \rightarrow 0$. However, it turns out that the set of functions for which L is discontinuous has zero Wiener measure, and so the functions L_N/N also converge in distribution to L .

A classical trick in analysis for proving convergence of a sequence x_n towards a limit x is to prove that (a) the sequence takes its values in a compact set, and (b) there can only be one subsequential limit. It is usually part (a) which demands slightly harder work to establish, as part (b) follows from usually softer arguments (you typically have identified the limit at this stage). Fortunately there is a general criterion and fairly easy

to use in practice, which tells us when the set $K = \{x_n\}_{n=1}^\infty$ is compact (or, actually, precompact, meaning that \bar{K} is compact). When this happens, we say that the sequence of processes $(X^n)_{n \geq 1}$ is *tight*. This criterion consists in, roughly speaking, showing that the process doesn't oscillate too wildly. This is the content of the following theorem. For a continuous path $\omega(t), t \in [0, 1]$, let

$$\text{osc}_\delta(\omega) = \sup\{|\omega(s) - \omega(t)| : |s - t| \leq \delta\}.$$

osc_δ is simply the modulus of continuity of the path ω , at precision δ .

Theorem 6.4 *Suppose that $(X^n)_{n \geq 1}$ is a sequence of processes with values in C . Then X^n is tight, if and only if for each $\varepsilon > 0$, there exists $n_0, M \geq 1$ and $\delta > 0$ such that:*

- (i) $\mathbb{P}(|X^n(0)| > M) \leq \varepsilon$ for all $n \geq n_0$.
- (ii) $\mathbb{P}(\text{osc}_\delta > \varepsilon) \leq \varepsilon$.

To summarize, to show that a sequence converges weakly in C , it suffices to prove that (i) and (ii) hold above and that there is a unique weak subsequential limit. This is for instance the case if one has already established convergence of the finite-dimensional distributions, i.e., convergence of the k -dimensional vector $(X_{t_1}^n, \dots, X_{t_k}^n)$ towards $(X_{t_1}, \dots, X_{t_k})$, for any $k \geq 1$ and any choice of "test times" t_1, \dots, t_k . This could have been a possible route for proving Donsker's theorem, as convergence of finite-dimensional distributions is easy to establish.

Note that condition (i) in the above theorem says that the starting point of the process $X^n(0)$ takes values in a compact set with arbitrarily high probability. This is usually trivial since in general, the starting point of a process is a deterministic point such as 0.

In the next section, we will prove weak convergence of certain rescaled Markov chains towards diffusion processes. For this, we will usually use the fact that any weak subsequential limit must satisfy the associated martingale problem $\mathbf{M}(a, b)$ for which sufficient smoothness of the coefficients proves uniqueness in distribution. However there is one (small) additional difficulty in this case: it will be more natural to work with right-continuous processes X^n rather than with the linear-interpolation of X_n , which typically loses some of the Markov property. Let D be the space of right-continuous paths on $[0, 1]$. Without entering into the details, D can also be equipped with a complete separable metric d , which is called the *Skorokhod topology*. It can also be proved that if a sequence of right-continuous processes X^n satisfy (i) and (ii) in Theorem 6.4, then X^n is also tight and any subsequential limit X must be continuous, in the sense that $\mathbb{P}(X \in C) = 1$. Furthermore, weak convergence with respect to the Skorokhod topology towards a continuous process X , implies weak convergence in C of the linear interpolations. Another fact which will be needed is that if $x_n \rightarrow x$ in the Skorokhod topology, then $x_n(t) \rightarrow x(t)$ for all $t \geq 0$.

6.3 Markov chains and diffusions

The result which we now discuss is due to Stroock and Varadhan (Chapter 11 in [4]), and shows a link between rescaled Markov chains and certain diffusion processes. It

is applicable in a remarkably wide variety of contexts, of which we will only have the time to give one example. Our treatment follows rather closely the book of Durrett [1] which can be used to look up additional details.

