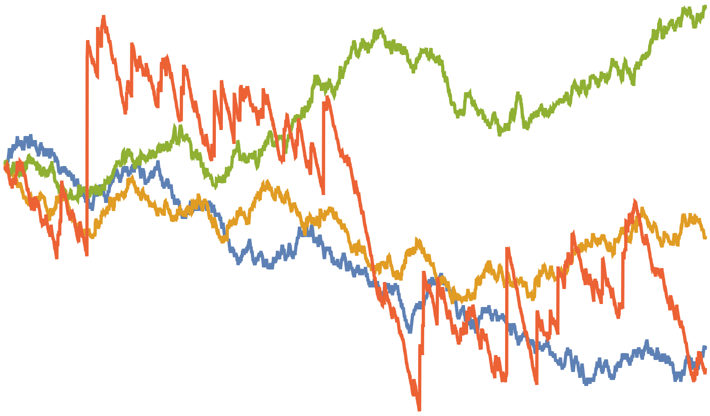
Stochastic Analysis



Andreas Eberle

February 20, 2016

Contents

Chapter 1

Brownian Motion

This introduction to stochastic analysis starts with an introduction to Brownian motion. Brownian Motion is a diffusion process, i.e. a continuous-time Markov process $(B_t)_{t\geq 0}$ with continuous sample paths $t \mapsto B_t(\omega)$. In fact, it is the only nontrivial continuous-time process that is a Lévy process as well as a martingale and a Gaussian process. A rigorous construction of this process has been carried out first by N. Wiener in 1923. Already about 20 years earlier, related models had been introduced independently for financial markets by L. Bachelier [*Théorie de la spéculation*, Ann. Sci. École Norm. Sup. 17, 1900], and for the velocity of molecular motion by A. Einstein [*Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Annalen der Physik 17, 1905*].
It has been a groundbreaking approach of K. Itô to construct general diffusion processes from Brownian motion, cf. [...]. In classical analysis, the solution of an ordinary differential equation $x'(t) = f(t, x(t))$ is a function, that can be approximated locally for t close to t_0 by the linear function $x(t_0) + f(t_0, x(t_0)) \cdot (t - t_0)$. Similarly, Itô showed, that a diffusion process behaves locally like a linear function of Brownian motion – the connection being described rigorously by a stochastic differential equation (SDE).

The fundamental rôle played by Brownian motion in stochastic analysis is due to the central limit Theorem. Similarly as the normal distribution arises as a universal scaling limit of standardized sums of independent, identically distributed, square integrable random variables, Brownian motion shows up as a universal scaling limit of Random Walks with square integrable increments.

1.1 From Random Walks to Brownian Motion

To motivate the definition of Brownian motion below, we first briefly discuss discrete-time stochastic processes and possible continuous-time scaling limits on an informal level.

A standard approach to model stochastic dynamics in discrete time is to start from a sequence of random variables $\eta_1, \eta_2, \ldots$ defined on a common probability space $(\Omega, \mathcal{A}, P)$. The random variables η_n describe the stochastic influences (*noise*) on the system. Often they are assumed to be *independent and identically distributed (i.i.d.)*. In this case the collection (η_n) is also called a **white noise**, whereas a **colored noise** is given by dependent random variables. A stochastic process $X_n, n = 0, 1, 2, \ldots$, taking values in $\mathbb{R}^d$ is then defined recursively on $(\Omega, \mathcal{A}, P)$ by

$$X_{n+1} \quad = \quad X_n + \Phi_{n+1}(X_n, \eta_{n+1}), \qquad n = 0, 1, 2, \ldots. \tag{1.1.1}$$

Here the Φ_n are measurable maps describing the *random law of motion*. If X_0 and $\eta_1, \eta_2, \ldots$ are independent random variables, then the process (X_n) is a Markov chain with respect to P.

Now let us assume that the random variables η_n are independent and identically distributed taking values in $\mathbb{R}$, or, more generally, $\mathbb{R}^d$. The easiest type of a nontrivial stochastic dynamics as described above is the Random Walk $S_n = \sum_{i=1}^{n} \eta_i$ which satisfies

$$S_{n+1} \quad = \quad S_n + \eta_{n+1} \qquad \text{for } n = 0, 1, 2, \ldots.$$

Since the noise random variables η_n are the increments of the Random Walk (S_n), the law of motion (1.1.1) in the general case can be rewritten as

$$X_{n+1} - X_n \quad = \quad \Phi_{n+1}(X_n, S_{n+1} - S_n), \qquad n = 0, 1, 2, \ldots. \tag{1.1.2}$$

This equation is a difference equation for (X_n) driven by the stochastic process (S_n).

Our aim is to carry out a similar construction as above for stochastic dynamics in continuous time. The stochastic difference equation (1.1.2) will then eventually be replaced by a *stochastic differential equation (SDE)*. However, before even being able to think about how to write down and make sense of such an equation, we have to identify a continuous-time stochastic process that takes over the rôle of the Random Walk. For this purpose, we first determine possible scaling limits of Random Walks when the time steps tend to 0. It will turn out that if the increments are square integrable and the size of the increments goes to 0 as the length of the time steps tends to 0, then by the Central Limit Theorem there is essentially only one possible limit process in continuous time: Brownian motion.

Central Limit Theorem

Suppose that $Y_{n,i} : \Omega \to \mathbb{R}^d, 1 \leq i \leq n < \infty$, are identically distributed, square-integrable random variables on a probability space $(\Omega, \mathcal{A}, P)$ such that $Y_{n,1}, \ldots, Y_{n,n}$ are independent for each $n \in \mathbb{N}$. Then the rescaled sums

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{n,i} - E[Y_{n,i}])$$

converge in distribution to a multivariate normal distribution $N(0, C)$ with covariance matrix

$$C_{kl} \quad = \quad \mathrm{Cov}[Y_{n,i}^{(k)}, Y_{n,i}^{(l)}].$$

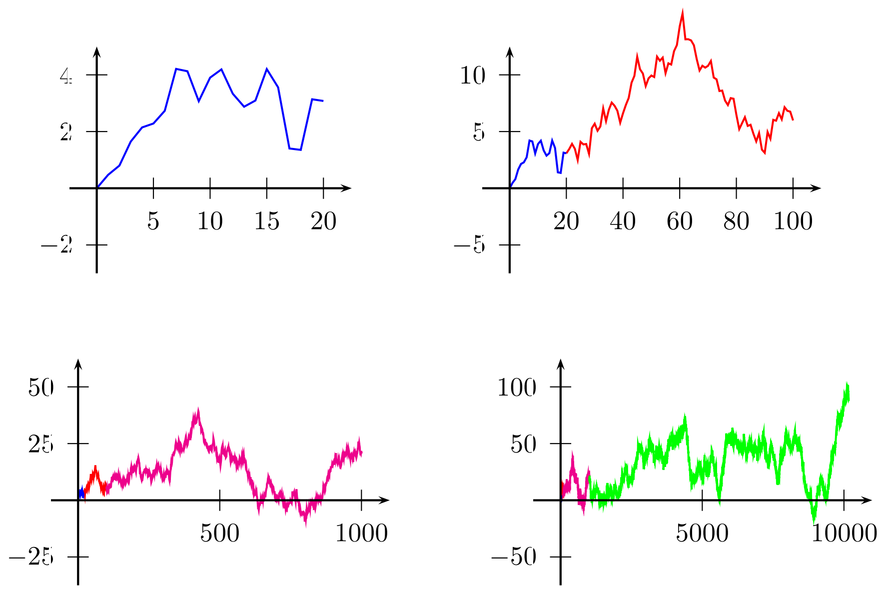
To see, how the CLT determines the possible scaling limits of Random Walks, let us consider a one-dimensional Random Walk

$$S_n = \sum_{i=1}^{n} \eta_i, \qquad n = 0, 1, 2, \ldots,$$

on a probability space $(\Omega, \mathcal{A}, P)$ with independent increments $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ normalized such that

$$E[\eta_i] \quad = \quad 0 \qquad \text{and} \qquad \mathrm{Var}[\eta_i] \quad = \quad 1. \tag{1.1.3}$$

Plotting many steps of the Random Walk seems to indicate that there is a limit process with continuous sample paths after appropriate rescaling:



To see what appropriate means, we fix a positive integer m, and try to define a rescaled Random Walk $S_t^{(m)}$ $(t = 0, 1/m, 2/m, \ldots)$ with time steps of size $1/m$ by

$$S_{k/m}^{(m)} \quad = \quad c_m \cdot S_k \qquad (k = 0, 1, 2, \ldots)$$

for some constants $c_m > 0$. If t is a multiple of $1/m$, then

$$\mathrm{Var}[S_t^{(m)}] \quad = \quad c_m^2 \cdot \mathrm{Var}[S_{mt}] \quad = \quad c_m^2 \cdot m \cdot t.$$

Hence in order to achieve convergence of $S_t^{(m)}$ as $m \to \infty$, we should choose c_m proportional to $m^{-1/2}$. This leads us to define a continuous time process $(S_t^{(m)})_{t \geq 0}$ by

$$S_t^{(m)}(\omega) \quad := \quad \frac{1}{\sqrt{m}} S_{mt}(\omega) \qquad \text{whenever } t = k/m \text{ for some integer } k,$$

and by linear interpolation for $t \in \left(\frac{k-1}{m}, \frac{k}{m}\right]$.

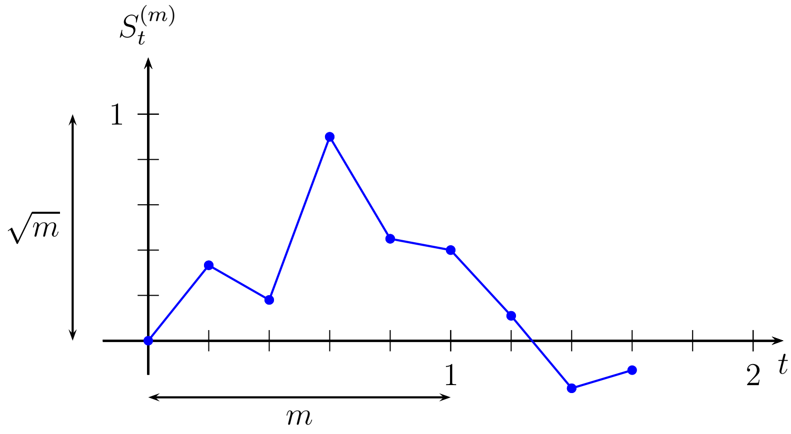


Figure 1.1: Rescaling of a Random Walk.

Clearly,

$$E[S_t^{(m)}] \quad = \quad 0 \qquad \text{for all } t \geq 0,$$

and

$$\mathrm{Var}[S_t^{(m)}] \quad = \quad \frac{1}{m}\mathrm{Var}[S_{mt}] \quad = \quad t$$

whenever t is a multiple of $1/m$. In particular, the expectation values and variances for a fixed time t do not depend on m. Moreover, if we fix a partition $0 \leq t_0 < t_1 < \ldots < t_n$ such that each t_i is a multiple of $1/m$, then the increments

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} \quad = \quad \frac{1}{\sqrt{m}}\left(S_{mt_{i+1}} - S_{mt_i}\right), \qquad i = 0, 1, 2, \ldots, n-1, \tag{1.1.4}$$

of the rescaled process $(S_t^{(m)})_{t\geq 0}$ are independent centered random variables with variances $t_{i+1} - t_i$. If t_i is not a multiple of $1/m$, then a corresponding statement holds approximately with an error that should be negligible in the limit $m \to \infty$. Hence, if the rescaled Random Walks $(S_t^{(m)})_{t\geq 0}$ converge in distribution to a limit process $(B_t)_{t\geq 0}$, then $(B_t)_{t\geq 0}$ should have *independent increments* $B_{t_{i+1}} - B_{t_i}$ *over disjoint time intervals with mean* 0 *and variances* $t_{i+1} - t_i$.

It remains to determine the precise distributions of the increments. Here the Central Limit Theorem applies. In fact, we can observe that by (1.1.4) each increment

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} \quad = \quad \frac{1}{\sqrt{m}} \sum_{k=mt_i+1}^{mt_{i+1}} \eta_k$$

of the rescaled process is a rescaled sum of $m \cdot (t_{i+1} - t_i)$ i.i.d. random variables with mean 0 and variance 1. Therefore, the CLT implies that the distributions of the increments converge weakly to a normal distribution:

$$S^{(m)}_{t_{i+1}} - S^{(m)}_{t_i} \quad \xrightarrow{\mathcal{D}} \quad N(0, t_{i+1} - t_i).$$

Hence if a limit process (B_t) exists, then it should have *independent, normally distributed increments*.

Our considerations motivate the following definition:

Definition (Brownian Motion).

(1). Let $a \in \mathbb{R}$. A continuous-time stochastic process $B_t : \Omega \to \mathbb{R}$, $t \geq 0$, defined on a probability space $(\Omega, \mathcal{A}, P)$, is called a ***Brownian motion (starting in** a**)*** *if and only if*

(a) $B_0(\omega) \; = \; a$ for each $\omega \in \Omega$.

(b) For any partition $0 \leq t_0 < t_1 < \ldots < t_n$, the increments $B_{t_{i+1}} - B_{t_i}$ are independent random variables with distribution

$$B_{t_{i+1}} - B_{t_i} \quad \sim \quad N(0, t_{i+1} - t_i).$$

(c) P-almost every sample path $t \mapsto B_t(\omega)$ is continuous.

(2). An $\mathbb{R}^d$-valued stochastic process $B_t(\omega) = (B^{(1)}_t(\omega), \ldots, B^{(d)}_t(\omega))$ is called a multi-dimensional Brownian motion if and only if the component processes $(B^{(1)}_t), \ldots, (B^{(d)}_t)$ are independent one-dimensional Brownian motions.

Thus the increments of a d-dimensional Brownian motion are independent over disjoint time intervals and have a multivariate normal distribution:

$$B_t - B_s \quad \sim \quad N(0, (t-s) \cdot I_d) \qquad \text{for any } 0 \leq s \leq t.$$

Remark. (1). *Continuity:* Continuity of the sample paths has to be assumed separately: If $(B_t)_{t\geq 0}$ is a one-dimensional Brownian motion, then the modified process $(\widetilde{B}_t)_{t\geq 0}$ defined by $\widetilde{B}_0 = B_0$ and

$$\widetilde{B}_t = B_t \cdot I_{\{B_t \in \mathbb{R}\setminus\mathbb{Q}\}} \qquad \text{for } t > 0$$

has almost surely discontinuous paths. On the other hand, it satisfies (a) and (b) since the distributions of $(\widetilde{B}_{t_1}, \ldots, \widetilde{B}_{t_n})$ and $(B_{t_1}, \ldots, B_{t_n})$ coincide for all $n \in \mathbb{N}$ and $t_1, \ldots, t_n \geq 0$.

(2). *Spatial Homogeneity:* If $(B_t)_{t\geq 0}$ is a Brownian motion starting at 0, then the translated process $(a + B_t)_{t\geq 0}$ is a Brownian motion starting at a.

(3). *Existence:* There are several constructions and existence proofs for Brownian motion. In Section 1.3 below we will discuss in detail the Wiener-Lévy construction of Brownian motion as a random superposition of infinitely many deterministic paths. This explicit construction is also very useful for numerical approximations. A more general (but less constructive) existence proof is based on Kolmogorov's extension Theorem, cf. e.g. [Klenke].

(4). *Functional Central Limit Theorem:* The construction of Brownian motion as a scaling limit of Random Walks sketched above can also be made rigorous. *Donsker's invariance principle* is a functional version of the central limit Theorem which states that the rescaled Random Walks $(S_t^{(m)})$ converge in distribution to a Brownian motion. As in the classical CLT the limit is universal, i.e., it does not depend on the distribution of the increments η_i provided (1.1.3) holds, cf. Section **??**.

Brownian motion as a Lévy process.

The definition of Brownian motion shows in particular that Brownian motion is a *Lévy process*, i.e., it has stationary independent increments (over disjoint time intervals). In fact, the analogues of Lévy processes in discrete time are Random Walks, and it is rather obvious, that all scaling limits of Random Walks should be Lévy processes. Brownian motion is the only Lévy process L_t in continuous time with paths such that $E[L_1] = 0$ and $\mathrm{Var}[L_1] = 1$. The normal distribution of the increments follows under these assumptions by an extension of the CLT, cf. e.g. [Breiman: Probability]. A simple example of a Lévy process with non-continuous paths is the Poisson process. Other examples are α-stable processes which arise as scaling limits of Random Walks when the increments are not square-integrable. Stochastic analysis based on general Lévy processes has attracted a lot of interest recently.

Let us now consider consider a Brownian motion $(B_t)_{t\geq 0}$ starting at a fixed point $a \in \mathbb{R}^d$, defined on a probability space $(\Omega, \mathcal{A}, P)$. The information on the process up to time t is encoded in the σ-algebra

$$\mathcal{F}_t^B \quad = \quad \sigma(B_s \,|\, 0 \leq s \leq t)$$

generated by the process. The independence of the increments over disjoint intervals immediately implies:

Lemma 1.1. *For any $0 \leq s \leq t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s^B$.*

Proof. For any partition $0 = t_0 \leq t_1 \leq \ldots \leq t_n = s$ of the interval $[0, s]$, the increment $B_t - B_s$ is independent of the σ-algebra

$$\sigma(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$$

generated by the increments up to time s. Since

$$B_{t_k} \quad = \quad B_{t_0} + \sum_{i=1}^{k} (B_{t_i} - B_{t_{i-1}})$$

and B_{t_0} is constant, this σ-algebra coincides with $\sigma(B_{t_0}, B_{t_1}, \ldots, B_{t_n})$. Hence $B_t - B_s$ is independent of all finite subcollections of $(B_u \,|\, 0 \leq u \leq s)$ and therefore independent of $\mathcal{F}_s^B$. □

Brownian motion as a Markov process.