While the idea for the statement of the result is in fact fairly simple, there is quite a bit of notation to introduce. We assume that a certain Markov chain is given to us. A certain scaling parameter $h > 0$ is going to 0, and we assume that the chain has already been rescaled, so it takes its values in a certain set $S_h \subseteq \mathbb{R}^d$. We will denote this Markov chain by $(Y_n^h, n \geq 1)$. The transition probabilities of Y are given by a transition kernel Π_h which may depend on $h > 0$:

$$\mathbb{P}(Y_{n+1}^h \in A | Y_n^h = x) = \Pi_h(x, A).$$

We define the random process on $[0, 1]$ by

$$X_t^h = Y_{h\lfloor t/h \rfloor}^h, t \in [0, 1]$$

so that X^h is almost surely right-continuous and is constant between two successive jumps of the chain, which may occur every h units of time for the process X^h . We let K_h denote the rescaled transition kernel:

$$K_h(x, dy) = \frac{1}{h} \Pi^h(x, dy).$$

Roughly, the conditions of the theorem states that “the infinitesimal mean variance of the jumps of X when $X = x$ are approximately given by $b(x)$ and $\sigma(x)$, respectively”. The conclusion states that X^h converges weakly towards the solution of $\mathbf{M}(a, b)$.

For $1 \leq i, j \leq d$, define:

$$a_{i,j} = \int_{|y-x| \leq 1} (y_i - x_i)(y_j - x_j) K_h(x, dy)$$

$$b_i^h(x) = \int_{|y-x| \leq 1} (y_i - x_i) K_h(x, dy)$$

$$\Delta_\varepsilon^h(x) = K^h(x, B(x, \varepsilon)^c).$$

Suppose that a_{ij} and b_i are continuous coefficients on \mathbb{R}^d for which the martingale problem $\mathbf{M}(a, b)$ is *well posed*, i.e., for each $x \in \mathbb{R}^d$ there is a unique in distribution process $(X_t, 0 \leq t \leq 1)$ such that $X_0 = x$ almost surely, and

$$X_t^i - \int_0^t b_i(X_s) ds \quad \text{and} \quad X_t^i X_t^j - \int_0^t a_{ij}(X_s) ds$$

are both local martingales.

Theorem 6.5 *Suppose that the above holds, and that for every $1 \leq i, j \leq d$, and every $R > 0$, every $\varepsilon > 0$,*

$$(i) \lim_{h \rightarrow 0} \sup_{|x| \leq R} |a_{ij}^h(x) - a_{ij}(x)| = 0.$$

$$(ii) \lim_{h \rightarrow 0} \sup_{|x| \leq R} |b_i^h(x) - b_i(x)| = 0.$$

$$(iii) \lim_{h \rightarrow 0} \sup_{|x| \leq R} \Delta_\varepsilon^h(x) = 0.$$

Then if $X_0^h = x_h \rightarrow x_0$, we have $(X_t^h, 0 \leq t \leq 1) \rightarrow (X_t, 0 \leq t \leq 1)$ weakly in D , and in particular, the linear interpolations of Y^h converge weakly in C .

The rest of this section is devoted to a proof of this result. By localization, one may replace (i), (ii) and (iii) by the following stronger conditions:

$$(i) \lim_{h \rightarrow 0} \sup_{x \in \mathbb{R}^d} |a_{ij}^h(x) - a_{ij}(x)| = 0.$$

$$(ii) \lim_{h \rightarrow 0} \sup_{x \in \mathbb{R}^d} |b_i^h(x) - b_i(x)| = 0.$$

$$(iii) \lim_{h \rightarrow 0} \sup_{x \in \mathbb{R}^d} \Delta_\varepsilon^h(x) = 0.$$

(iv) Moreover, $a_{i,j}^h$, b_i^h , Δ^h are uniformly bounded in h and x .

Step 1. Tightness

Let f be a bounded and measurable function. Define the operator L^h by

$$L^h f(x) = \int K_h(x, dy)(f(y) - f(x)). \quad (6.2)$$

This is the “generator” of the process: this represents the infinitesimal change in the function f when the process is at x . In particular, note that the process

$$f(Y_k^h) - \sum_{j=0}^{k-1} h L^h f(Y_j^h), \quad k = 0, 1, \dots \quad (6.3)$$

is a (discrete-time) martingale. For our proof of tightness we are going to need an estimate on the time needed by the chain to make a deviation of size roughly $\varepsilon > 0$, when it starts at position $y \in \mathbb{R}^d$. To do this, we introduce a function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $g \in C^2$, $0 \leq g \leq 1$ and $g(x) = 0$ if $x \geq 1$, while $g(0) = 1$. We also define for $x \in \mathbb{R}^d$, $f_\varepsilon(x) = g(|x|^2/\varepsilon^2)$ which is also C^2 , and becomes 0 when $|x| \geq \varepsilon$, and let $f_{y,\varepsilon}(x) = f_\varepsilon(y - x)$.

Lemma 6.6 *There exists $C_\varepsilon > 0$ such that $|L^h f_{y,\varepsilon}(x)| \leq C_\varepsilon$ for all $y, x \in \mathbb{R}^d$.*

Proof. This is simply an application of Taylor expansion. For $t \in [0, 1]$ let $\phi(t) = f_{y,\varepsilon}(x + t(y - x))$. Then by Taylor’s theorem, there exists $c_{x,y} \in [0, 1]$ such that

$$\begin{aligned} f_{y,\varepsilon}(y) - f_{y,\varepsilon}(x) &= \phi(1) - \phi(0) = \phi'(0) + \frac{1}{2!} \phi''(c_{x,y}) \\ &= \sum_{i=1}^d (y_i - x_i) D_i f(x) + \sum_{1 \leq i,j}^d (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) \end{aligned}$$

where $f \equiv f_{y,\varepsilon}$ and D_i and D_{ij} stand for $\frac{\partial f}{\partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ respectively, while $z_{xy} = x + c_{xy}(y - x)$.

To obtain $L^h f$, we integrate the above with respect to $K_h(x, dy)$, and get:

$$\begin{aligned}
L^h f_{y,\varepsilon}(x) &= \int K_h(x, dy) (f_{y,\varepsilon}(y) - f_{y,\varepsilon}(x)) \\
&\leq \left| \nabla f_{y,\varepsilon}(x) \cdot \int_{|y-x|\leq 1} (y-x) K_h(x, dy) \right| \\
&\quad + \left| \int_{|y-x|\leq 1} \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{y,\varepsilon}(z_{xy}) K_h(x, dy) \right| \\
&\quad + 2\|f_{y,\varepsilon}\|_\infty K_h(x, B(x, 1)^c).
\end{aligned}$$

Let $A_\varepsilon = \sup_x \|\nabla f_{y,\varepsilon}(x)\|$, let $B_\varepsilon = \sup_z \|D_{ij} f(z)\|$, where:

$$\|m_{ij}\| := \sup_{u \in \mathbb{R}^d: |u|=1} |\langle u, Mu \rangle|; \quad M = (m_{ij}).$$

Thus by Cauchy-Schwarz,

$$\left| \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{y,\varepsilon}(z_{xy}) \right| \leq |y-x|^2 \|D_{ij} f_{y,\varepsilon}(z_{xy})\|$$

hence

$$\begin{aligned}
L^h f_{y,\varepsilon}(x) &\leq A_\varepsilon \left| \int_{|y-x|\leq 1} (y-x) K_h(x, dy) \right| \\
&\quad + B_\varepsilon \int_{|y-x|\leq 1} |y-x|^2 K_h(x, dy) + 2K_h(x, B(x, 1)^c).
\end{aligned}$$

Since $\int_{|y-x|\leq 1} (y-x) K_h(x, dy) = b(x)$ and $\int_{|y-x|\leq 1} |y-x|^2 K_h(x, dy) = \sum_i a_{ii}^h(x)$ and since we have assumed in (iv) that all those quantities were uniformly bounded, we have proved the lemma. \square