As a process with stationary increments, Brownian motion is in particular a time-homogeneous Markov process. In fact, we have:

Theorem 1.2 (Markov property). *A Brownian motion $(B_t)_{t\ge 0}$ in $\mathbb{R}^d$ is a time-homogeneous Markov process with transition densities*

$$p_t(x,y) \quad = \quad (2\pi t)^{-d/2} \cdot \exp\left(-\frac{|x-y|^2}{2t}\right), \qquad t>0, \quad x,y\in\mathbb{R}^d,$$

i.e., for any Borel set $A \subseteq \mathbb{R}^d$ and $0 \le s < t$,

$$P[B_t \in a \mid \mathcal{F}_s^B] \quad = \quad \int_A p_{t-s}(B_s,y)\,dy \qquad P\text{-almost surely.}$$

Proof. For $0 \le s < t$ we have $B_t = B_s + (B_t - B_s)$ where B_s is $\mathcal{F}_s^B$-measurable, and $B_t - B_s$ is independent of $\mathcal{F}_s^B$ by Lemma 1.1. Hence

$$\begin{aligned} P[B_t \in A \mid \mathcal{F}_s^B](\omega) \quad &= \quad P[B_s(\omega) + B_t - B_s \in A] \quad = \quad N(B_s(\omega),(t-s)\cdot I_d)[A] \\ = \quad &\int_A (2\pi(t-s))^{-d/2}\cdot \exp\left(-\frac{|y-B_s(\omega)|^2}{2(t-s)}\right)\,dy \qquad P\text{-almost surely.} \end{aligned}$$

□

Remark (Heat equation as backward equation and forward equation). The transition function of Brownian motion is the *heat kernel* in $\mathbb{R}^d$, i.e., it is the fundamental solution of the heat equation

$$\frac{\partial u}{\partial t} \quad = \quad \frac{1}{2}\Delta u.$$

More precisely, $p_t(x,y)$ solves the initial value problem

$$\begin{aligned} \frac{\partial}{\partial t}p_t(x,y) \ &= \ \frac{1}{2}\Delta_x p_t(x,y) \qquad &&\text{for any } t>0, x,y\in\mathbb{R}^d, \\ \lim_{t\searrow 0}\int p_t(x,y)f(y)\,dy \ &= \ f(x) &&\text{for any } f\in C_b(\mathbb{R}^d), x\in\mathbb{R}^d, \end{aligned} \tag{1.1.5}$$

where $\Delta_x = \sum\limits_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ denotes the action of the Laplace operator on the x-variable. The equation (1.1.5) can be viewed as a version of *Kolmogorov's backward equation* for

Brownian motion as a time-homogeneous Markov process, which states that for each $t > 0, y \in \mathbb{R}^d$ and $f \in C_b(\mathbb{R}^d)$, the function

$$v(s,x) \quad = \quad \int p_{t-s}(x,y) f(y)\, dy$$

solves the terminal value problem

$$\frac{\partial v}{\partial s}(s,x) \quad = \quad -\frac{1}{2}\Delta_x v(s,x) \qquad \text{for } s \in [0,t), \qquad \lim_{s \nearrow t} v(s,x) \;=\; f(x). \tag{1.1.6}$$

Note that by the Markov property, $v(s,x) = (p_{t-s}f)(x)$ is a version of the conditional expectation $E[f(B_t) \mid B_s = x]$. Therefore, the backward equation describes the dependence of the expectation value on starting point and time.

By symmetry, $p_t(x,y)$ also solves the initial value problem

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x,y) \;&=\; \frac{1}{2}\Delta_y p_t(x,y) && \text{for any } t > 0, \quad \text{and} \quad x, y \in \mathbb{R}^d, \\ \lim_{t \searrow 0} \int g(x) p_t(x,y)\, dx \;&=\; g(y) && \text{for any } g \in C_b(\mathbb{R}^d), y \in \mathbb{R}^d. \end{aligned} \tag{1.1.7}$$

The equation (1.1.7) is a version of *Kolmogorov's forward equation*, stating that for $g \in C_b(\mathbb{R}^d)$, the function $u(t,y) = \int g(x) p_t(x,y)\, dx$ solves

$$\frac{\partial u}{\partial t}(t,y) \quad = \quad \frac{1}{2}\Delta_y u(t,y) \qquad \text{for } t > 0, \qquad \lim_{t \searrow 0} u(t,y) \;=\; g(y). \tag{1.1.8}$$

The forward equation describes the forward time evolution of the transition densities $p_t(x,y)$ for a given starting point x.

The Markov property enables us to compute the marginal distributions of Brownian motion:

Corollary 1.3 (Finite dimensional marginals). *Suppose that $(B_t)_{t \geq 0}$ is a Brownian motion starting at $x_0 \in \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{A}, P)$. Then for any*

$n \in \mathbb{N}$ and $0 = t_0 < t_1 < t_2 < \ldots < t_n$, the joint distribution of $B_{t_1}, B_{t_2}, \ldots, B_{t_n}$ is absolutely continuous with density

$$\begin{aligned} f_{B_{t_1},\ldots,B_{t_n}}(x_1,\ldots,x_n) &= p_{t_1}(x_0,x_1)p_{t_2-t_1}(x_1,x_2)p_{t_3-t_2}(x_2,x_3)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n) \\ &= \prod_{i=1}^n (2\pi(t_i-t_{i-1}))^{-d/2}\cdot\exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{|x_i-x_{i-1}|^2}{t_i-t_{i-1}}\right) \quad (1.1.9) \end{aligned}$$

Proof. By the Markov property and induction on n, we obtain

$$\begin{aligned} &P[B_{t_1}\in A_1,\ldots,B_{t_n}\in A_n] \\ &= E[P[B_{t_n}\in A_n \mid \mathcal{F}^B_{t_{n-1}}]\ ;\ B_{t_1}\in A_1,\ldots,B_{t_{n-1}}\in A_{n-1}] \\ &= E[p_{t_n-t_{n-1}}(B_{t_{n-1}},A_n)\ ;\ B_{t_1}\in A_1,\ldots,B_{t_{n-1}}\in A_{n-1}] \\ &= \int\limits_{A_1}\cdots\int\limits_{A_{n-1}} p_{t_1}(x_0,x_1)p_{t_2-t_1}(x_1,x_2)\cdots \\ &\quad \cdot p_{t_{n-1}-t_{n-2}}(x_{n-2},x_{n-1})p_{t_n-t_{n-1}}(x_{n-1},A_n)\,dx_{n-1}\cdots dx_1 \\ &= \int\limits_{A_1}\cdots\int\limits_{A_n}\left(\prod_{i=1}^n p_{t_i-t_{i-1}}(x_{n-1},x_n)\right)\,dx_n\cdots dx_1 \end{aligned}$$

for all $n \geq 0$ and $A_1,\ldots,A_n \in \mathcal{B}(\mathbb{R}^d)$. □

Remark (Brownian motion as a Gaussian process). The corollary shows in particular that Brownian motion is a Gaussian process, i.e., all the marginal distributions in (1.1.9) are multivariate normal distributions. We will come back to this important aspect in the next section.

Wiener Measure

The distribution of Brownian motion could be considered as a probability measure on the product space $(\mathbb{R}^d)^{[0,\infty)}$ consisting of all maps $x : [0,\infty) \to \mathbb{R}^d$. A disadvantage of this approach is that the product space is far too large for our purposes: It contains extremely irregular paths $x(t)$, although at least almost every path of Brownian motion is continuous by definition. Actually, since $[0,\infty)$ is uncountable, the subset of all continuous paths is not even measurable w.r.t. the product σ-algebra on $(\mathbb{R}^d)^{[0,\infty)}$.

Instead of the product space, we will directly consider the distribution of Brownian motion on the continuous path space $C([0,\infty),\mathbb{R}^d)$. For this purpose, we fix a Brownian motion $(B_t)_{t\geq 0}$ starting at $x_0 \in \mathbb{R}^d$ on a probability space $(\Omega, \mathcal{A}, P)$, and we **assume** that **every** sample path $t \mapsto B_t(\omega)$ is continuous. This assumption can always be fulfilled by modifying a given Brownian motion on a set of measure zero. The full process $(B_t)_{t\geq 0}$ can then be interpreted as a single path-space valued random variable (or a *"random path"*).

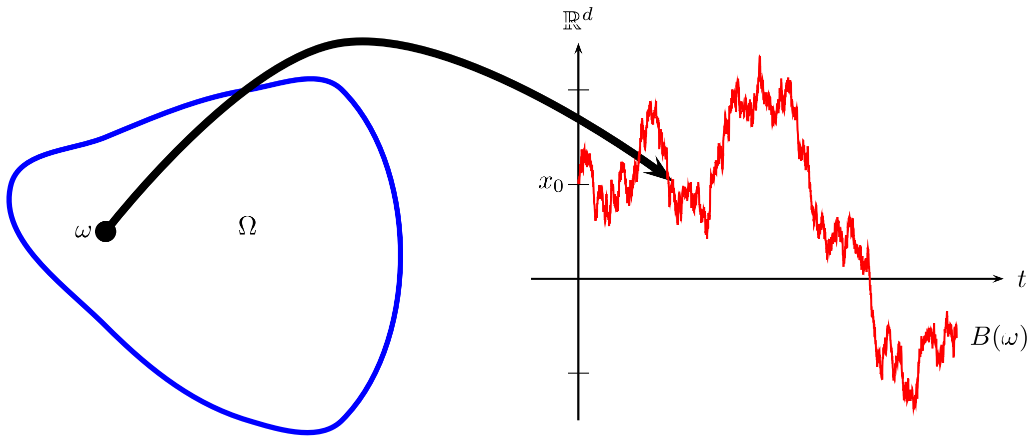


Figure 1.2: $B : \Omega \to C([0,\infty),\mathbb{R}^d), \quad B(\omega) \;=\; (B_t(\omega))_{t\geq 0}$.

We endow the space of continuous paths $x : [0,\infty) \to \mathbb{R}^d$ with the σ-algebra

$$\mathcal{B} \quad = \quad \sigma(X_t \,|\, t \geq 0)$$

generated by the coordinate maps

$$X_t \;:\; C([0,\infty),\mathbb{R}^d) \;\to\; \mathbb{R}^d, \quad X_t(x) = x_t, \qquad t \geq 0.$$

Note that we also have

$$\mathcal{B} \quad = \quad \sigma(X_t \,|\, t \in \mathcal{D})$$

for any dense subset $\mathcal{D}$ of $[0,\infty)$, because $X_t = \lim_{s\to t} X_s$ for each $t \in [0,\infty)$ by continuity. Furthermore, it can be shown that $\mathcal{B}$ is the Borel σ-algebra on $C([0,\infty),\mathbb{R}^d)$ endowed with the topology of uniform convergence on finite intervals.

Theorem 1.4 (Distribution of Brownian motion on path space). *The map $B : \Omega \to C([0,\infty), \mathbb{R}^d)$ is measurable w.r.t. the σ-algebras $\mathcal{A}/\mathcal{B}$. The distribution $P \circ B^{-1}$ of B is the unique probability measure μ_{x_0} on $(C([0,\infty), \mathbb{R}^d), \mathcal{B})$ with marginals*

$$\begin{aligned} &\mu_{x_0}\left[\{x \in C([0,\infty), \mathbb{R}^d) : x_{t_1} \in A_1, \ldots, x_{t_n} \in A_n\}\right] \qquad (1.1.10)\\ &\quad = \prod_{i=1}^{n} (2\pi(t_i - t_{i-1}))^{-d/2} \int_{A_1} \cdots \int_{A_n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) dx_n \cdots dx_1 \end{aligned}$$

for any $n \in \mathbb{N}$, $0 < t_1 < \ldots < t_n$, and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$.

Definition. *The probability measure μ_{x_0} on the path space $C([0,\infty), \mathbb{R}^d)$ determined by (1.1.10) is called* ***Wiener measure*** *(with start in x_0).*

Remark (Uniqueness in distribution). The Theorem asserts that the path space distribution of a Brownian motion starting at a given point x_0 is the corresponding Wiener measure. In particular, it is uniquely determined by the marginal distributions in (1.1.9).

Proof of Theorem 1.4. For $n \in \mathbb{N}, 0 < t_1 < \ldots < t_n$, and $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\begin{aligned} B^{-1}(\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}) &= \{\omega : X_{t_1}(B(\omega)) \in A_1, \ldots, X_{t_n}(B(\omega)) \in A_n\} \\ &= \{B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n\} \quad \in \quad \mathcal{A}. \end{aligned}$$

Since the cylinder sets of type $\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}$ generate the σ-algebra $\mathcal{B}$, the map B is $\mathcal{A}/\mathcal{B}$-measurable. Moreover, by corollary 1.3, the probabilities

$$P[B \in \{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}] \quad = \quad P[B_{t_1} \in A_1, \ldots, B_{t_n} \in A_n],$$

are given by the right hand side of (1.1.10). Finally, the measure μ_{x_0} is uniquely determined by (1.1.10), since the system of cylinder sets as above is stable under intersections and generates the σ-algebra $\mathcal{B}$. $\square$

Definition (Canonical model for Brownian motion.). *By (1.1.10), the coordinate process*

$$X_t(x) \quad = \quad x_t, \qquad t \geq 0,$$

on $C([0,\infty),\mathbb{R}^d)$ is a Brownian motion starting at x_0 w.r.t. Wiener measure μ_{x_0}. We refer to the stochastic process $(C([0,\infty),\mathbb{R}^d),\mathcal{B},\mu_{x_0},(X_t)_{t\geq 0})$ as the ***canonical model for Brownian motion starting at*** x_0.

1.2 Brownian Motion as a Gaussian Process

We have already verified that Brownian motion is a Gaussian process, i.e., the finite dimensional marginals are multivariate normal distributions. We will now exploit this fact more thoroughly.

Multivariate normals

Let us first recall some basics on normal random vectors:

Definition. *Suppose that $m \in \mathbb{R}^n$ is a vector and $C \in \mathbb{R}^{n\times n}$ is a symmetric non-negative definite matrix. A random variable $Y : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega,\mathcal{A},P)$ has a* ***multivariate normal distribution*** $\mathrm{N}(\mathrm{m},\mathrm{C})$ ***with mean*** m ***and covariance matrix*** C *if and only if its characteristic function is given by*

$$E[e^{ip\cdot Y}] \quad = \quad e^{ip\cdot m - \frac{1}{2}p\cdot Cp} \qquad \text{for any } p \in \mathbb{R}^n. \tag{1.2.1}$$

If C is non-degenerate, then a multivariate normal random variable Y is absolutely continuous with density

$$f_Y(x) \quad = \quad (2\pi \det C)^{-1/2} \exp\left(-\frac{1}{2}(x-m)\cdot C^{-1}(x-m)\right).$$

A degenerate normal distribution with vanishing covariance matrix is a Dirac measure:

$$N(m,0) \quad = \quad \delta_m.$$

Differentiating (1.2.1) w.r.t. p shows that for a random variable $Y \sim N(m,C)$, the mean vector is m and $C_{i,j}$ is the covariance of the components Y_i and Y_j. Moreover, the following important facts hold:

Theorem 1.5 (Properties of normal random vectors).

(1). A random variable $Y : \Omega \to \mathbb{R}^n$ has a multivariate normal distribution if and only if any linear combination

$$p \cdot Y \quad = \quad \sum_{i=1}^{n} p_i Y_i, \qquad p \in \mathbb{R}^n,$$

of the components Y_i has a one dimensional normal distribution.

(2). Any affine function of a normally distributed random vector Y is again normally distributed:

$$Y \sim N(m,C) \quad \Longrightarrow \quad AY+b \sim N(Am+b, ACA^\top)$$

for any $d \in \mathbb{N}$, $A \in \mathbb{R}^{d\times n}$ and $b \in \mathbb{R}^d$.

(3). If $Y = (Y_1, \dots, Y_n)$ has a multivariate normal distribution, and the components $Y_1, \dots, Y_n$ are uncorrelated random variables, then $Y_1, \dots, Y_n$ are independent.

Proof. (1). follows easily from the definition.

(2). For $Y \sim N(m,C)$, $A \in \mathbb{R}^{d\times n}$ and $b \in \mathbb{R}^d$ we have

$$\begin{aligned} E[e^{ip\cdot(AY+b)}] &= e^{ip\cdot b} E[e^{i(A^\top p)\cdot Y}] \\ &= e^{ip\cdot b} e^{i(A^\top p)\cdot m - \frac{1}{2}(A^\top p)\cdot CA^\top p} \\ &= e^{ip\cdot(Am+b) - \frac{1}{2} p \cdot ACA^\top} \qquad \text{for any } p \in \mathbb{R}^d, \end{aligned}$$

i.e., $AY+b \sim N(Am+b, ACA^\top)$.

(3). If $Y_1, \dots, Y_n$ are uncorrelated, then the covariance matrix $C_{i,j} = \mathrm{Cov}[Y_i, Y_j]$ is a diagonal matrix. Hence the characteristic function

$$E[e^{ip\cdot Y}] \quad = \quad e^{ip\cdot m - \frac{1}{2}p\cdot Cp} \quad = \quad \prod_{k=1}^{n} e^{im_k p_k - \frac{1}{2}C_{k,k}p_k^2}$$

is a product of characteristic functions of one-dimensional normal distributions. Since a probability measure on $\mathbb{R}^n$ is uniquely determined by its characteristic function, it follows that the adjoint distribution of $Y_1, \dots, Y_n$ is a product measure, i.e. $Y_1, \dots, Y_n$ are independent.

□

If Y has a multivariate normal distribution $N(m, C)$ then for any $p, q \in \mathbb{R}^n$, the random variables $p \cdot Y$ and $q \cdot Y$ are normally distributed with means $p \cdot m$ and $q \cdot m$, and covariance

$$\mathrm{Cov}[p \cdot Y, q \cdot Y] \quad = \quad \sum_{i,j=1}^{n} p_i C_{i,j} q_j \quad = \quad p \cdot Cq.$$

In particular, let $\{e_1, \dots, e_n\} \subseteq \mathbb{R}^n$ be an orthonormal basis consisting of eigenvectors of the covariance matrix C. Then the components $e_i \cdot Y$ of Y in this basis are uncorrelated and therefore independent, jointly normally distributed random variables with variances given by the corresponding eigenvectors λ_i:

$$\mathrm{Cov}[e_i \cdot Y, e_j \cdot Y] \quad = \quad \lambda_i \delta_{i,j}, \qquad 1 \le i, j \le n. \tag{1.2.2}$$

Correspondingly, the contour lines of the density of a non-degenerate multivariate normal distribution $N(m, C)$ are ellipsoids with center at m and principal axes of length $\sqrt{\lambda_i}$ given by the eigenvalues e_i of the covariance matrix C.