To estimate $\text{osc}_\delta(X^h)$, we introduce the following random variables: $\tau_0 = 0$,

$$\tau_n = \inf\{t \geq \tau_{n-1} : |X_t^h - X_{\tau_{n-1}}^h| \geq \varepsilon\},$$

$$N = \min\{n : \tau_n > 1\}$$

$$\sigma = \min\{\tau_n - \tau_{n-1} : 1 \leq n \leq N\}$$

and, finally

$$\theta = \max\{|X^h(t) - X^h(t^-)| : 0 < t \leq 1\}.$$

The relation between these random variables and tightness is provided by the following lemma:

Lemma 6.7 *Assume that $\sigma > \delta$ and that $\theta \leq \varepsilon$. Then $\text{osc}_\delta(X^h) \leq 4\varepsilon$.*

Proof. The proof is straightforward. We want to show that for all $s, t \in [0, 1]$ with $|s - t| \leq \delta$, $|X^h(s) - X^h(t)| \leq 4\varepsilon$. The point is that since $|s - t| \leq \delta < \sigma$, s and t can only span at most one of the interval τ_{n-1}, τ_n , and by definition of these stopping times, everything behaves well on those intervals. Thus if $\tau_{n-1} \leq s < t < \tau_n$, then $|f(s) - f(t)| \leq 2\varepsilon$. If on the other hand, $\tau_{n-1} \leq s < \tau_n \leq t$, then

$$\begin{aligned} |f(s) - f(t)| &\leq |f(s) - f(\tau_{n-1})| + |f(t) - f(\tau_n)| \\ &\quad + |f(\tau_n) - f(\tau_n^-)| + |f(\tau_n^-) - f(\tau_{n-1})| \\ &\leq 4\varepsilon. \end{aligned}$$

We now use this to prove the tightness estimate. Since it is assumed that the starting point $X_0^h = x_0^h$ is nonrandom and converges towards a fixed x_0 , it suffices to prove the statement about oscillations: for all ε , there exists $\delta > 0$ and h_0 such that for all $h \leq h_0$,

$$\mathbb{P}(\text{osc}_\delta(X^h) \geq \varepsilon) \leq \varepsilon.$$

Thus it follows to prove that for all h sufficiently small and for δ small enough, $\mathbb{P}_y(\theta > \varepsilon/4) \rightarrow 0$ as $h \rightarrow 0$, and $\mathbb{P}_y(\sigma > \delta) \rightarrow 0$ for $h \rightarrow 0$ for all $y \in \mathbb{R}^d$. The first one is very simple: since there are at most $1/h$ time steps in the unit interval $[0, 1]$, a simple union bound yields

$$\mathbb{P}_y(\theta > \varepsilon) \leq \frac{1}{h} \sup_x \Pi_h(x, B(x, \varepsilon)^c) \leq \sup_x \Delta_\varepsilon^h(x) \rightarrow 0$$

by (iii). The second one requires more arguments. We follow the elegant argument introduced by Stroock and Varadhan (Theorem 1.4.6). The first step is to estimate $\mathbb{P}_y(\tau_1 \leq u)$ for small u . Note that by Lemma 6.6, the process

$$f_{y,\varepsilon}(Y_k^h) + C_\varepsilon k, \quad k = 0, 1, \dots$$

is a submartingale. Thus letting $\tau = \tau_1$, and using the Optional stopping theorem at $\tau \wedge u$,

$$\mathbb{E}_y \left\{ f_{y,\varepsilon}(Y_{\tau \wedge u}^h) + C_\varepsilon(\tau \wedge u) \right\} \geq 1.$$

Since $\tau \wedge u \leq u$ and since on the event that $\tau \leq u$, we have that $|Y_{\tau \wedge u}^h - y| \geq \varepsilon$, so $f_{y,\varepsilon}(Y_{\tau \wedge u}^h) = 0$, we have:

$$\mathbb{P}_y(\tau \leq u) \leq \mathbb{E}_y \left\{ 1 - f_{y,\varepsilon}(Y_{\tau \wedge u}^h) \right\} \leq C_\varepsilon u.$$