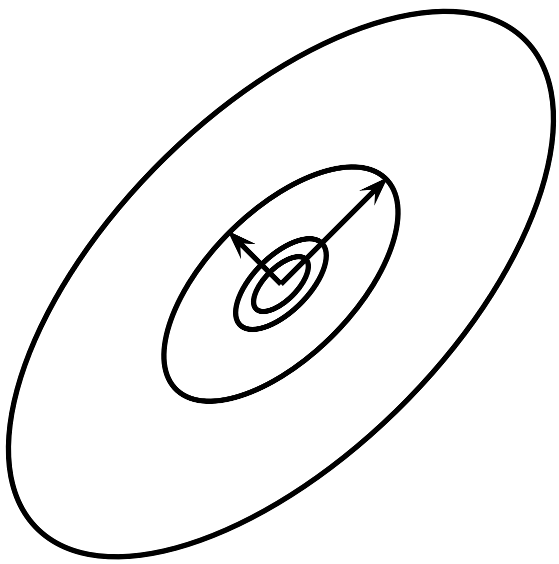


Figure 1.3: Level lines of the density of a normal random vector $Y \sim N\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}1 & 1\\-1 & 1\end{pmatrix}\right)$.

Conversely, we can generate a random vector Y with distribution $N(m, C)$ from i.i.d. standard normal random variables $Z_1, \dots, Z_n$ by setting

$$Y \quad = \quad m + \sum_{i=1}^{n} \sqrt{\lambda_i} Z_i e_i. \tag{1.2.3}$$

More generally, we have:

Corollary 1.6 (Generating normal random vectors). *Suppose that $C = U\Lambda U^\top$ with a matrix $U \in \mathbb{R}^{n\times d}$, $d \in \mathbb{N}$, and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d\times d}$ with nonnegative entries λ_i. If $Z = (Z_1, \dots, Z_d)$ is a random vector with i.i.d. standard normal random components $Z_1, \dots, Z_d$ then*

$$Y \quad = \quad U\Lambda^{1/2} Z + m$$

has distribution $N(m, C)$.

Proof. Since $Z \sim N(0, I_d)$, the second assertion of Theorem 1.5 implies

$$Y \sim N(m, U\Lambda U^\top).$$

□

Choosing for U the matrix $(e_1, \ldots, e_n)$ consisting of the orthonormal eigenvectors $e_1, \ldots, e_n$ of C, we obtain (1.2.3) as a special case of the corollary. For computational purposes it is often more convenient to use the Cholesky decomposition

$$C = LL^\top$$

of the covariance matrix as a product of a lower triangular matrix L and the upper triangular transpose $L^\top$:

Algorithm 1.7 (Simulation of multivariate normal random variables).

Given: $m \in \mathbb{R}^n, C \in \mathbb{R}^{n\times n}$ symmetric and non-negative definite.

Output: Sample $y \sim N(m, C)$.

(1). Compute the Cholesky decomposition $C = LL^\top$.

(2). Generate independent samples $z_1, \ldots, z_n \sim N(0, 1)$ (e.g. by the Box-Muller method).

(3). Set $y := Lz + m$.

Gaussian processes

Let I be an arbitrary index set, e.g. $I = \mathbb{N}, I = [0, \infty)$ or $I = \mathbb{R}^n$.

Definition. *A collection $(Y_t)_{t\in I}$ of random variables $Y_t : \Omega \to \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is called a* ***Gaussian process*** *if and only if the joint distribution of any finite subcollection $Y_{t_1}, \ldots, Y_{t_n}$ with $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in I$ is a multivariate normal distribution.*

The distribution of a Gaussian process $(Y_t)_{t\in I}$ on the path space $\mathbb{R}^I$ or $C(I,\mathbb{R})$ endowed with the σ-algebra generated by the maps $x \mapsto x_t$, $t \in I$, is uniquely determined by the multinormal distributions of finite subcollections $Y_{t_1}, \ldots, Y_{t_n}$ as above, and hence by the expectation values

$$m(t) \quad = \quad E[Y_t], \qquad t \in I,$$

and the covariances

$$c(s,t) \quad = \quad \mathrm{Cov}[Y_s, Y_t], \qquad s,t \in I.$$

A Gaussian process is called **centered**, if $m(t) = 0$ for any $t \in I$.

Example (AR(1) process). The autoregressive process $(Y_n)_{n=0,1,2,\ldots}$ defined recursively by $Y_0 \sim N(0, v_0)$,

$$Y_n = \alpha Y_{n-1} + \varepsilon \eta_n \qquad \text{for } n \in \mathbb{N},$$

with parameters $v_0 > 0,\ \alpha, \varepsilon \in \mathbb{R},\ \eta_n$ i.i.d. $\sim N(0,1)$, is a centered Gaussian process. The covariance function is given by

$$c(n, n+k) \quad = \quad v_0 + \varepsilon^2 n \qquad \text{for any } n,k \geq 0 \quad \text{if } \alpha = 1,$$

and

$$c(n, n+k) \quad = \quad \alpha^k \cdot \left(\alpha^{2n} v_0 + (1-\alpha^{2n}) \cdot \frac{\varepsilon^2}{1-\alpha^2}\right) \qquad \text{for } n,k \geq 0 \quad \text{otherwise.}$$

This is easily verified by induction. We now consider some special cases:

$\alpha = \mathbf{0}$**:** In this case $Y_n = \varepsilon \eta_n$. Hence (Y_n) is a *white noise*, i.e., a sequence of independent normal random variables, and

$$\mathrm{Cov}[Y_n, Y_m] \quad = \quad \varepsilon^2 \cdot \delta_{n,m} \qquad \text{for any } n, m \geq 1.$$

$\alpha = \mathbf{1}$**:** Here $Y_n = Y_0 + \varepsilon \sum\limits_{i=1}^{n} \eta_i$, i.e., the process (Y_n) is a *Gaussian Random Walk*, and

$$\mathrm{Cov}[Y_n, Y_m] \quad = \quad v_0 + \varepsilon^2 \cdot \min(n,m) \qquad \text{for any } n,m \geq 0.$$

We will see a corresponding expression for the covariances of Brownian motion.

$\alpha < 1$: For $\alpha < 1$, the covariances $\mathrm{Cov}[Y_n, Y_{n+k}]$ decay exponentially fast as $k \to \infty$. If $v_0 = \frac{\varepsilon^2}{1-\alpha^2}$, then the covariance function is translation invariant:

$$c(n, n+k) \quad = \quad \frac{\varepsilon^2 \alpha^k}{1-\alpha^2} \qquad \text{for any } n, k \geq 0.$$

Therefore, in this case the process (Y_n) is *stationary*, i.e., $(Y_{n+k})_{n\geq 0} \sim (Y_n)_{n\geq 0}$ for all $k \geq 0$.

Brownian motion is our first example of a nontrivial Gaussian process in continuous time. In fact, we have:

Theorem 1.8 (Gaussian characterization of Brownian motion). *A real-valued stochastic process $(B_t)_{t\in[0,\infty)}$ with continuous sample paths $t \mapsto B_t(\omega)$ and $B_0 = 0$ is a Brownian motion if and only if (B_t) is a centered Gaussian process with covariances*

$$\mathrm{Cov}[B_s, B_t] \quad = \quad \min(s,t) \qquad \textit{for any } s, t \geq 0. \tag{1.2.4}$$

Proof. For a Brownian motion (B_t) and $0 = t_0 < t_1 < \ldots < t_n$, the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq n$, are independent random variables with distribution $N(0, t_i - t_{i-1})$. Hence,

$$(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}) \quad \sim \quad \bigotimes_{i=1}^{n} N(0, t_i - t_{i-1}),$$

which is a multinormal distribution. Since $B_{t_0} = B_0 = 0$, we see that

$$\begin{pmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \end{pmatrix} \quad = \quad \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ & & \ddots & & & \\ & & & \ddots & & \\ 1 & 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix} \begin{pmatrix} B_{t_1} - B_{t_0} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

also has a multivariate normal distribution, i.e., (B_t) is a Gaussian process. Moreover, since $B_t = B_t - B_0$, we have $E[B_t] = 0$ and

$$\mathrm{Cov}[B_s, B_t] \quad = \quad \mathrm{Cov}[B_s, B_s] + \mathrm{Cov}[B_s, B_t - B_s] \quad = \quad \mathrm{Var}[B_s] \quad = \quad s$$

for any $0 \leq s \leq t$, i.e., (1.2.4) holds.

Conversely, if (B_t) is a centered Gaussian process satisfying (1.2.4), then for any $0 = t_0 < t_1 < \ldots < t_n$, the vector $(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}})$ has a multivariate normal distribution with

$$E[B_{t_i} - B_{t_{i-1}}] \quad = \quad E[B_{t_i}] - E[B_{t_{i-1}}] \quad = \quad 0, \qquad \text{and}$$

$$\begin{aligned} \mathrm{Cov}[B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}] &= \min(t_i, t_j) - \min(t_i, t_{j-1}) \\ &\quad - \min(t_{i-1}, t_j) + \min(t_{i-1}, t_{j-1}) \\ &= (t_i - t_{i-1}) \cdot \delta_{i,j} \qquad \text{for any } i, j = 1, \ldots, n. \end{aligned}$$

Hence by Theorem 1.5 (3), the increments $B_{t_i} - B_{t_{i-1}}, 1 \leq i \leq n$, are independent with distribution $N(0, t_i - t_{i-1})$, i.e., (B_t) is a Brownian motion. □

Symmetries of Brownian motion

A first important consequence of the Gaussian characterization of Brownian motion are several symmetry properties of Wiener measure:

Theorem 1.9 (Invariance properties of Wiener measure). *Let $(B_t)_{t\geq 0}$ be a Brownian motion starting at 0 defined on a probability space $(\Omega, \mathcal{A}, P)$. Then the following processes are again Brownian motions:*

(1). $(-B_t)_{t\geq 0}$ *(Reflection invariance)*

(2). $(B_{t+h} - B_h)_{t\geq 0}$ *for any* $h \geq 0$ *(Stationarity)*

(3). $(a^{-1/2} B_{at})_{t\geq 0}$ *for any* $a > 0$ *(Scale invariance)*

(4). The time inversion $(\widetilde{B}_t)_{t\geq 0}$ *defined by*

$$\widetilde{B}_0 = 0, \qquad \widetilde{B}_t = t \cdot B_{1/t} \quad \textit{for } t > 0.$$

Proof. The proofs of (1), (2) and (3) are left as an exercise to the reader. To show (4), we first note that for each $n \in \mathbb{N}$ and $0 \leq t_1 < \ldots < t_n$, the vector $(\widetilde{B}_{t_1}, \ldots, \widetilde{B}_{t_n})$ has a multivariate normal distribution since it is a linear transformation of $(B_{1/t_1}, \ldots, B_{1/t_n})$, $(B_0, B_{1/t_2}, \ldots, B_{1/t_n})$ respectively. Moreover,

$$\begin{aligned} E[\widetilde{B}_t] &= 0 \qquad \text{for any } t \geq 0, \\ \mathrm{Cov}[\widetilde{B}_s, \widetilde{B}_t] &= st \cdot \mathrm{Cov}[B_{1/s}, B_{1/t}] \\ &= st \cdot \min(\frac{1}{s}, \frac{1}{t}) \quad = \quad \min(t,s) \qquad \text{for any } s,t > 0, \quad \text{and} \\ \mathrm{Cov}[\widetilde{B}_0, \widetilde{B}_t] &= 0 \qquad \text{for any } t \geq 0. \end{aligned}$$

Hence $(\widetilde{B}_t)_{t\geq 0}$ is a centered Gaussian process with the covariance function of Brownian motion. By Theorem 1.8, it only remains to show that P-almost every sample path $t \mapsto \widetilde{B}_t(\omega)$ is continuous. This is obviously true for $t > 0$. Furthermore, since the finite dimensional marginals of the processes $(\widetilde{B}_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ are multivariate normal distributions with the same means and covariances, the distributions of $(\widetilde{B}_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ on the product space $\mathbb{R}^{(0,\infty)}$ endowed with the product σ-algebra generated by the cylinder sets agree. To prove continuity at 0 we note that the set

$$\left\{ x : (0,\infty) \to \mathbb{R} \;\middle|\; \lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} x_t = 0 \right\}$$

is measurable w.r.t. the product σ-algebra on $\mathbb{R}^{(0,\infty)}$. Therefore,

$$P\left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} \widetilde{B}_t = 0 \right] \quad = \quad P\left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} B_t = 0 \right] \quad = \quad 1.$$

Since $\widetilde{B}_t$ is almost surely continuous for $t > 0$, we can conclude that outside a set of measure zero,

$$\sup_{s \in (0,t)} |\widetilde{B}_s| \quad = \quad \sup_{s \in (0,t) \cap \mathbb{Q}} |\widetilde{B}_s| \quad \longrightarrow \quad 0 \qquad \text{as } t \searrow 0,$$

i.e., $t \mapsto \widetilde{B}_t$ is almost surely continuous at 0 as well. □

Remark (Long time asymptotics versus local regularity, LLN). The time inversion invariance of Wiener measure enables us to translate results on the long time asymptotics of Brownian motion ($t \nearrow \infty$) into local regularity results for Brownian paths ($t \searrow 0$) and vice versa. For example, the continuity of the process $(\widetilde{B}_t)$ at 0 is equivalent to the *law of large numbers*:

$$P\left[\lim_{t\to\infty} \frac{1}{t} B_t \;=\; 0\right] \quad = \quad P\left[\lim_{s\searrow 0} sB_{1/s} \;=\; 0\right] \quad = \quad 1.$$

At first glance, this looks like a simple proof of the LLN. However, the argument is based on the existence of a continuous Brownian motion, and the existence proof requires similar arguments as a direct proof of the law of large numbers.

Wiener measure as a Gaussian measure, path integral heuristics

Wiener measure (with start at 0) is the unique probability measure μ on the continuous path space $C([0,\infty),\mathbb{R}^d)$ such that the coordinate process

$$X_t : C([0,\infty),\mathbb{R}^d) \to \mathbb{R}^d, \qquad X_t(x) \;=\; x_t,$$

is a Brownian motion starting at 0. By Theorem 1.8, Wiener measure is a centered **Gaussian measure** on the infinite dimensional space $C([0,\infty),\mathbb{R}^d)$, i.e., for any $n \in \mathbb{N}$ and $t_1,\dots,t_n \in \mathbb{R}_+$, $(X_{t_1},\dots,X_{t_n})$ is normally distributed with mean 0. We now "derive" a heuristic representation of Wiener measure that is not mathematically rigorous but nevertheless useful:
Fix a constant $T > 0$. Then for $0 = t_0 < t_1 < \dots < t_n \le T$, the distribution of $(X_{t_1},\dots,X_{t_n})$ w.r.t. Wiener measure is

$$\mu_{t_1,\dots,t_n}(dx_{t_1},\dots,dx_{t_n}) \quad = \quad \frac{1}{Z(t_1,\dots,t_n)} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{|x_{t_i}-x_{t_{i-1}}|^2}{t_i - t_{i-1}}\right) \prod_{i=1}^{n} dx_{t_i}, \tag{1.2.5}$$

where $Z(t_1,\dots,t_n)$ is an appropriate finite normalization constant, and $x_0 := 0$. Now choose a sequence $(\tau_k)_{k\in\mathbb{N}}$ of partitions $0 = t_0^{(k)} < t_1^{(k)} < \dots < t_{n(k)}^{(k)} = T$ of the interval

$[0,T]$ such that the mesh size $\max_i |t_{i+1}^{(k)} - t_i^{(k)}|$ tends to zero. Taking informally the limit in (1.2.5), we obtain the heuristic asymptotic representation

$$\mu(dx) \quad = \quad \frac{1}{Z_\infty} \exp\left(-\frac{1}{2} \int_0^T \left| \frac{dx}{dt} \right|^2 dt \right) \delta_0(dx_0) \prod_{t \in (0,T]} dx_t \tag{1.2.6}$$

for Wiener measure on continuous paths $x : [0,T] \to \mathbb{R}^d$ with a "normalizing constant" Z_∞. Trying to make the informal expression (1.2.6) rigorous fails for several reasons:

- The normalizing constant $Z_\infty = \lim_{k\to\infty} Z(t_1^{(k)}, \ldots, t_{n(k)}^{(k)})$ is infinite.

- The integral $\int_0^T \left| \frac{dx}{dt} \right|^2 dt$ is also infinite for μ-almost every path x, since typical paths of Brownian motion are nowhere differentiable, cf. below.

- The product measure $\prod_{t\in(0,T]} dx_t$ can be defined on cylinder sets but an extension to the σ-algebra generated by the coordinate maps on $C([0,\infty),\mathbb{R}^d)$ does not exist.

Hence there are several infinities involved in the informal expression (1.2.6). These infinities magically balance each other such that the measure μ is well defined in contrast to all of the factors on the right hand side.

In physics, R. Feynman introduced correspondingly integrals w.r.t. "Lebesgue measure on path space", cf. e.g. the famous Feynman Lecture notes [...], or Glimm and Jaffe [...].

Although not mathematically rigorous, the heuristic expression (1.2.5) can be a very useful guide for intuition. Note for example that (1.2.5) takes the form

$$\mu(dx) \quad \propto \quad \exp(-\|x\|_H^2/2)\, \lambda(dx), \tag{1.2.7}$$

where $\|x\|_H = (x,x)_H^{1/2}$ is the norm induced by the inner product

$$(x,y)_H \quad = \quad \int_0^T \frac{dx}{dt}\frac{dy}{dt} dt \tag{1.2.8}$$

of functions $x, y : [0,T] \to \mathbb{R}^d$ vanishing at 0, and λ is a corresponding "infinite-dimensional Lebesgue measure" (which does not exist!). The vector space

$$H = \{x : [0,T] \to \mathbb{R}^d \ : \ x(0) = 0, x \text{ is absolutely continuous with } \frac{dx}{dt} \in L^2\}$$

is a Hilbert space w.r.t. the inner product (1.2.8). Therefore, (1.2.7) suggests to consider Wiener measure as a *standard normal distribution on H*. It turns out that this idea can be made rigorous although not as easily as one might think at first glance. The difficulty is that a standard normal distribution on an infinite-dimensional Hilbert space does not exist on the space itself but only on a larger space. In particular, we will see in the next sections that Wiener measure μ can indeed be realized on the continuous path space $C([0,T], \mathbb{R}^d)$, but μ-almost every path is not contained in H!

Remark (Infinite-dimensional standard normal distributions). The fact that a standard normal distribution on an infinite dimensional separable Hilbert space H can not be realized on the space H itself can be easily seen by contradiction: Suppose that μ is a standard normal distribution on H, and $e_n, n \in \mathbb{N}$, are infinitely many orthonormal vectors in H. Then by rotational symmetry, the balls

$$B_n = \left\{ x \in H \ : \ \|x - e_n\|_H < \frac{1}{2} \right\}, \qquad n \in \mathbb{N},$$

should all have the same measure. On the other hand, the balls are disjoint. Hence by σ-additivity,

$$\sum_{n=1}^{\infty} \mu[B_n] = \mu\left[\bigcup B_n\right] \leq \mu[H] = 1,$$

and therefore $\mu[B_n] = 0$ for all $n \in \mathbb{N}$. A scaling argument now implies

$$\mu[\{x \in H \ : \ \|x - h\| \leq \|h\|/2\}] = 0 \qquad \text{for all } h \in H,$$

and hence $\mu \equiv 0$.

1.3 The Wiener-Lévy Construction

In this section we discuss how to construct Brownian motion as a random superposition of deterministic paths. The idea already goes back to N. Wiener, who constructed

Brownian motion as a random Fourier series. The approach described here is slightly different and due to P. Lévy: The idea is to approximate the paths of Brownian motion on a finite time interval by their piecewise linear interpolations w.r.t. the sequence of dyadic partitions. This corresponds to a development of the Brownian paths w.r.t. Schauder functions ("wavelets") which turns out to be very useful for many applications including numerical simulations.

Our aim is to construct a one-dimensional Brownian motion B_t starting at 0 for $t \in [0,1]$. By stationarity and independence of the increments, a Brownian motion defined for all $t \in [0,\infty)$ can then easily be obtained from infinitely many independent copies of Brownian motion on $[0,1]$. We are hence looking for a random variable

$$B \;=\; (B_t)_{t\in[0,1]} \;:\Omega \;\longrightarrow\; C([0,1])$$

defined on a probability space $(\Omega, \mathcal{A}, P)$ such that the distribution $P\circ B^{-1}$ is Wiener measure μ on the continuous path space $C([0,1])$.

A first attempt

Recall that μ_0 should be a kind of standard normal distribution w.r.t. the inner product

$$(x,y)_H \quad = \quad \int_0^1 \frac{dx}{dt}\frac{dy}{dt}\,dt \tag{1.3.1}$$

on functions $x, y : [0,1] \to \mathbb{R}$. Therefore, we could try to define

$$B_t(\omega) \quad := \quad \sum_{i=1}^{\infty} Z_i(\omega)e_i(t) \qquad \text{for } t\in[0,1] \text{ and } \omega\in\Omega, \tag{1.3.2}$$

where $(Z_i)_{i\in\mathbb{N}}$ is a sequence of independent standard normal random variables, and $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis in the Hilbert space

$$H \quad = \quad \{x : [0,1]\to\mathbb{R} \mid x(0) = 0,\ x \text{ is absolutely continuous with } (x,x)_H < \infty\}. \tag{1.3.3}$$

However, the resulting series approximation does not converge in H:

Theorem 1.10. *Suppose $(e_i)_{i\in\mathbb{N}}$ is a sequence of orthonormal vectors in a Hilbert space H and $(Z_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with $P[Z_i \neq 0] > 0$. Then the series $\sum\limits_{i=1}^{\infty} Z_i(\omega)e_i$ diverges with probability 1 w.r.t. the norm on H.*

Proof. By orthonormality and by the law of large numbers,

$$\left\|\sum_{i=1}^{n} Z_i(\omega)e_i\right\|_H^2 \quad = \quad \sum_{i=1}^{n} Z_i(\omega)^2 \quad \longrightarrow \quad \infty$$

P-almost surely as $n \to \infty$. □

The Theorem again reflects the fact that a standard normal distribution on an infinite-dimensional Hilbert space can not be realized on the space itself.

To obtain a positive result, we will replace the norm

$$\|x\|_H \quad = \quad \left(\int_0^1 \left|\frac{dx}{dt}\right|^2 \, dt\right)^{\frac{1}{2}}$$

on H by the supremum norm

$$\|x\|_{\sup} \quad = \quad \sup_{t\in[0,1]} |x_t|,$$

and correspondingly the Hilbert space H by the Banach space $C([0,1])$. Note that the supremum norm is weaker than the H-norm. In fact, for $x \in H$ and $t \in [0,1]$, the Cauchy-Schwarz inequality implies

$$|x_t|^2 \quad = \quad \left|\int_0^t x'_s \, ds\right|^2 \quad \leq \quad t \cdot \int_0^t |x'_s|^2 \, ds \quad \leq \quad \|x\|_H^2,$$

and therefore

$$\|x\|_{\sup} \quad \leq \quad \|x\|_H \qquad \text{for any } x \in H.$$

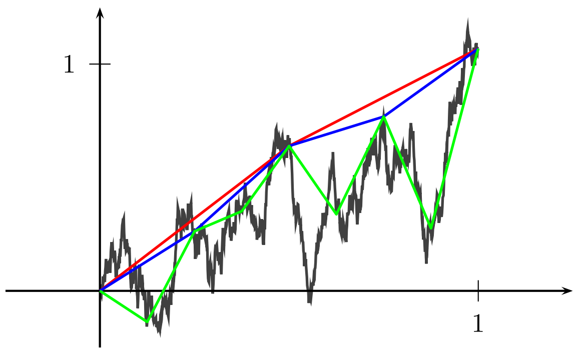
There are two choices for an orthonormal basis of the Hilbert space H that are of particular interest: The first is the Fourier basis given by

$$e_0(t) \quad = \quad t, \qquad e_n(t) \quad = \quad \frac{\sqrt{2}}{\pi n}\sin(\pi n t) \qquad \text{for } n \geq 1.$$

With respect to this basis, the series in (1.3.2) is a Fourier series with random coefficients. Wiener's original construction of Brownian motion is based on a *random Fourier series*. A second convenient choice is the basis of *Schauder functions* ("wavelets") that has been used by P. Lévy to construct Brownian motion. Below, we will discuss Lévy's construction in detail. In particular, we will prove that for the Schauder functions, the series in (1.3.2) converges almost surely w.r.t. the supremum norm towards a continuous (but not absolutely continuous) random path $(B_t)_{t\in[0,1]}$. It is then not difficult to conclude that $(B_t)_{t\in[0,1]}$ is indeed a Brownian motion.

The Wiener-Lévy representation of Brownian motion

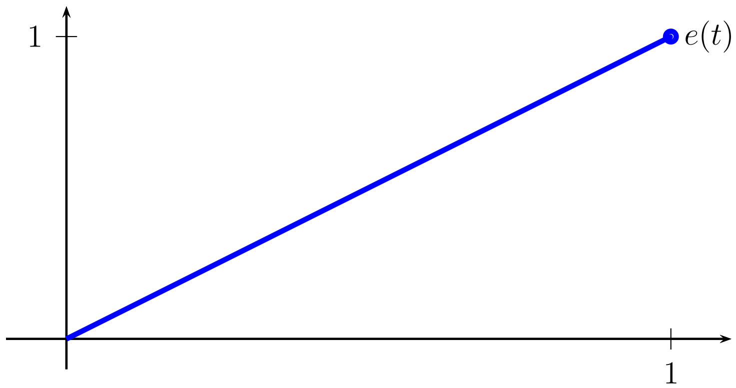
Before carrying out Lévy's construction of Brownian motion, we introduce the Schauder functions, and we show how to expand a given Brownian motion w.r.t. this basis of function space. Suppose we would like to approximate the paths $t \mapsto B_t(\omega)$ of a Brownian motion by their piecewise linear approximations adapted to the sequence of dyadic partitions of the interval $[0, 1]$.

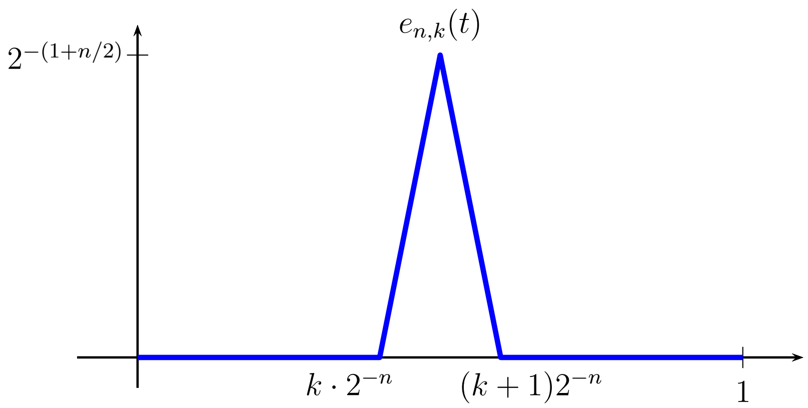


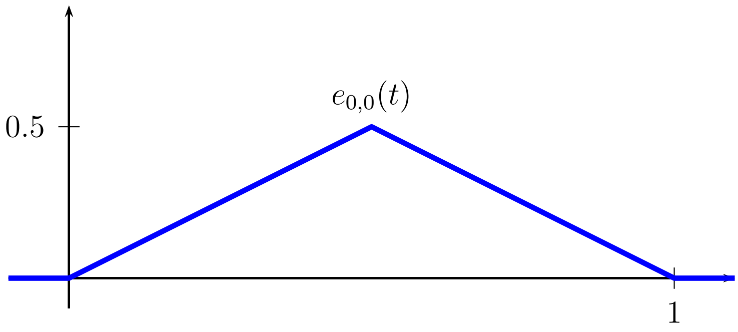
An obvious advantage of this approximation over a Fourier expansion is that the values of the approximating functions at the dyadic points remain fixed once the approximating partition is fine enough. The piecewise linear approximations of a continuous function on $[0,1]$ correspond to a series expansion w.r.t. the base functions

$$\begin{aligned} e(t) &= t \quad \text{, and} \\ e_{n,k}(t) &= 2^{-n/2} e_{0,0}(2^n t - k), \quad n = 0,1,2,\ldots, k = 0,1,2,\ldots,2^n - 1, \quad \text{, where} \end{aligned}$$

$$e_{0,0}(t) = \min(t, 1-t)^+ = \begin{cases} t & \text{for } t \in [0,1/2] \\ 1-t & \text{for } t \in (1/2,1] \\ 0 & \text{for } t \in \mathbb{R} \setminus [0,1] \end{cases} .$$







The functions $e_{n,k}$ $(n \geq 0, 0 \leq k < 2^n)$ are called **Schauder functions**. It is rather obvious that piecewise linear approximation w.r.t. the dyadic partitions corresponds to the expansion of a function $x \in C([0,1])$ with $x(0) = 0$ in the basis given by $e(t)$ and the Schauder functions. The normalization constants in defining the functions $e_{n,k}$ have been chosen in such a way that the $e_{n,k}$ are orthonormal w.r.t. the H-inner product introduced above.

Definition. *A sequence $(e_i)_{i\in\mathbb{N}}$ of vectors in an infinite-dimensional Hilbert space H is called an **orthonormal basis** (or **complete orthonormal system**) of H if and only if*

(1). Orthonormality: $\quad (e_i, e_j) \;=\; \delta_{ij} \qquad$ *for any $i, j \in \mathbb{N}$, $\qquad$ and*

(2). Completeness: Any $h \in H$ can be expressed as

$$h \;=\; \sum_{i=1}^{\infty} (h, e_i)_H e_i.$$

Remark (Equivalent characterizations of orthonormal bases). Let $e_i, i \in \mathbb{N}$, be orthonormal vectors in a Hilbert space H. Then the following conditions are equivalent:

(1). $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis of H.

(2). The linear span

$$\mathrm{span}\{e_i \mid i \in \mathbb{N}\} \;=\; \left\{ \sum_{i=1}^{k} c_i e_i \;\middle|\; k \in \mathbb{N}, c_1, \ldots, c_k \in \mathbb{R} \right\}$$

is a dense subset of H.

(3). There is no element $x \in H, x \neq 0$, such that $(x, e_i)_H = 0$ for every $i \in \mathbb{N}$.

(4). For any element $x \in H$, *Parseval's relation*

$$\|x\|_H^2 \quad = \quad \sum_{i=1}^{\infty} (x, e_i)_H^2 \tag{1.3.4}$$

holds.

(5). For any $x, y \in H$,

$$(x, y)_H \quad = \quad \sum_{i=1}^{\infty} (x, e_i)_H (y, e_i)_H. \tag{1.3.5}$$

For the proofs we refer to any book on functional analysis, cf. e.g. [Reed and Simon: Methods of modern mathematical physics, Vol. I].

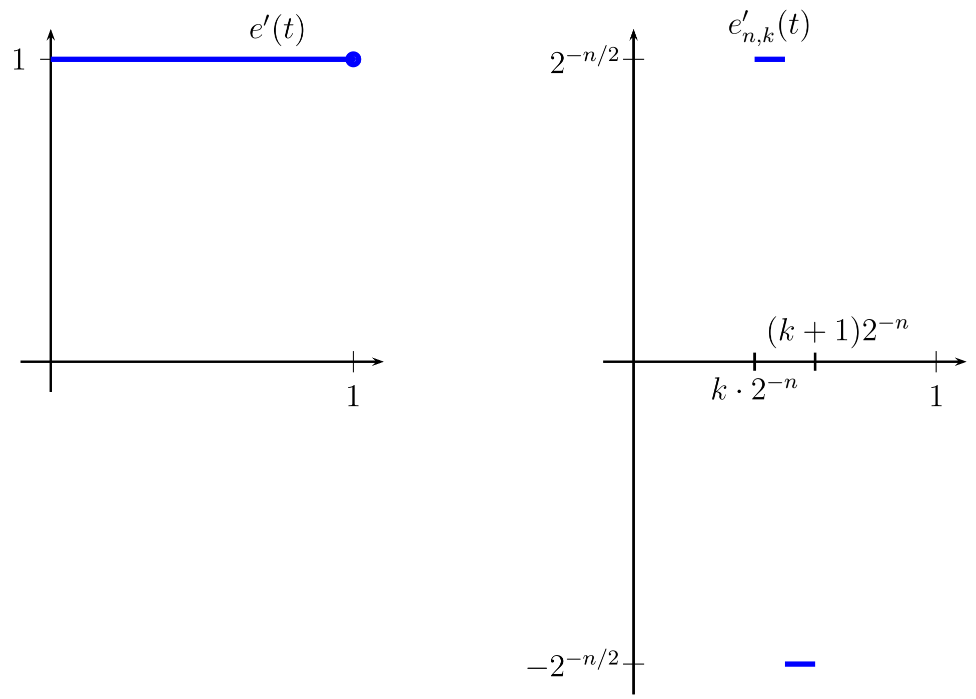
Lemma 1.11. *The Schauder functions e and $e_{n,k}$ $(n \geq 0, 0 \leq k < 2^n)$ form an orthonormal basis in the Hilbert space H defined by (1.3.3).*

Proof. By definition of the inner product on H, the linear map d/dt which maps an absolutely continuous function $x \in H$ to its derivative $x' \in L^2(0,1)$ is an isometry from H onto $L^2(0,1)$, i.e.,

$$(x, y)_H \quad = \quad (x', y')_{L^2(0,1)} \qquad \text{for any } x, y \in H.$$

The derivatives of the Schauder functions are the Haar functions

$$\begin{aligned} e'(t) &\equiv 1, \\ e'_{n,k}(t) &= 2^{n/2}(I_{[k\cdot 2^{-n},(k+1/2)\cdot 2^{-n})}(t) - I_{[(k+1/2)\cdot 2^{-n},(k+1)\cdot 2^{-n})}(t)) \qquad \text{for a.e. } t. \end{aligned}$$



It is easy to see that these functions form an orthonormal basis in $L^2(0,1)$. In fact, orthonormality w.r.t. the L^2 inner product can be verified directly. Moreover, the linear span of the functions e' and $e'_{n,k}$ for $n = 0, 1, \ldots, m$ and $k = 0, 1, \ldots, 2^n - 1$ consists of all step functions that are constant on each dyadic interval $[j \cdot 2^{-(m+1)}, (j+1) \cdot 2^{-(m+1)})$. An arbitrary function in $L^2(0,1)$ can be approximated by dyadic step functions w.r.t. the L^2 norm. This follows for example directly from the L^2 martingale convergence Theorem, cf. ... below. Hence the linear span of e' and the Haar functions $e'_{n,k}$ is dense in $L^2(0,1)$, and therefore these functions form an orthonormal basis of the Hilbert space $L^2(0,1)$. Since $x \mapsto x'$ is an isometry from H onto $L^2(0,1)$, we can conclude that e and the Schauder functions $e_{n,k}$ form an orthonormal basis of H. $\square$

The expansion of a function $x : [0,1] \to \mathbb{R}$ in the basis of Schauder functions can now be made explicit. The coefficients of a function $x \in H$ in the expansion are

$$(x,e)_H \quad = \quad \int_0^1 x'e' \, dt \quad = \quad \int_0^1 x' \, dt \quad = \quad x(1) - x(0) \quad = \quad x(1)$$

$$(x, e_{n,k})_H = \int_0^1 x' e'_{n,k}\, dt \quad = \quad 2^{n/2} \int_0^1 x'(t) e'_{0,0}(2^n t - k)\, dt$$
$$= \quad 2^{n/2} \left[(x((k+\frac{1}{2}) \cdot 2^{-n}) - x(k \cdot 2^{-n})) - (x((k+1) \cdot 2^{-n}) - x((k+\frac{1}{2}) \cdot 2^{-n})) \right].$$

Theorem 1.12. *Let $x \in C([0,1])$. Then the expansion*

$$x(t) \quad = \quad x(1)e(t) - \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} 2^{n/2} \Delta_{n,k} x \cdot e_{n,k}(t),$$

$$\Delta_{n,k} x \ = \ \left[(x((k+1) \cdot 2^{-n}) - x((k+\frac{1}{2}) \cdot 2^{-n})) - (x((k+\frac{1}{2}) \cdot 2^{-n}) - x(k \cdot 2^{-n})) \right]$$

holds w.r.t. uniform convergence on $[0,1]$. For $x \in H$ the series also converges w.r.t. the stronger H-norm.