This has the following consequence: for all $u > 0$, letting $p = \mathbb{P}_y(\tau \leq u)$:

$$\begin{aligned} \mathbb{E}_y(e^{-\tau}) &\leq \mathbb{P}(\tau \leq u) + e^{-u} \mathbb{P}_y(\tau > u) \\ &\leq p + e^{-u}(1 - p) \leq e^{-u} + p(1 - e^{-u}) \\ &\leq e^{-u} + pu + o(u) \leq 1 - u + pu + o(u) \\ &\leq 1 - u(1 - p) + o(u). \end{aligned}$$

Since $p \leq C_\varepsilon u$, this means $\mathbb{E}_y(e^{-\tau}) \leq \lambda$ where $\lambda < 1$ is independent of y and δ : indeed, one can choose u small enough that $\mathbb{E}_y(e^{-\tau}) \leq 1 - u/2 < 1$. Now, iterating and using the strong Markov property at times τ_1, \dots, τ_n , which are stopping times,

$$\mathbb{E}_y(e^{-\tau_n}) \leq \lambda^n$$

since λ does not depend on y , and thus by Markov's inequality:

$$\begin{aligned}\mathbb{P}_y(N > n) &= \mathbb{P}_y(\tau_n < 1) \leq \mathbb{P}_y(e^{-\tau_n} \geq e^{-1}) \\ &\leq e\mathbb{E}_y(e^{-\tau_n}) \leq e\lambda^n.\end{aligned}$$

We finish by saying that

$$\begin{aligned}\mathbb{P}_y(\sigma \leq \delta) &\leq k \sup_x \mathbb{P}_x(\tau \leq \delta) + \mathbb{P}_y(N > k) \\ &\leq C_\varepsilon k\delta + e\lambda^k.\end{aligned}$$

Thus we take k large enough that $e\lambda^k < \varepsilon/2$ and then pick δ small enough that $C_\varepsilon k\delta < \varepsilon/2$. We are then done for the proof of tightness.

Step 2. Uniqueness of the weak subsequential limits.

Since we have assumed that the martingale problem $\mathbf{M}(a, b)$ was well posed, it suffices to show that the limit of any weakly convergent subsequence solves the martingale problem $\mathbf{M}(a, b)$. Our first step for doing so is to show that the generator of the Markov chain L^h converges in a suitable sense to the generator L of the diffusion:

$$Lf(x) = \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x) \frac{\partial f}{\partial x_i}(x)$$

Lemma 6.8 *Let $f \in \mathcal{C}_K^2$ be twice differentiable and with compact support. Then $L^h f(x) \rightarrow Lf(x)$ uniformly over $x \in \mathbb{R}^d$ as $h \rightarrow 0$.*

Proof. Going back to our Taylor expansion of $L^h f(x)$, and recalling the definition of $b_i^h(x)$ and $a_{ij}(x)$, we may write:

$$\begin{aligned}L^h f(x) &= \sum_{i=1}^d b_i^h(x) D_i f(x) \\ &+ \int_{|y-x| \leq 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) K_h(x, dy) \\ &+ \int_{|y-x| > 1} [f(y) - f(x)] K_h(x, dy)\end{aligned}$$