Proof. It can be easily verified that by definition of the Schauder functions, for each $m \in \mathbb{N}$ the partial sum

$$x^{(m)}(t) \quad := \quad x(1)e(t) - \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} 2^{n/2} \Delta_{n,k} x \cdot e_{n,k}(t) \tag{1.3.6}$$

is the polygonal interpolation of $x(t)$ w.r.t. the $(m+1)$-th dyadic partition of the interval $[0,1]$. Since the function x is uniformly continuous on $[0,1]$, the polygonal interpolations converge uniformly to x. This proves the first statement. Moreover, for $x \in H$, the series is the expansion of x in the orthonormal basis of H given by the Schauder functions, and therefore it also converges w.r.t. the H-norm. □

Applying the expansion to the paths of a Brownian motions, we obtain:

Corollary 1.13 (Wiener-Lévy representation). *For a Brownian motion $(B_t)_{t\in[0,1]}$ the series representation*

$$B_t(\omega) \quad = \quad Z(\omega)e(t) + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k}(\omega) e_{n,k}(t), \qquad t \in [0,1], \tag{1.3.7}$$

holds w.r.t. uniform convergence on $[0,1]$ for P-almost every $\omega \in \Omega$, where

$$Z \;:=\; B_1, \quad \textit{and} \quad Z_{n,k} \;:=\; -2^{n/2}\Delta_{n,k}B \qquad (n \geq 0, 0 \leq k \leq 2^n - 1)$$

are independent random variables with standard normal distribution.

Proof. It only remains to verify that the coefficients Z and $Z_{n,k}$ are independent with standard normal distribution. A vector given by finitely many of these random variables has a multivariate normal distribution, since it is a linear transformation of increments of the Brownian motion B_t. Hence it suffices to show that the random variables are uncorrelated with variance 1. This is left as an exercise to the reader. □

Lévy's construction of Brownian motion

The series representation (1.3.7) can be used to construct Brownian motion starting from independent standard normal random variables. The resulting construction does not only prove existence of Brownian motion but it is also very useful for numerical implementations:

Theorem 1.14 (P. Lévy 1948). *Let Z and $Z_{n,k}$ $(n \geq 0, 0 \leq k \leq 2^n - 1)$ be independent standard normally distributed random variables on a probability space $(\Omega, \mathcal{A}, P)$. Then the series in (1.3.7) converges uniformly on $[0,1]$ with probability 1. The limit process $(B_t)_{t\in[0,1]}$ is a Brownian motion.*

The convergence proof relies on a combination of the Borel-Cantelli Lemma and the Weierstrass criterion for uniform convergence of series of functions. Moreover, we will need the following result to identify the limit process as a Brownian motion:

Lemma 1.15 (Parseval relation for Schauder functions). *For any $s,t \in [0,1]$,*

$$e(t)e(s) + \sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s) \quad = \quad \min(t,s).$$

Proof. Note that for $g \in H$ and $s \in [0,1]$, we have

$$g(s) \quad = \quad g(s) - g(0) \quad = \quad \int_0^1 g' \cdot I_{(0,s)} \quad = \quad (g, h^{(s)})_H,$$

where $h^{(s)}(t) \;:=\; \int_0^t I_{(0,s)} \;=\; \min(s,t)$. Hence the Parseval relation (1.3.4) applied to the functions $h^{(s)}$ and $h^{(t)}$ yields

$$\begin{aligned}
&e(t)e(s) + \sum_{n,k} e_{n,k}(t)e_{n,k}(s) \\
&= \quad (e, h^{(t)})(e, h^{(s)}) + \sum_{n,k} (e_{n,k}, h^{(t)})(e_{n,k}, h^{(s)}) \\
&= \quad (h^{(t)}, h^{(s)}) \quad = \quad \int_0^1 I_{(0,t)} I_{(0,s)} \quad = \quad \min(t,s).
\end{aligned}$$

□

Proof of Theorem 1.14. We proceed in 4 steps:

(1). *Uniform convergence for P-a.e. ω:* By the Weierstrass criterion, a series of functions converges uniformly if the sum of the supremum norms of the summands is finite. To apply the criterion, we note that for any fixed $t \in [0,1]$ and $n \in \mathbb{N}$, only one of the functions $e_{n,k}$, $k = 0, 1, \ldots, 2^n - 1$, does not vanish at t. Moreover, $|e_{n,k}(t)| \le 2^{-n/2}$. Hence

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n - 1} Z_{n,k}(\omega) e_{n,k}(t) \right| \quad \le \quad 2^{-n/2} \cdot M_n(\omega), \tag{1.3.8}$$

where

$$M_n \quad := \quad \max_{0 \le k < 2^n} |Z_{n,k}|.$$

We now apply the Borel-Cantelli Lemma to show that with probability 1, M_n grows at most linearly. Let Z denote a standard normal random variable. Then we have

$$\begin{aligned}
P[M_n > n] \quad &\le \quad 2^n \cdot P[|Z| > n] \quad \le \quad \frac{2^n}{n} \cdot E[|Z| \,;\, |Z| > n] \\
&= \quad \frac{2 \cdot 2^n}{n \cdot \sqrt{2\pi}} \int_n^\infty x e^{-x^2/2}\, dx \quad = \quad \sqrt{\frac{2}{\pi}} \frac{2^n}{n} \cdot e^{-n^2/2}
\end{aligned}$$

for any $n \in \mathbb{N}$. Since the sequence on the right hand side is summable, $M_n \leq n$ holds eventually with probability one. Therefore, the sequence on the right hand side of (1.3.8) is also summable for P-almost every ω. Hence, by (1.3.8) and the Weierstrass criterion, the partial sums

$$B_t^{(m)}(\omega) \quad = \quad Z(\omega)e(t) + \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} Z_{n,k}(\omega)e_{n,k}(t), \qquad m \in \mathbb{N},$$

converge almost surely uniformly on $[0, 1]$. Let

$$B_t \quad = \quad \lim_{m\to\infty} B_t^{(m)}$$

denote the almost surely defined limit.

(2). *L^2 convergence for fixed t:* We now want to prove that the limit process (B_t) is a Brownian motion, i.e., a continuous Gaussian process with $E[B_t] = 0$ and $\mathrm{Cov}[B_t, B_s] = \min(t, s)$ for any $t, s \in [0, 1]$. To compute the covariances we first show that for a given $t \in [0, 1]$ the series approximation $B_t^{(m)}$ of B_t converges also in L^2. Let $l, m \in \mathbb{N}$ with $l < m$. Since the $Z_{n,k}$ are independent (and hence uncorrelated) with variance 1, we have

$$E[(B_t^{(m)} - B_t^{(l)})^2] \;=\; E\left[\left(\sum_{n=l+1}^{m} \sum_{k=0}^{2^n-1} Z_{n,k}e_{n,k}(t)\right)^2\right] \;=\; \sum_{n=l+1}^{m} \sum_{k} e_{n,k}(t)^2.$$

The right hand side converges to 0 as $l, m \to \infty$ since $\sum_{n,k} e_{n,k}(t)^2 < \infty$ by Lemma 1.15. Hence $B_t^{(m)}, m \in \mathbb{N}$, is a Cauchy sequence in $L^2(\Omega, \mathcal{A}, P)$. Since $B_t = \lim_{m\to\infty} B_t^{(m)}$ almost surely, we obtain

$$B_t^{(m)} \quad \overset{m\to\infty}{\longrightarrow} \quad B_t \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

(3). *Expectations and Covariances:* By the L^2 convergence we obtain for any $s, t \in [0, 1]$:

$$E[B_t] \quad = \quad \lim_{m\to\infty} E[B_t^{(m)}] \quad = \quad 0, \qquad \text{and}$$

$$\begin{aligned} \mathrm{Cov}[B_t, B_s] \quad &= \quad E[B_t B_s] \quad = \quad \lim_{m\to\infty} E[B_t^{(m)} B_s^{(m)}] \\ &= \quad e(t)e(s) + \lim_{m\to\infty} \sum_{n=0}^{m} \sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s). \end{aligned}$$

Here we have used again that the random variables Z and $Z_{n,k}$ are independent with variance 1. By Parseval's relation (Lemma 1.15), we conclude

$$\mathrm{Cov}[B_t, B_s] \quad = \quad \min(t, s).$$

Since the process $(B_t)_{t\in[0,1]}$ has the right expectations and covariances, and, by construction, almost surely continuous paths, it only remains to show that (B_t) is a Gaussian process in oder to complete the proof:

(4). $(B_t)_{t\in[0,1]}$ *is a Gaussian process:* We have to show that $(B_{t_1}, \ldots, B_{t_l})$ has a multivariate normal distribution for any $0 \leq t_1 < \ldots < t_l \leq 1$. By Theorem 1.5, it suffices to verify that any linear combination of the components is normally distributed. This holds by the next Lemma since

$$\sum_{j=1}^{l} p_j B_{t_j} \quad = \quad \lim_{m\to\infty} \sum_{j=1}^{l} p_j B_{t_j}^{(m)} \qquad P\text{-a.s.}$$

is an almost sure limit of normally distributed random variables for any $p_1, \ldots, p_l \in \mathbb{R}$.

Combining Steps $3, 4$ and the continuity of sample paths, we conclude that $(B_t)_{t\in[0,1]}$ is indeed a Brownian motion. □

Lemma 1.16. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of normally distributed random variables defined on a joint probability space $(\Omega, \mathcal{A}, P)$, and X_n converges almost surely to a random variable X. Then X is also normally distributed.*

Proof. Suppose $X_n \sim N(m_n, \sigma_n^2)$ with $m_n \in \mathbb{R}$ and $\sigma_n \in (0, \infty)$. By the Dominated Convergence Theorem,

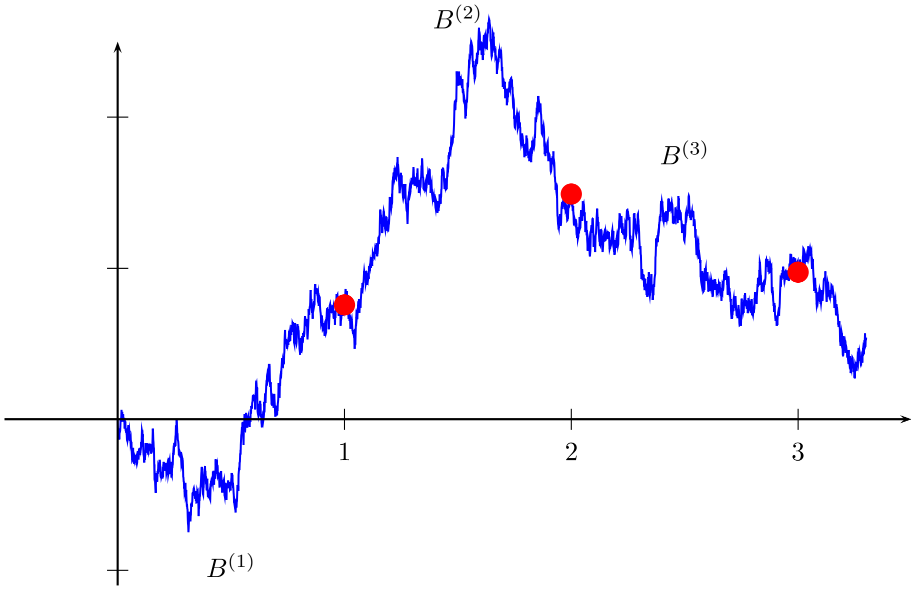
$$E[e^{ipX}] \;=\; \lim_{n\to\infty} E[e^{ipX_n}] \;=\; \lim_{n\to\infty} e^{ipm_n} e^{-\frac{1}{2}\sigma_n^2 p^2}.$$

The limit on the right hand side only exists for all p, if either $\sigma_n \to \infty$, or the sequences σ_n and m_n both converge to finite limits $\sigma \in [0, \infty)$ and $m \in \mathbb{R}$. In the first case, the limit would equal 0 for $p \neq 0$ and 1 for $p = 0$. This is a contradiction, since characteristic functions are always continuous. Hence the second case occurs, and, therefore

$$E[e^{ipX}] \quad = \quad e^{ipm-\frac{1}{2}\sigma^2 p^2} \qquad \text{for any } p \in \mathbb{R},$$

i.e., $X \sim N(m, \sigma^2)$. □

So far, we have constructed Brownian motion only for $t \in [0,1]$. Brownian motion on any finite time interval can easily be obtained from this process by rescaling. Brownian motion defined for all $t \in \mathbb{R}_+$ can be obtained by joining infinitely many Brownian motions on time intervals of length 1:



Theorem 1.17. *Suppose that $B_t^{(1)}, B_t^{(2)}, \ldots$ are independent Brownian motions starting at 0 defined for $t \in [0,1]$. Then the process*

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

is a Brownian motion defined for $t \in [0,\infty)$.

The proof is left as an exercise.

1.4 The Brownian Sample Paths

In this section we study some properties of Brownian sample paths in dimension one. We show that a typical Brownian path is nowhere differentiable, and Hölder-continuous with parameter α if and only if $\alpha < 1/2$. Furthermore, the set $\Lambda_a = \{t \geq 0 : B_t = a\}$ of all passage times of a given point $a \in \mathbb{R}$ is a fractal. We will show that almost surely, Λ_a has Lebesgue measure zero but any point in Λ_a is an accumulation point of Λ_a.
We consider a one-dimensional Brownian motion $(B_t)_{t\geq 0}$ with $B_0 = 0$ defined on a probability space $(\Omega, \mathcal{A}, P)$. Then:

Typical Brownian sample paths are nowhere differentiable

For any $t \geq 0$ and $h > 0$, the difference quotient $\frac{B_{t+h}-B_t}{h}$ is normally distributed with mean 0 and standard deviation

$$\sigma[(B_{t+h} - B_t)/h] \quad = \quad \sigma[B_{t+h} - B_t]/h \quad = \quad 1/\sqrt{h}.$$

This suggests that the derivative

$$\frac{d}{dt}B_t \quad = \quad \lim_{h\searrow 0} \frac{B_{t+h} - B_t}{h}$$

does not exist. Indeed, we have the following stronger statement.

Theorem 1.18 (Paley, Wiener, Zygmund 1933). *Almost surely, the Brownian sample path $t \mapsto B_t$ is nowhere differentiable, and*

$$\limsup_{s\searrow t} \left| \frac{B_s - B_t}{s - t} \right| \quad = \quad \infty \qquad \textit{for any } t \geq 0.$$

Note that, since there are uncountably many $t \geq 0$, the statement is stronger than claiming only the almost sure non-differentiability for any given $t \geq 0$.

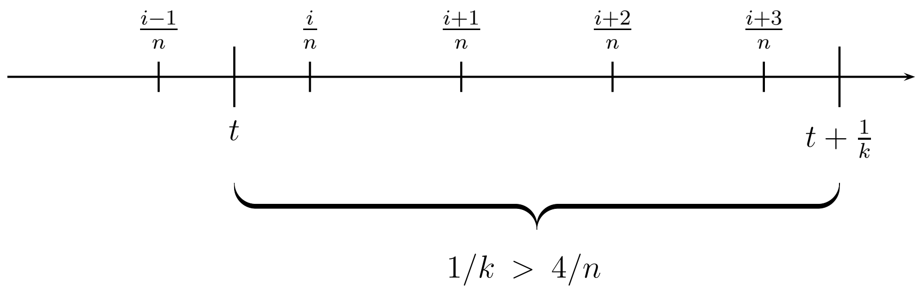
Proof. It suffices to show that the set

$$N = \left\{ \omega \in \Omega \,\middle|\, \exists\, t \in [0,T], k, L \in \mathbb{N}\ \forall\, s \in (t, t+\frac{1}{k}) \,:\, |B_s(\omega) - B_t(\omega)| \leq L|s-t| \right\}$$

is a null set for any $T \in \mathbb{N}$. Hence fix $T \in \mathbb{N}$, and consider $\omega \in N$. Then there exist $k, L \in \mathbb{N}$ and $t \in [0,T]$ such that

$$|B_s(\omega) - B_t(\omega)| \quad \leq \quad L \cdot |s-t| \qquad \text{holds for } s \in (t, t+\frac{1}{k}). \tag{1.4.1}$$

To make use of the independence of the increments over disjoint intervals, we note that for any $n > 4k$, we can find an $i \in \{1, 2, \dots, nT\}$ such that the intervals $(\frac{i}{n}, \frac{i+1}{n})$, $(\frac{i+1}{n}, \frac{i+2}{n})$, and $(\frac{i+2}{n}, \frac{i+3}{n})$ are all contained in $(t, t+\frac{1}{k})$:



Hence by (1.4.1), the bound

$$\begin{aligned} \left| B_{\frac{j+1}{n}}(\omega) - B_{\frac{j}{n}}(\omega) \right| &\leq \left| B_{\frac{j+1}{n}}(\omega) - B_t(\omega) \right| + \left| B_t(\omega) - B_{\frac{j}{n}}(\omega) \right| \\ &\leq L \cdot (\frac{j+1}{n} - t) + L \cdot (\frac{j}{n} - t) \quad \leq \quad \frac{8L}{n} \end{aligned}$$

holds for $j = i, i+1, i+2$. Thus we have shown that N is contained in the set

$$\widetilde{N} \quad := \quad \bigcup_{k,L \in \mathbb{N}} \bigcap_{n > 4k} \bigcup_{i=1}^{nT} \left\{ \left| B_{\frac{j+1}{n}} - B_{\frac{j}{n}} \right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2 \right\}.$$

We now prove $P[\widetilde{N}] = 0$. By independence and stationarity of the increments we have

$$\begin{aligned} & P\left[\left\{\left|B_{\frac{j+1}{n}} - B_{\frac{j}{n}}\right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2\right\}\right] \\ = \; & P\left[\left|B_{\frac{1}{n}}\right| \leq \frac{8L}{n}\right]^3 \quad = \quad P\left[|B_1| \leq \frac{8L}{\sqrt{n}}\right]^3 \qquad (1.4.2) \\ \leq \; & \left(\frac{1}{\sqrt{2\pi}}\frac{16L}{\sqrt{n}}\right)^3 \quad = \quad \frac{16^3}{\sqrt{2\pi}^3} \cdot \frac{L^3}{n^{3/2}} \end{aligned}$$

for any i and n. Here we have used that the standard normal density is bounded from above by $1/\sqrt{2\pi}$. By (1.4.2) we obtain

$$\begin{aligned} & P\left[\bigcap_{n>4k}\bigcup_{i=1}^{nT}\left\{\left|B_{\frac{j+1}{n}} - B_{\frac{j}{n}}\right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2\right\}\right] \\ \leq \; & \frac{16^3}{\sqrt{2\pi}^3} \cdot \inf_{n>4k} nTL^3/n^{3/2} \quad = \quad 0. \end{aligned}$$

Hence, $P[\widetilde{N}] = 0$, and therefore N is a null set. □

Hölder continuity

The statement of Theorem 1.18 says that a typical Brownian path is not Lipschitz continuous on any non-empty open interval. On the other hand, the Wiener-Lévy construction shows that the sample paths are continuous. We can almost close the gap between these two statements by arguing in both cases slightly more carefully:

Theorem 1.19. *The following statements hold almost surely:*

(1). For any $\alpha > 1/2$,

$$\limsup_{s \searrow t} \frac{|B_s - B_t|}{|s-t|^\alpha} \quad = \quad \infty \qquad \textit{for all } t \geq 0.$$

(2). For any $\alpha < 1/2$,

$$\sup_{\substack{s,t \in [0,T] \\ s \neq t}} \frac{|B_s - B_t|}{|s-t|^\alpha} \quad < \quad \infty \qquad \textit{for all } T > 0.$$

Hence a typical Brownian path is nowhere Hölder continuous with parameter $\alpha > 1/2$, but it is Hölder continuous with parameter $\alpha < 1/2$ on any finite interval. The critical case $\alpha = 1/2$ is more delicate, and will be briefly discussed below.