The final term in the right-hand side converges to 0 uniformly in x by assumption (iii) with $\varepsilon = 1$. To deal with the first term, note that

$$\left| \sum_{i=1}^d b_i^h(x) D_i f(x) - \sum_{i=1}^d b_i(x) D_i f(x) \right| \leq \sup_{1 \leq i \leq d} |b_i^h(x) - b_i(x)| \sum_{i=1}^d \|D_i f\|_\infty$$

which converges to 0 uniformly in x by assumption (ii) (since $f \in \mathcal{C}_K^2$). It remains to

deal with the central term. Recalling the definition of $a_{ij}^h(x)$, we get:

$$\begin{aligned} & \left| \int_{|y-x|\leq 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) D_{ij}f(z_{xy}) K_h(x, dy) - \sum_{i,j=1}^d a_{ij}(x) D_{ij}f(x) \right| \\ & \leq \left| \sum_{i,j=1}^d a_{ij}^h(x) D_{ij}f(x) - \sum_{i,j=1}^d a_{ij}(x) D_{ij}f(x) \right| \\ & \quad + \left| \int_{|y-x|\leq 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) [D_{ij}f(z_{xy}) - D_{ij}f(x)] K_h(x, dy) \right|. \end{aligned}$$

The first term converges to 0 uniformly in x by (i) and the fact that the derivatives of f are uniformly bounded. The second term can be split in an integral over $|y - x| > \varepsilon$ and $|y - x| \leq \varepsilon$. The first one converges to 0 uniformly in $x \in \mathbb{R}^d$ thanks to (iii) and the fact that the integrand is bounded. For the other term, let

$$\Gamma(\varepsilon) = \sup_{1 \leq i, j \leq d} \sup_{|y-x| \leq \varepsilon} |D_{ij}f(z_{xy}) - D_{ij}f(x)|.$$

Then since z_{xy} lies on the segment between x and y , and since $D_{ij}f$ is continuous on the compact set K (and hence uniformly continuous), $\Gamma(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. On the other hand,

$$\begin{aligned} & \left| \int_{|y-x|\leq 1} \sum_{i,j=1}^d (y_i - x_i)(y_j - x_j) [D_{ij}f(z_{xy}) - D_{ij}f(x)] K_h(x, dy) \right| \\ & \leq \Gamma(\varepsilon) \int_{|y-x|\leq \varepsilon} |y - x|^2 K_h(x, dy) \text{ by Cauchy-Schwarz's inequality,} \end{aligned}$$

so the proof of the lemma is complete. \square

We now use this lemma to conclude the proof of Theorem 6.5. Fix $h_n \rightarrow 0$ such that $X^{h_n} \rightarrow X$ weakly (in D) as $n \rightarrow \infty$. (Recall that X^h is defined as $X_t^h = Y_{h\lfloor t/h \rfloor}^h$.) Fix $s < t$. Then for any continuous functional $F : D \rightarrow \mathbb{R}$ which is measurable with respect to \mathcal{F}_s , we have, since L^h is the discrete generator of Y ,

$$f(X_{kh_n}^{h_n}) - \sum_{j=0}^{k-1} h_n L^{h_n} f(X_{jh_n}^{h_n}), \quad k = 0, 1, \dots$$

is a martingale. In particular, taking $k = k_n$ such that $kh_n > s$, i.e., $k_n = \lceil s/h_n \rceil$, and taking ℓ_n similarly so that $\ell_n h_n > t$, i.e., $\ell_n = \lceil t/h_n \rceil$, we get

$$\mathbb{E}_x \left(F(X^{h_n}) \left\{ f(X_{\ell_n h_n}^{h_n}) - f(X_{k_n h_n}^{h_n}) - \sum_{j=k_n}^{\ell_n-1} h_n L^{h_n} f(X_{jh_n}^{h_n}) \right\} \right) = 0.$$

By using the Skorokhod representation theorem, one may find Y^n such that $Y^n \stackrel{d}{=} X^{h_n}$ and $Y^n \rightarrow Y$ almost surely, where $Y \stackrel{d}{=} X$. We recognize a Riemann sum in this

expectation. Since almost sure convergence in D implies almost sure convergence of the marginals, we conclude by the Lebesgue convergence theorem that

$$\mathbb{E}_x \left(F(X) \left\{ f(X_t) - f(X_s) - \int_s^t Lf(X_u) du \right\} \right) = 0.$$

Since F is an arbitrary continuous function on D , it follows that

$$f(X_t) - \int_0^t Lf(X_u) du, \quad t \geq 0$$

is a martingale for all $f \in \mathcal{C}_K^2$. Since the martingale problem has a unique solution, the desired conclusion follows. This ends the proof of Theorem 6.5. \square

6.4 Example.

This result has literally thousands of practical applications, and we show one particularly simple such application.

The Ehrenfest chain. This is a Markov chain which models a box filled with gas molecules which is divided in two equal pieces, and where gas molecules can be exchanged between the two pieces through a small hole. Mathematically, we have two urns with a total of $2n$ balls (molecules). At each time step we pick one ball uniformly at random among the $2n$ balls of the urn, and move it to the other urn (we think of this event as a certain gas molecule going through that hole). Let Y_t^n denote the number of molecules in the left urn.

Define a normalized process $X_t^n = (Y_{\lfloor tn \rfloor}^n - n)/\sqrt{n}$, and assume for instance that $Y_0^n = n$, i.e., equal number of molecules in each urn.

Theorem 6.9 *The process $(X_t^n, 0 \leq t \leq 1)$ converges weakly to an Ornstein-Uhlenbeck diffusion $(X_t, 0 \leq t \leq 1)$ with unit viscosity, i.e., the pathwise unique solution to*

$$dX_t = -X_t dt + dB_t, \quad X_0 = 0.$$

Thus the number of molecules in each urn never deviates too much from n . Writing $K^n(x, dy) = n\Pi^n(x, dy)$,

Proof. The state space for Y^n is $S_n = \{k/\sqrt{n} : -n \leq k \leq n\}$. The transition probability Π^n of Y^n is given

$$\Pi^n(x, x + n^{-1/2}) = \frac{n - x\sqrt{n}}{2n}, \quad \Pi^n(x, x - n^{-1/2}) = \frac{n + x\sqrt{n}}{2n}$$

Here $d = 1$, and the expected infinitesimal drift

$$\hat{b}^n(x) = \int (y - x) K^n(x, dy) = n \left\{ n^{-1/2} \frac{n - x\sqrt{n}}{2n} - n^{-1/2} \frac{n + x\sqrt{n}}{2n} \right\} = -x,$$

while the infinitesimal variance

$$\hat{a}^n(x) = \int (y - x)^2 K^n(x, dy) = n \left\{ n^{-1} \frac{n - x\sqrt{n}}{2n} + n^{-1} \frac{n + x\sqrt{n}}{2n} \right\} = 1.$$

It follows without difficulty that the truncated expected drift and variance, respectively $b^n(x) = \int_{|y-x|\leq 1} (y-x)^2 K^n(x, dy)$ and $a^n(x) = \int_{|y-x|\leq 1} (y-x)^2 K^n(x, dy)$, satisfy:

$$a^n(x) \rightarrow 1; b^n(x) \rightarrow -x$$

uniformly on every compact set. Since the coefficients of the Ornstein-Uhlenbeck diffusion are Lipschitz, there is pathwise uniqueness for the associated SDE and thus uniqueness in distribution. Therefore, $(X_t^n, 0 \leq t \leq 1)$ converges to $(X_t, 0 \leq t \leq 1)$ weakly, by Theorem 6.5.

This finishes the lecture notes for this course. I hope you've enjoyed it!

References

- [1] R. Durrett. *Stochastic calculus: a practical introduction*. CRC Press, Probability and Stochastics Series, 1996.
- [2] D. Revuz and M. Yor (1999). *Continuous martingales and Brownian motion*. 3rd edition. Springer, Grundlehren der mathematischen Wissenschaften, Vol. 293.
- [3] L.C.G. Rogers and D. Williams (1987). *Diffusions, Markov Processes and Martingales*, Volume 2. 2nd edition. Cambridge University Press.
- [4] D.W. Stroock and S.R.S. Varadhan (1997). *Multidimensional Diffusion Processes*. Springer, Classics in Mathematics. Reprint of Grundlehren der mathematischen Wissenschaften, Vol. 233.