Proof of Theorem 1.19. The first statement can be shown by a similar argument as in the proof of Theorem 1.18. The details are left to the reader.

To prove the second statement for $T = 1$, we use the Wiener-Lévy representation

$$B_t \quad = \quad Z \cdot t + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k} e_{n,k}(t) \qquad \text{for any } t \in [0,1]$$

with independent standard normal random variables $Z, Z_{n,k}$. For $t, s \in [0,1]$ we obtain

$$|B_t - B_s| \quad \leq \quad |Z| \cdot |t-s| + \sum_n M_n \sum_k |e_{n,k}(t) - e_{n,k}(s)|,$$

where $M_n := \max_k |Z_{n,k}|$ as in the proof of Theorem 1.14. We have shown above that by the Borel-Cantelli Lemma, $M_n \leq n$ eventually with probability one, and hence

$$M_n(\omega) \quad \leq \quad C(\omega) \cdot n$$

for some almost surely finite constant $C(\omega)$. Moreover, note that for each s, t and n, at most two summands in $\sum_k |e_{n,k}(t) - e_{n,k}(s)|$ do not vanish. Since $|e_{n,k}(t)| \leq \frac{1}{2} \cdot 2^{-n/2}$ and $|e'_{n,k}(t)| \leq 2^{n/2}$, we obtain the estimates

$$|e_{n,k}(t) - e_{n,k}(s)| \;\leq\; 2^{-n/2}, \qquad \text{and} \tag{1.4.3}$$

$$|e_{n,k}(t) - e_{n,k}(s)| \;\leq\; 2^{n/2} \cdot |t-s|. \tag{1.4.4}$$

For given $s, t \in [0,1]$, we now choose $N \in \mathbb{N}$ such that

$$2^{-N} \quad \leq \quad |t-s| \quad < \quad 2^{1-N}. \tag{1.4.5}$$

By applying (1.4.3) for $n > N$ and (1.4.4) for $n \leq N$, we obtain

$$|B_t - B_s| \quad \leq \quad |Z| \cdot |t-s| + 2C \cdot \left(\sum_{n=1}^{N} n 2^{n/2} \cdot |t-s| + \sum_{n=N+1}^{\infty} n 2^{-n/2} \right).$$

By (1.4.5) the sums on the right hand side can both be bounded by a constant multiple of $|t-s|^\alpha$ for any $\alpha < 1/2$. This proves that $(B_t)_{t \in [0,1]}$ is almost surely Hölder-continuous of order α. $\square$

Law of the iterated logarithm

Khintchine's version of the law of the iterated logarithm is a much more precise statement on the local regularity of a typical Brownian path at a fixed time $s \geq 0$. It implies in particular that almost every Brownian path is not Hölder continuous with parameter $\alpha = 1/2$. We state the result without proof:

Theorem 1.20 (Khintchine 1924). *For $s \geq 0$, the following statements hold almost surely:*

$$\limsup_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log\log(1/t)}} = +1, \quad \textit{and} \quad \liminf_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log\log(1/t)}} = -1.$$

For the proof cf. e.g. Breiman, Probability, Section 12.9.
By a time inversion, the Theorem translates into a statement on the global asymptotics of Brownian paths:

Corollary 1.21. *The following statements hold almost surely:*

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = +1, \quad \textit{and} \quad \liminf_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log t}} = -1.$$

Proof. This follows by applying the Theorem above to the Brownian motion $\widehat{B}_t = t \cdot B_{1/t}$. For example, substituting $h = 1/t$, we have

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log\log(t)}} = \limsup_{h \searrow 0} \frac{h \cdot B_{1/h}}{\sqrt{2h \log\log 1/h}} = +1$$

almost surely. □

The corollary is a continuous time analogue of Kolmogorov's law of the iterated logarithm for Random Walks stating that for $S_n = \sum_{i=1}^{n} \eta_i$, η_i i.i.d. with $E[\eta_i] = 0$ and $\mathrm{Var}[\eta_i] = 1$, one has

$$\limsup_{n \to \infty} \frac{S_n}{\sqrt{2n \log\log n}} = +1 \quad \text{and} \quad \liminf_{n \to \infty} \frac{S_n}{\sqrt{2n \log\log n}} = -1$$

almost surely. In fact, one way to prove Kolmogorov's LIL is to embed the Random Walk into a Brownian motion, cf. e.g. Rogers and Williams, Vol. I, Ch. 7 or Section 3.3

Passage times

We now study the set of passage times to a given level a for a one-dimensional Brownian motion $(B_t)_{t\geq 0}$. This set has interesting properties – in particular it is a random fractal. Fix $a \in \mathbb{R}$, and let

$$\Lambda_a(\omega) \quad = \quad \{t \geq 0 \, : \, B_t(\omega) = a\} \quad \subseteq \quad [0, \infty).$$

Assuming that every path is continuous, the random set $\Lambda_a(\omega)$ is *closed* for every ω. Moreover, scale invariance of Brownian motion implies a *statistical self similarity* property for the sets of passage times: Since the rescaled process $(c^{-1/2}B_{ct})_{t\geq 0}$ has the same distribution as $(B_t)_{t\geq 0}$ for any $c > 0$, we can conclude that the set valued random variable $c \cdot \Lambda_{a/\sqrt{c}}$ has the same distribution as Λ_a. In particular, Λ_0 is a *fractal* in the sense that

$$\Lambda_0 \quad \sim \quad c \cdot \Lambda_0 \qquad \text{for any } c > 0.$$

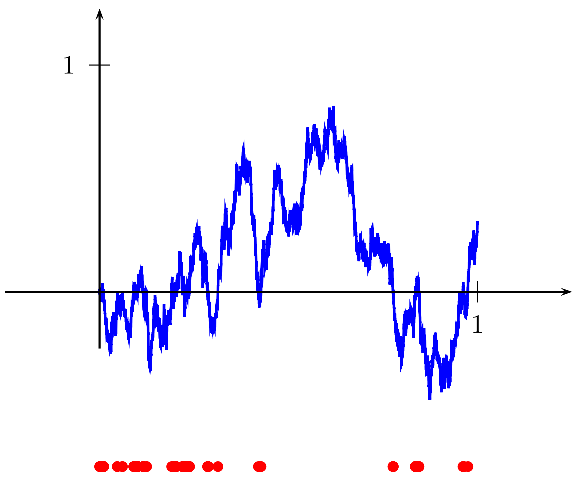


Figure 1.4: Brownian motion with corresponding level set Λ_0.

Moreover, by Fubini's Theorem one easily verifies that Λ_a has almost surely Lebesgue measure zero. In fact, continuity of $t \mapsto B_t(\omega)$ for any ω implies that $(t, \omega) \mapsto B_t(\omega)$ is product measurable (Exercise). Hence $\{(t, \omega) : B_t(\omega) = a\}$ is contained in the product σ-algebra, and

$$E[\lambda(\Lambda_a)] \quad = \quad E\left[\int_0^\infty I_{\{a\}}(B_t)\,dt\right] \quad = \quad \int_0^\infty P[B_t = a]\,dt \quad = \quad 0.$$

Theorem 1.22 (Unbounded oscillations, recurrence).

$$P\left[\sup_{t\geq 0} B_t \;=\; +\infty\right] \quad = \quad P\left[\inf_{t\geq 0} B_t \;=\; -\infty\right] \quad = \quad 1.$$

In particular, for any $a \in \mathbb{R}$, the random set Λ_a is almost surely unbounded, i.e. Brownian motion is recurrent.

Proof. By scale invariance,

$$\sup_{t\geq 0} B_t \quad \sim \quad c^{-1/2} \sup_{t\geq 0} B_{ct} \quad = \quad c^{-1/2} \sup_{t\geq 0} B_t \qquad \text{for any } c > 0.$$

Hence,

$$P\left[\sup_{t\geq 0} B_t \geq a\right] \;=\; P\left[\sup_{t\geq 0} B_t \geq a \cdot \sqrt{c}\right]$$

for any $c > 0$, and therefore $\sup B_t \in \{0, \infty\}$ almost surely. The first part of the assertion now follows since $\sup B_t$ is almost surely strictly positive. By reflection symmetry, we also obtain $\inf B_t = -\infty$ with probability one. ☐

The last Theorem makes a statement on the global structure of the set Λ_a. By invariance w.r.t. time inversion this again translates into a local regularity result:

Theorem 1.23 (Fine structure of Λ_a)**.** *The set Λ_a is almost surely a* ***perfect set****, i.e., any $t \in \Lambda_a$ is an accumulation point of Λ_a.*

Proof. We prove the statement for $a = 0$, the general case being left as an exercise. We proceed in three steps:

STEP 1: 0 *is almost surely an accumulation point of* Λ_0*:* This holds by time-reversal. Setting $\widehat{B}_t = t \cdot B_{1/t}$, we see that 0 is an accumulation point of Λ_0 if and only of for any $n \in \mathbb{N}$ there exists $t > n$ such that $\widehat{B}_t = 0$, i.e., if and only if the zero set of $\widehat{B}_t$ is unbounded. By Theorem 1.22, this holds almost surely.

STEP 2: *For any* $s \geq 0$, $T_s := \min(\Lambda_a \cap [s, \infty)) = \min\{t \geq s : B_t = a\}$ *is almost surely an accumulation point of* Λ_a*:* For the proof we need the strong Markov property of Brownian motion which will be proved in the next section. By Theorem 1.22, the random variable T_s is almost surely finite. Hence, by continuity, $B_{T_s} = a$ almost surely. The strong Markov property says that the process

$$\widetilde{B}_t := B_{T_s+t} - B_{T_s}, \qquad t \geq 0,$$

is again a Brownian motion starting at 0. Therefore, almost surely, 0 is an accumulation point of the zero set of $\widetilde{B}_t$ by Step 1. The claim follows since almost surely

$$\{t \geq 0 : \widetilde{B}_t = 0\} = \{t \geq 0 : B_{T_s+t} = B_{T_s}\} = \{t \geq T_s : B_t = a\} \subseteq \Lambda_a.$$

STEP 3: To complete the proof note that we have shown that the following properties hold with probability one:

(1). Λ_a is closed.

(2). $\min(\Lambda_a \cap [s, \infty))$ is an accumulation point of Λ_a for any $s \in \mathbb{Q}_+$.

Since $\mathbb{Q}_+$ is a dense subset of $\mathbb{R}_+$, (1) and (2) imply that any $t \in \Lambda_a$ is an accumulation point of Λ_a. In fact, for any $s \in [0, t] \cap \mathbb{Q}$, there exists an accumulation point of Λ_a in $(s, t]$ by (2), and hence t is itself an accumulation point.

□

Remark. It can be shown that the set Λ_a has Hausdorff dimension $1/2$.

1.5 Strong Markov property and reflection principle

In this section we prove a strong Markov property for Brownian motion. Before, we give another motivation for our interest in an extension of the Markov property to random times.

Maximum of Brownian motion

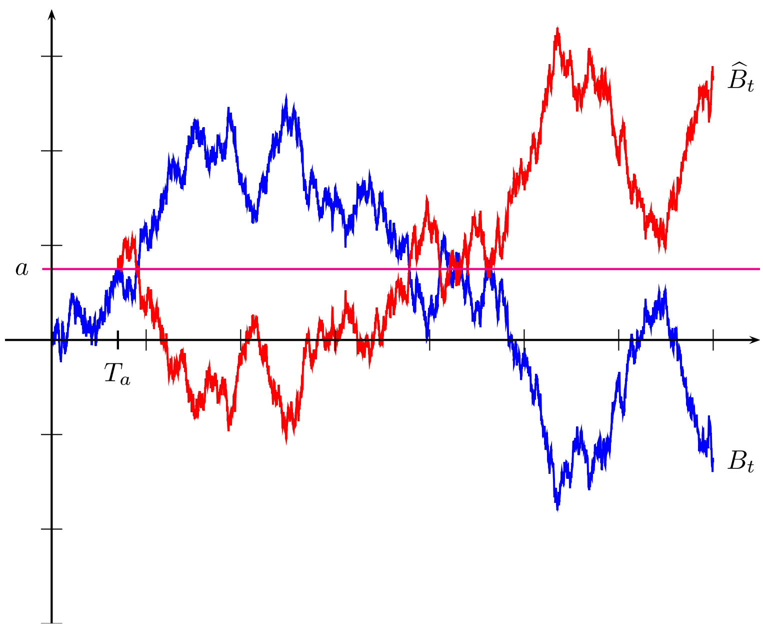
Suppose that $(B_t)_{t\geq 0}$ is a one-dimensional continuous Brownian motion starting at 0 defined on a probability space $(\Omega, \mathcal{A}, P)$. We would like to compute the distribution of the maximal value

$$M_s \quad = \quad \max_{t\in[0,s]} B_t$$

attained before a given time $s \in \mathbb{R}_+$. The idea is to proceed similarly as for Random Walks, and to reflect the Brownian path after the first passage time

$$T_a \quad = \quad \min\{t \geq 0 : B_t = a\}$$

to a given level $a > 0$:



It seems plausible (e.g. by the heuristic path integral representation of Wiener measure, or by a Random Walk approximation) that the reflected process $(\widehat{B}_t)_{t\geq 0}$ defined by

$$\widehat{B}_t \quad := \quad \begin{cases} B_t & \text{for } t \leq T_a \\ a-(B_t-a) & \text{for } t > T_a \end{cases}$$

is again a Brownian motion. At the end of this section, we will prove this reflection principle rigorously by the strong Markov property. Assuming the reflection principle is true, we can compute the distribution of M_s in the following way:

$$\begin{aligned} P[M_s \geq a] &= P[M_s \geq a, B_s \leq a] + P[M_s \geq a, B_s > a] \\ &= P[\widehat{B}_s \geq a] + P[B_s > a] \\ &= 2 \cdot P[B_s \geq a] \\ &= P[|B_s| \geq a]. \end{aligned}$$

Thus M_s has the same distribution as $|B_s|$.
Furthermore, since $M_s \geq a$ if and only if $\widehat{M_s} = \max\{\widehat{B}_t : t \in [0,s]\} \geq a$, we obtain the stronger statement

$$\begin{aligned} P[M_s \geq a, B_s \leq c] &= P[\widehat{M_s} \geq a, \widehat{B}_s \geq 2a-c] \quad = \quad P[\widehat{B}_s \geq 2a-c] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} \exp(-x^2/2s)\, dx \end{aligned}$$

for any $a \geq 0$ and $c \leq a$. As a consequence, we have:

Theorem 1.24 (Maxima of Brownian paths)**.**

(1). For any $s \geq 0$, the distribution of M_s is absolutely continuous with density

$$f_{M_s}(x) \quad = \quad \frac{2}{\sqrt{2\pi s}} \exp(-x^2/2s) \cdot I_{(0,\infty)}(x).$$

(2). The joint distribution of M_s and B_s is absolutely continuous with density

$$f_{M_s,B_s}(x,y) \quad = \quad 2\,\frac{2x-y}{\sqrt{2\pi s^3}} \exp\left(-\frac{(2x-y)^2}{2s}\right) I_{(0,\infty)}(x) I_{(-\infty,x)}(y).$$

Proof. (1) holds since $M_s \sim |B_s|$. For the proof of (2) we assume w.l.o.g. $s = 1$. The general case can be reduced to this case by the scale invariance of Brownian motion (Exercise). For $a \geq 0$ and $c \leq a$ let

$$G(a,c) \quad := \quad P[M_1 \geq a, B_1 \leq c].$$

By the reflection principle,

$$G(a,c) \quad = \quad P[B_1 \geq 2a - c] \quad = \quad 1 - \Phi(2a - c),$$

where Φ denotes the standard normal distribution function. Since $\lim\limits_{a\to\infty} G(a,c) = 0$ and $\lim\limits_{c\to-\infty} G(a,c) = 0$, we obtain

$$\begin{aligned} P[M_1 \geq a, B_1 \leq c] \;=\; G(a,c) \quad &= \quad -\int_{x=a}^{\infty}\int_{y=-\infty}^{c} \frac{\partial^2 G}{\partial x \partial y}(x,y)\, dydx \\ &= \quad \int_{x=a}^{\infty}\int_{y=-\infty}^{c} 2 \cdot \frac{2x - y}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(2x-y)^2}{2}\right)\, dydx. \end{aligned}$$

This implies the claim for $s = 1$, since $M_1 \geq 0$ and $B_1 \leq M_1$ by definition of M_1. □

The Theorem enables us to compute the distributions of the first passage times T_a. In fact, for $a > 0$ and $s \in [0, \infty)$ we obtain

$$\begin{aligned} P[T_a \leq s] \;=\; P[M_s \geq a] \quad &= \quad 2 \cdot P[B_s \geq a] \;=\; 2 \cdot P[B_1 \geq a/\sqrt{s}] \\ &= \quad \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{s}}^{\infty} e^{-x^2/2}\, dx. \end{aligned} \tag{1.5.1}$$

Corollary 1.25 (Distribution of T_a). *For any $a \in \mathbb{R} \setminus \{0\}$, the distribution of T_a is absolutely continuous with density*

$$f_{T_a}(s) \quad = \quad \frac{|a|}{\sqrt{2\pi s^3}} \cdot e^{-a^2/2s}.$$

Proof. For $a > 0$, we obtain

$$f_{T_a}(s) \quad = \quad F'_{T_a}(s) \quad = \quad \frac{a}{\sqrt{2\pi s^3}} e^{-a^2/2s}$$

by (1.5.1). For $a < 0$ the assertion holds since $T_a \sim T_{-a}$ by reflection symmetry of Brownian motion. □

Next, we prove a strong Markov property for Brownian motion. Below we will then complete the proof of the reflection principle and the statements above by applying the strong Markov property to the passage time T_a.

Strong Markov property for Brownian motion

Suppose again that $(B_t)_{t\geq 0}$ is a d-dimensional continuous Brownian motion starting at 0 on a probability space $(\Omega, \mathcal{A}, P)$, and let

$$\mathcal{F}_t^B \quad = \quad \sigma(B_s \, : \, 0 \leq s \leq t), \qquad t \geq 0,$$

denote the σ-algebras generated by the process up to time t.

Definition. *A random variable $T : \Omega \to [0, \infty]$ is called an $(\mathcal{F}_t^B)$-**stopping time** if and only if*

$$\{T \leq t\} \in \mathcal{F}_t^B \qquad \textit{for any } t \geq 0.$$

Example. Clearly, for any $a \in \mathbb{R}$, the first passage time

$$T_a \; = \; \min\{t \geq 0 \, : \, B_t = a\}$$

to a level a is an $(\mathcal{F}_t^B)$-stopping time.

The σ-algebra $\mathcal{F}_T^B$ describing the information about the process up to a stopping time T is defined by

$$\mathcal{F}_T^B \; = \; \{A \in \mathcal{A} \, : \, A \cap \{T \leq t\} \in \mathcal{F}_t^B \quad \text{for any } t \geq 0\}.$$

Note that for $(\mathcal{F}_t^B)$ stopping times S and T with $S \leq T$ we have $\mathcal{F}_S^B \subseteq \mathcal{F}_T^B$, since for $t \geq 0$

$$A \cap \{S \leq t\} \in \mathcal{F}_t^B \qquad \Longrightarrow \qquad A \cap \{T \leq t\} \;=\; A \cap \{S \leq t\} \cap \{T \leq t\} \;\in\; \mathcal{F}_t^B.$$

For any constant $s \in \mathbb{R}_+$, the process $(B_{s+t} - B_s)_{t \geq 0}$ is a Brownian motion independent of $\mathcal{F}_s^B$.

A corresponding statement holds for stopping times:

Theorem 1.26 (Strong Markov property). *Suppose that T is an almost surely finite $(\mathcal{F}_t^B)$ stopping time. Then the process $(\widetilde{B}_t)_{t \geq 0}$ defined by*

$$\widetilde{B}_t \;=\; B_{T+t} - B_T \quad \text{if } T < \infty, \qquad 0 \quad \text{otherwise,}$$

is a Brownian motion independent of $\mathcal{F}_T^B$.

Proof. We first assume that T takes values only in $C \cup \{\infty\}$ where C is a countable subset of $[0, \infty)$. Then for $A \in \mathcal{F}_T^B$ and $s \in C$, we have $A \cap \{T = s\} \in \mathcal{F}_s^B$ and $\widetilde{B}_t = B_{t+s} - B_s$ on $A \cap \{T = s\}$. By the Markov property, $(B_{t+s} - B_s)_{t \geq 0}$ is a Brownian motion independent of $\mathcal{F}_s^B$. Hence for any measurable subset Γ of $C([0, \infty], \mathbb{R}^d)$, we have

$$\begin{aligned} P[\{(\widetilde{B}_t)_{t \geq 0} \in \Gamma\} \cap A] &= \sum_{s \in C} P[\{(B_{t+s} - B_s)_{t \geq 0} \in \Gamma\} \cap A \cap \{T = s\}] \\ &= \sum_{s \in C} \mu_0[\Gamma] \cdot P[A \cap \{T = s\}] \quad = \quad \mu_0[\Gamma] \cdot P[A] \end{aligned}$$

where μ_0 denotes the distribution of Brownian motion starting at 0. This proves the assertion for discrete stopping times.

For an arbitrary $(\mathcal{F}_t^B)$ stopping time T that is almost surely finite and $n \in \mathbb{N}$, we set $T_n = \frac{1}{n} \lceil nT \rceil$, i.e.,

$$T_n \;=\; \frac{k}{n} \quad \text{on} \quad \left\{ \frac{k-1}{n} < T \leq \frac{k}{n} \right\} \quad \text{for any } k \in \mathbb{N}.$$

Since the event $\{T_n = k/n\}$ is $\mathcal{F}^B_{k/n}$-measurable for any $k \in \mathbb{N}$, T_n is a discrete $(\mathcal{F}^B_t)$ stopping time. Therefore, $(B_{T_n+t} - B_{T_n})_{t\geq 0}$ is a Brownian motion that is independent of $\mathcal{F}^B_{T_n}$, and hence of the smaller σ-algebra $\mathcal{F}^B_T$. As $n \to \infty$, $T_n \to T$, and thus, by continuity,

$$\widetilde{B}_t \;=\; B_{T+t} - B_T \;=\; \lim_{n\to\infty}(B_{T_n+t} - B_{T_n}).$$

Now it is easy to verify that $(\widetilde{B}_t)_{t\geq 0}$ is again a Brownian motion that is independent of $\mathcal{F}^B_T$. □

A rigorous reflection principle

We now apply the strong Markov property to prove a reflection principle for Brownian motion. Consider a one-dimensional continuous Brownian motion $(B_t)_{t\geq 0}$ starting at 0. For $a \in \mathbb{R}$ let

$$\begin{array}{rcll} T_a &=& \min\{t \geq 0 \,:\, B_t = a\} & \text{(first passage time),} \\ B^{T_a}_t &=& B_{\min\{t,T_a\}} & \text{(process stopped at } T_a\text{),} \quad \text{and} \\ \widetilde{B}_t &=& B_{T_a+t} - B_{T_a} & \text{(process after } T_a\text{).} \end{array}$$

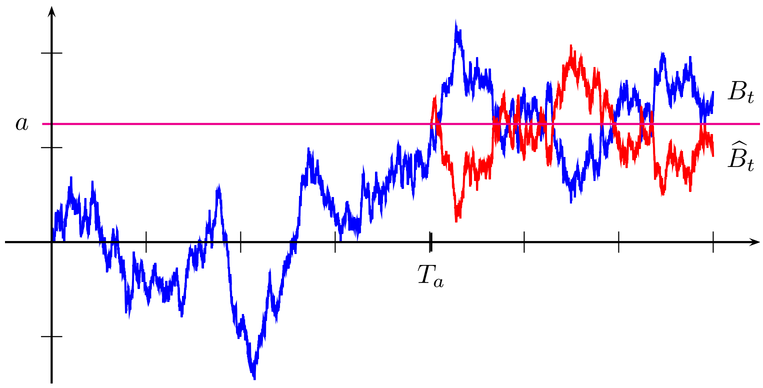
Theorem 1.27 (Reflection principle). *The joint distributions of the following random variables with values in* $\mathbb{R}_+ \times C([0,\infty)) \times C([0,\infty))$ *agree:*

$$(T_a, (B^{T_a}_t)_{t\geq 0}, (\widetilde{B}_t)_{t\geq 0}) \quad \sim \quad (T_a, (B^{T_a}_t)_{t\geq 0}, (-\widetilde{B}_t)_{t\geq 0})$$

Proof. By the strong Markov property, the process $\widetilde{B}$ is a Brownian motion starting at 0 independent of $\mathcal{F}_{T_a}$, and hence of T_a and $B^{T_a} = (B^{T_a}_t)_{t\geq 0}$. Therefore,

$$P \circ (T_a, B^{T_a}, \widetilde{B})^{-1} \;=\; P \circ (T_a, B^{T_a})^{-1} \otimes \mu_0 \;=\; P \circ (T_a, B^{T_a}, -\widetilde{B})^{-1}.$$

□



As a consequence of the theorem, we can complete the argument given at the beginning of this section: The "shadow path" $\widehat{B}_t$ of a Brownian path B_t with reflection when reaching the level a is given by

$$\widehat{B}_t = \begin{cases} B_t^{T_a} & \text{for } t \le T_a \\ a - \widetilde{B}_{t-T_a} & \text{for } t > T_a \end{cases},$$

whereas

$$B_t = \begin{cases} B_t^{T_a} & \text{for } t \le T_a \\ a + \widetilde{B}_{t-T_a} & \text{for } t > T_a \end{cases}.$$

By the Theorem 1.27, $(\widehat{B}_t)_{t\ge 0}$ has the same distribution as $(B_t)_{t\ge 0}$. Therefore, and since $\max_{t\in[0,s]} B_t \ge a$ if and only if $\max_{t\in[0,s]} \widehat{B}_t \ge a$, we obtain for $a \ge c$:

$$\begin{aligned} P\left[\max_{t\in[0,s]} B_t \ge a, B_s \le c\right] &= P\left[\max_{t\in[0,s]} \widehat{B}_t \ge a, \widehat{B}_s \ge 2a - c\right] \\ &= P\left[\widehat{B}_s \ge 2a - c\right] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} e^{-x^2/2s}\, dx. \end{aligned}$$

Chapter 2

Martingales in discrete time

Classical analysis starts with studying convergence of sequences of real numbers. Similarly, stochastic analysis relies on basic statements about sequences of real-valued random variables. Any such sequence can be decomposed uniquely into a martingale, i.e., a real.valued stochastic process that is "constant on average", and a predictable part. Therefore, estimates and convergence theorems for martingales are crucial in stochastic analysis.

2.1 Definitions and examples

We fix a probability space $(\Omega, \mathcal{A}, P)$. Moreover, we assume that we are given an increasing sequence $\mathcal{F}_n$ $(n = 0, 1, 2, \ldots)$ of sub-σ-algebras of $\mathcal{A}$. Intuitively, we often think of $\mathcal{F}_n$ as describing the information available to us at time n. Formally, we define:

Definition (Filtration, adapted process). *(1). A **filtration** on $(\Omega, \mathcal{A})$ is an increasing sequence*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$$

of σ-algebras $\mathcal{F}_n \subseteq \mathcal{A}$.

*(2). A stochastic process $(X_n)_{n\geq 0}$ is **adapted** to a filtration $(\mathcal{F}_n)_{n\geq 0}$ iff each X_n is $\mathcal{F}_n$-measurable.*

Example. (1). The **canonical filtration** $(\mathcal{F}_n^X)$ generated by a stochastic process (X_n) is given by

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \dots, X_n).$$

If the filtration is not specified explicitly, we will usually consider the canonical filtration.

(2). Alternatively, filtrations containing additional information are of interest, for example the filtration

$$\mathcal{F}_n = \sigma(Z, X_0, X_1, \dots, X_n)$$

generated by the process (X_n) and an additional random variable Z, or the filtration

$$\mathcal{F}_n = \sigma(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n)$$

generated by the process (X_n) and a further process (Y_n).

Clearly, the process (X_n) is adapted to any of these filtrations. In general, (X_n) is adapted to a filtration $(\mathcal{F}_n)$ if and only if $\mathcal{F}_n^X \subseteq \mathcal{F}_n$ for any $n \geq 0$.

Martingales and supermartingales

We can now formalize the notion of a real-valued stochastic process that is constant (respectively decreasing or increasing) on average:

Definition (Martingale, supermartingale, submartingale). *(1). A sequence of real-valued random variables $M_n : \Omega \to \mathbb{R} \quad (n = 0, 1, \dots)$ on the probability space $(\Omega, \mathcal{A}, P)$ is called a **martingale w.r.t. the filtration** $(\mathcal{F}_n)$ if and only if*

(a) (M_n) is adapted w.r.t. $(\mathcal{F}_n)$,

(b) M_n is integrable for any $n \geq 0$, and

(c) $E[M_n \mid \mathcal{F}_{n-1}] = M_{n-1}$ for any $n \in \mathbb{N}$.

*(2). Similarly, (M_n) is called a **supermartingale** (resp. a **submartingale**) w.r.t. $(\mathcal{F}_n)$ if and only if (a) holds, the positive part M_n^+ (resp. the negative part M_n^-) is integrable for any $n \geq 0$, and (c) holds with "$=$" replaced by "$\leq$", "$\geq$" respectively.*

Condition (c) in the martingale definition can equivalently be written as

(c') $E[M_{n+1} - M_n \mid \mathcal{F}_n] \;=\; 0$ for any $n \in \mathbb{Z}_+$,

and correspondingly with "$=$" replaced by "$\leq$" or "$\geq$" for super- or submartingales.

Intuitively, a martingale is a "fair game´´, i.e., M_{n-1} is the best prediction (w.r.t. the mean square error) for the next value M_n given the information up to time $n-1$. A supermartingale is *"decreasing on average"*, a submartingale is *"increasing on average"*, and a martingale is both "decreasing" and "increasing", i.e., *"constant on average"*. In particular, by induction on n, a martingale satisfies

$$E[M_n] = E[M_0] \qquad \text{for any } n \geq 0.$$

Similarly, for a supermartingale, the expectation values $E[M_n]$ are decreasing. More generally, we have:

Lemma 2.1. *If (M_n) is a martingale (respectively a supermartingale) w.r.t. a filtration $(\mathcal{F}_n)$ then*

$$E[M_{n+k} \mid \mathcal{F}_n] \overset{(\leq)}{=} M_n \qquad P\text{-almost surely for any } n, k \geq 0.$$

Proof. By induction on k: The assertion holds for $k = 0$, since M_n is $\mathcal{F}_n$-measurable. Moreover, the assertion for $k-1$ implies

$$\begin{aligned} E[M_{n+k} \mid \mathcal{F}_n] &= E\big[E[M_{n+k} \mid \mathcal{F}_{n+k-1}] \,\big|\, \mathcal{F}_n\big] \\ &= E[M_{n+k-1} \mid \mathcal{F}_n] \;=\; M_n \qquad P\text{-a.s.} \end{aligned}$$

by the tower property for conditional expectations. □

Remark (Supermartingale Convergence Theorem). A key fact in analysis is that any lower bounded decreasing sequence of real numbers converges to its infimum. The counterpart of this result in stochastic analysis is the Supermartingale Convergence Theorem: Any lower bounded supermartingale converges almost surely, cf. Theorem 4.5 below.

Some fundamental examples

a) Sums of independent random variables

A Random Walk

$$S_n \;=\; \sum_{i=1}^{n} \eta_i, \qquad n = 0, 1, 2, \ldots,$$

with independent increments $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale w.r.t. to the filtration

$$\mathcal{F}_n \;=\; \sigma(\eta_1, \ldots, \eta_n) \;=\; \sigma(S_0, S_1, \ldots, S_n)$$

if and only if the increments η_i are centered random variables. In fact, for any $n \in \mathbb{N}$,

$$E[S_n - S_{n-1} \,|\, \mathcal{F}_{n-1}] \;=\; E[\eta_n \,|\, \mathcal{F}_{n-1}] \;=\; E[\eta_n]$$

by independence of the increments. Correspondingly, (S_n) is an $(\mathcal{F}_n)$ supermartingale if and only if $E[\eta_i] \leq 0$ for any $i \in \mathbb{N}$.

b) Products of independent non-negative random variables

A stochastic process

$$M_n \;=\; \prod_{i=1}^{n} Y_i, \qquad n = 0, 1, 2, \ldots,$$

with independent non-negative factors $Y_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale respectively a supermartingale w.r.t. the filtration

$$\mathcal{F}_n \;=\; \sigma(Y_1, \ldots, Y_n)$$

if and only if $E[Y_i] = 1$ for any $i \in \mathbb{N}$, or $E[Y_i] \leq 1$ for any $i \in \mathbb{N}$ respectively. In fact, as M_n is $\mathcal{F}_n$-measurable and Y_{n+1} is independent of $\mathcal{F}_n$, we have

$$E[M_{n+1} \,|\, \mathcal{F}_n] \;=\; E[M_n \cdot Y_{n+1} \,|\, \mathcal{F}_n] \;=\; M_n \cdot E[Y_{n+1}] \qquad \text{for any } n \geq 0.$$

Martingales and supermartingales of this type occur naturally in stochastic growth models.

Example (**Exponential martingales**). Consider a Random Walk $S_n = \sum_{i=1}^n \eta_i$ with i.i.d. increments η_i, and let

$$Z(\lambda) \;=\; E[\exp(\lambda\eta_i)], \qquad \lambda \in \mathbb{R},$$

denote the moment generating function of the increments. Then for any $\lambda \in \mathbb{R}$ with $Z(\lambda) < \infty$, the process

$$M_n^\lambda \;:=\; e^{\lambda S_n}/Z(\lambda)^n \;=\; \prod_{i=1}^n \big(e^{\lambda\eta_i}/Z(\lambda)\big)$$

is a martingale. This martingale can be used to prove exponential bounds for Random Walks, cf. e.g. Chernov's theorem ["Einführung in die Wahrscheinlichkeitstheorie", Theorem 8.3].

Example (**CRR model of stock market**). In the Cox-Ross-Rubinstein binomial model of mathematical finance, the price of an asset is changing during each period either by a factor $1 + a$ or by a factor $1 + b$ with $a, b \in (-1, \infty)$ such that $a < b$. We can model the price evolution in a fixed number N of periods by a stochastic process

$$S_n = S_0 \cdot \prod_{i=1}^n X_i, \qquad n = 0, 1, 2, \ldots, N,$$

defined on $\Omega = \{1 + a, 1 + b\}^N$, where the initial price S_0 is a given constant, and $X_i(\omega) = \omega_i$. Taking into account a constant interest rate $r > 0$, the discounted stock price after n periods is

$$\widetilde{S}_n \;=\; S_n/(1+r)^n \;=\; S_0 \cdot \prod_{i=1}^n \frac{X_i}{1+r}.$$

A probability measure P on Ω is called a **martingale measure** if the discounted stock price is a martingale w.r.t. P and the filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Martingale measures are important for option pricing under no arbitrage assumptions, cf. Section 2.3 below. For $1 \le n \le N$,

$$E[\widetilde{S}_n \mid \mathcal{F}_{n-1}] \;=\; E\left[\widetilde{S}_{n-1} \cdot \frac{X_n}{1+r} \;\middle|\; \mathcal{F}_{n-1}\right] \;=\; \widetilde{S}_{n-1} \cdot \frac{E[X_n \mid \mathcal{F}_{n-1}]}{1+r}.$$

Hence $(\widetilde{S}_n)$ is an $(\mathcal{F}_n)$ martingale w.r.t. P if and only if

$$E[X_n \mid \mathcal{F}_{n-1}] \quad = \quad 1+r \qquad \text{for any } 1 \le n \le N. \tag{2.1.1}$$

On the other hand, since in the CRR model X_n only takes the values $1+a$ and $1+b$, we have

$$\begin{aligned} E[X_n \mid \mathcal{F}_{n-1}] &= (1+a) \cdot P[X_n = 1+a \mid \mathcal{F}_{n-1}] + (1+b) \cdot P[X_n = 1+b \mid \mathcal{F}_{n-1}] \\ &= 1+a+(b-a) \cdot P[X_n = 1+b \mid \mathcal{F}_{n-1}]. \end{aligned}$$

Therefore, by (2.1.1), $(\widetilde{S}_n)$ is a martingale if and only if

$$P[X_n = 1+b \mid \mathcal{F}_{n-1}] \; = \; \frac{r-a}{b-a} \qquad \text{for any } n = 1, \ldots, N,$$

i.e., if and only if the growth factors $X_1, \ldots, X_N$ are independent with

$$P[X_n = 1+b] \; = \; \frac{r-a}{b-a} \qquad \text{and} \qquad P[X_n = 1+a] \; = \; \frac{b-r}{b-a}. \tag{2.1.2}$$

Hence for $r \notin [a,b]$, a martingale measure does not exist, and for $r \in [a,b]$, the product measure P on Ω satisfying (2.1.2) is the unique martingale measure. Intuitively this is plausible: If $r < a$ or $r > b$ respectively, then the stock price is always growing more or less than the discount factor $(1+r)^n$, so the discounted stock price can not be a martingale. If, on the other hand, $a < r < b$, then $(\widetilde{S}_n)$ is a martingale provided the growth factors are independent with

$$\frac{P[X_n = 1+b]}{P[X_n = 1+a]} \quad = \quad \frac{(1+r)-(1+a)}{(1+b)-(1+r)}.$$

We remark, however, that uniqueness of the martingale measure only follows from (2.1.1) since we have assumed that each X_n takes only two possible values (binomial model). In a corresponding trinomial model there are infinitely many martingale measures!

c) Successive prediction values

Let F be an integrable random variable, and let $(\mathcal{F}_n)$ be a filtration on a probability space $(\Omega, \mathcal{A}, P)$. Then the process

$$M_n \; := \; E[F \mid \mathcal{F}_n], \qquad n = 0, 1, 2, \ldots,$$

of successive prediction values for F based on the information up to time n is a martingale. Indeed, by the tower property for conditional expectations, we have

$$E[M_n \mid \mathcal{F}_{n-1}] \;=\; E\big[E[F \mid \mathcal{F}_n] \,\big|\, \mathcal{F}_{n-1}\big] \;=\; E\big[F \,\big|\, \mathcal{F}_{n-1}\big] \;=\; M_{n-1}$$

almost surely for any $n \in \mathbb{N}$.

Remark (Representing martingales as successive prediction values). The class of martingales that have a representation as successive prediction values almost contains general martingales. In fact, for an arbitrary $(\mathcal{F}_n)$ martingale (M_n) and any finite integer $m \geq 0$, the representation

$$M_n \;=\; E[M_m \mid \mathcal{F}_n]$$

holds for any $n = 0, 1, \ldots, m$. Moreover, the L^1 Martingale Convergence Theorem implies that under a uniform integrability assumption, the limit $M_\infty = \lim\limits_{m\to\infty} M_m$ exists in $\mathcal{L}^1$, and the representation

$$M_n \;=\; E[M_\infty \mid \mathcal{F}_n]$$

holds for any $n \geq 0$, see Section 4.3 below .

d) Functions of martingales

By Jensen's inequality for conditional expectations, convex functions of martingales are submartingales, and concave functions of martingales are supermartingales:

Theorem 2.2 (Convex functions of martingales). *Suppose that $(M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)$ martingale, and $u : \mathbb{R} \to \mathbb{R}$ is a convex function that is bounded from below. Then $(u(M_n))$ is an $(\mathcal{F}_n)$ submartingale.*

Proof. Since u is lower bounded, $u(M_n)^-$ is integrable for any n. Jensen's inequality for conditional expectations now implies

$$E[u(M_{n+1}) \mid \mathcal{F}_n] \;\geq\; u\big(E[M_{n+1} \mid \mathcal{F}_n]\big) \;=\; u(M_n)$$

almost surely for any $n \geq 0$. □

Example. If (M_n) is a martingale then $(|M_n|^p)$ is a submartingale for any $p \geq 1$.

e) Functions of Markov chains

Let $p(x, dy)$ be a transition kernel on a measurable space $(S, \mathcal{B})$.

Definition (Markov chain, superharmonic function). *(1). A discrete time stochastic process $(X_n)_{n\geq 0}$ with state space $(S, \mathcal{B})$ defined on the probability space $(\Omega, \mathcal{A}, P)$ is called a **(time-homogeneous) Markov chain with transition kernel** p **w.r.t. the filtration** $(\mathcal{F}_\mathbf{n})$, if and only if*

(a) (X_n) is $(\mathcal{F}_n)$ adapted, and

(b) $P[X_{n+1} \in B \,|\, \mathcal{F}_n] \;=\; p(X_n, B)$ P-almost surely for any $B \in \mathcal{B}$ and $n \geq 0$.

*(2). A measurable function $h : S \to \mathbb{R}$ is called **superharmonic** (resp. **subharmonic**) w.r.t. p if and only if the integrals*

$$(ph)(x) \;:=\; \int p(x, dy)h(y), \qquad x \in S,$$

exist, and

$$(ph)(x) \leq h(x) \qquad \text{(respectively } (ph)(x) \geq h(x)\text{)}$$

holds for any $x \in S$.
*The function h is called **harmonic** iff it is both super- and subharmonic, i.e., iff*

$$(ph)(x) \;=\; h(x) \qquad \text{for any } x \in S.$$

By the tower property for conditional expectations, any $(\mathcal{F}_n)$ Markov chain is also a Markov chain w.r.t. the canonical filtration generated by the process.

Example (Classical Random Walk on $\mathbb{Z}^d$). The standard Random Walk $(X_n)_{n\geq 0}$ on $\mathbb{Z}^d$ is a Markov chain w.r.t. the filtration $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$ with transition probabilities $p(x, x+e) = 1/2d$ for any unit vector $e \in \mathbb{Z}^d$. The coordinate processes $(X_n^i)_{n\geq 0}$, $i = 1, \ldots, d$, are Markov chains w.r.t. the same filtration with transition probabilities

$$\overline{p}(x, x+1) \;=\; \overline{p}(x, x-1) \;=\; \frac{1}{2d}, \quad \overline{p}(x, x) \;=\; \frac{2d-2}{2d}.$$

A function $h : \mathbb{Z}^d \to \mathbb{R}$ is superharmonic w.r.t. p if and only if

$$\Delta_{\mathbb{Z}^d} h(x) \;=\; = \; \sum_{i=1}^{d} \big(h(x+e_i) - 2h(x) + h(x-e_i)\big) \;=\; 2d\,((ph)(x) - h(x)) \;\leq\; 0$$

for any $x \in \mathbb{Z}^d$.
A function $h : \mathbb{Z} \to \mathbb{R}$ is harmonic w.r.t. $\overline{p}$ if and only if $h(x) = ax + b$ with $a, b \in \mathbb{R}$, and h is superharmonic if and only if it is concave.

It is easy to verify that (super-)harmonic functions of Markov chains are (super-)martingales:

Theorem 2.3 (**Superharmonic functions of Markov chains are supermartingales**). *Suppose that (X_n) is an $(\mathcal{F}_n)$ Markov chain. Then the real-valued process*

$$M_n \;:=\; h(X_n), \quad n = 0, 1, 2, \ldots,$$

is a martingale (resp. a supermartingale) w.r.t. $(\mathcal{F}_n)$ for every harmonic (resp. superharmonic) function $h : S \to \mathbb{R}$ such that $h(X_n)$ (resp. $h(X_n)^+$) is integrable for all n.

Proof. Clearly, (M_n) is again $(\mathcal{F}_n)$ adapted. Moreover,

$$E[M_{n+1} \,|\, \mathcal{F}_n] \;=\; E[h(X_{n+1}) \,|\, \mathcal{F}_n] \;=\; (ph)(X_n) \qquad P\text{-a.s.}$$

The assertion now follows immediately from the definitions. □

Below, we will show how to construct more general martingales from Markov chains, cf. Theorem 2.5. At first, however, we consider a simple example that demonstrates the usefulness of martingale methods in analyzing Markov chains:

Example (**Wright model for evolution**). In the Wright model for a population of N individuals (replicas) with a finite number of possible types, each individual in generation $n + 1$ inherits a type from a randomly chosen predecessor in the n th generation.

The number X_n of individuals of a given type in generation n is a Markov chain with state space $S = \{0, 1, \ldots, N\}$ and transition kernel

$$p(k, \bullet) \;=\; \mathrm{Bin}(N, k/N).$$

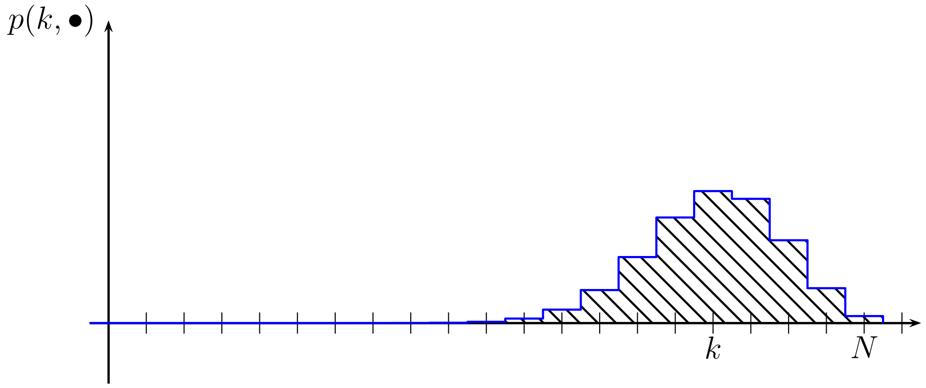


Figure 2.1: Transition function of (X_n).

Moreover, as the average of this binomial distribution is k, the function $h(x) = x$ is harmonic, and the expected number of individuals in generation $n+1$ given $X_0, \ldots, X_n$ is

$$E[X_{n+1} \mid X_0, \ldots, X_n] \;=\; X_n.$$

Hence, the process (X_n) is a bounded martingale. The Martingale Convergence Theorem now implies that the limit $X_\infty = \lim X_n$ exists almost surely, cf. Section 4.2 below. Since X_n takes discrete values, we can conclude that $X_n = X_\infty$ eventually with probability one. In particular, X_∞ is almost surely an absorbing state. Hence

$$P\big[X_n = 0 \;\text{ or }\; X_n = N \;\text{ eventually}\big] \;=\; 1. \tag{2.1.3}$$

In order to compute the probabilities of the events "$X_n = 0$ eventually" and "$X_n = N$ eventually" we can apply the Optional Stopping Theorem for martingales, cf. Section 2.3 below. Let

$$T \;:=\; \min\{n \geq 0 \;:\; X_n = 0 \;\text{ or }\; X_n = N\}, \qquad \min\emptyset \;:=\; \infty,$$

denote the first hitting time of the absorbing states. If the initial number X_0 of individuals of the given type is k, then by the Optional Stopping Theorem,

$$E[X_T] \;=\; E[X_0] \;=\; k.$$

Hence by (2.1.3) we obtain

$$P\big[X_n = N \text{ eventually}\big] = P[X_T = N] = \frac{1}{N}E[X_T] = \frac{k}{N}, \qquad \text{and}$$
$$P\big[X_n = 0 \text{ eventually}\big] = 1 - \frac{k}{N} = \frac{N-k}{N}.$$

Hence eventually all individuals have the same type, and a given type occurs eventually with probability determined by its initial relative frequency in the population.

2.2 Doob Decomposition and Martingale Problem

We will show now that any adapted sequence of real-valued random variables can be decomposed into a martingale and a predictable process. In particular, the variance process of a martingale (M_n) is the predictable part in the corresponding Doob decomposition of the process (M_n^2). The Doob decomposition for functions of Markov chains implies the martingale problem characterization of Markov chains.

Doob Decomposition

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(\mathcal{F}_n)_{n\geq 0}$ a filtration on $(\Omega, \mathcal{A})$.

Definition (Predictable process). *A stochastic process $(A_n)_{n\geq 0}$ is called **predictable w.r.t.** $(\mathcal{F}_\mathbf{n})$ if and only if A_0 is constant and A_n is measurable w.r.t. $\mathcal{F}_{n-1}$ for any $n \in \mathbb{N}$.*

Intuitively, the value $A_n(\omega)$ of a predictable process can be predicted by the information available at time $n-1$.

Theorem 2.4 (Doob decomposition). *Every $(\mathcal{F}_n)$ adapted sequence of integrable random variables Y_n $(n \geq 0)$ has a unique decomposition (up to modification on null sets)*

$$Y_n = M_n + A_n \tag{2.2.1}$$

into an $(\mathcal{F}_n)$ martingale (M_n) and a predictable process (A_n) such that $A_0 = 0$. Explicitly, the decomposition is given by

$$A_n \quad = \quad \sum_{k=1}^{n} E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}], \qquad \textit{and} \quad M_n \;=\; Y_n - A_n. \tag{2.2.2}$$

Remark. (1). The increments $E[Y_k - Y_{k-1} | \mathcal{F}_{k-1}]$ of the process (A_n) are the predicted increments of (Y_n) given the previous information.

(2). The process (Y_n) is a supermartingale (resp. a submartingale) if and only if the predictable part (A_n) is decreasing (resp. increasing).

Proof of Theorem 2.4. Uniqueness: For any decomposition as in (2.2.1) we have

$$Y_k - Y_{k-1} \quad = \quad M_k - M_{k-1} + A_k - A_{k-1} \qquad \text{for any } k \in \mathbb{N}.$$

If (M_n) is a martingale and (A_n) is predictable then

$$E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad E[A_k - A_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad A_k - A_{k-1} \qquad P\text{-a.s.}$$

This implies that (2.2.2) holds almost surely if $A_0 = 0$.

Existence: Conversely, if (A_n) and (M_n) are defined by (2.2.2) then (A_n) is predictable with $A_0 = 0$ and (M_n) is a martingale, since

$$E[M_k - M_{k-1} \,|\, \mathcal{F}_{k-1}] \;=\; 0 \qquad P\text{-a.s. for any } k \in \mathbb{N}.$$

□

Conditional Variance Process

Consider a martingale (M_n) such that M_n is square integrable for any $n \geq 0$. Then, by Jensen's inequality, (M_n^2) is a submartingale and can again be decomposed into a martingale $(\widetilde{M}_n)$ and a predictable process $\langle M \rangle_n$ such that $\langle M \rangle_0 = 0$:

$$M_n^2 \quad = \quad \widetilde{M}_n \;+\; \langle M \rangle_n \qquad \text{for any } n \geq 0.$$

The increments of the predictable process are given by

$$\begin{aligned}\langle M\rangle_k - \langle M\rangle_{k-1} &= E[M_k^2 - M_{k-1}^2 \mid \mathcal{F}_{k-1}] \\ &= E\big[(M_k - M_{k-1})^2 \,\big|\, \mathcal{F}_{k-1}\big] + 2 \cdot E\big[M_{k-1} \cdot (M_k - M_{k-1}) \,\big|\, \mathcal{F}_{k-1}\big] \\ &= \mathrm{Var}\big[M_k - M_{k-1} \,\big|\, \mathcal{F}_{k-1}\big] \qquad \text{for any } k \in \mathbb{N}.\end{aligned}$$

Here we have used in the last step that $E[M_k - M_{k-1} \mid \mathcal{F}_{k-1}]$ vanishes since (M_n) is a martingale.

Definition (Conditional variance process). *The predictable process*

$$\langle M\rangle_n \quad := \quad \sum_{k=1}^{n} \mathrm{Var}\left[M_k - M_{k-1} \mid \mathcal{F}_{k-1}\right], \qquad n \geq 0,$$

is called the ***conditional variance process*** *of the square integrable martingale (M_n).*

Example (Random Walks). If $M_n = \sum_{i=1}^n \eta_i$ is a sum of independent centered random variables η_i and $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$ then the conditional variance process is given by $\langle M\rangle_n = \sum_{i=1}^n \mathrm{Var}[\eta_i]$.

The conditional variance process is crucial for generalizations of classical limit theorems such as the Law of Large Numbers or the Central Limit Theorem from sums of independent random variables to martingales. A direct consequence of the fact that $M_n^2 - \langle M\rangle_n$ is a martingale is that

$$E[M_n^2] \;=\; E[M_0^2] + E[\langle M\rangle_n] \qquad \text{for any } n \geq 0.$$

This can often be used to derive L^2-estimates for martingales.

Example (Discretizations of stochastic differential equations). Consider an ordinary differential equation

$$\frac{dX_t}{dt} \;=\; b(X_t), \qquad t \geq 0, \tag{2.2.3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a given vector field. In order to take into account unpredictable effects on a system, one is frequently interested in studying random perturbations of the dynamics (2.2.3) of type

$$dX_t \quad = \quad b(X_t)\, dt + \text{“noise”} \tag{2.2.4}$$

with a random noise term. The solution $(X_t)_{t\geq 0}$ of such a stochastic differential equation (SDE) is a stochastic process in continuous time defined on a probability space $(\Omega, \mathcal{A}, P)$ where also the random variables describing the noise effects are defined. The vector field b is called the (deterministic) “drift”. We will make sense of general SDE later, but we can already consider time discretizations.

For simplicity let us assume $d = 1$. Let $b, \sigma : \mathbb{R} \to \mathbb{R}$ be continuous functions, and let $(\eta_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ describing the noise effects. We assume

$$E[\eta_i] \;=\; 0 \qquad \text{and} \qquad \mathrm{Var}[\eta_i] = 1 \qquad \text{for any } i \in \mathbb{N}.$$

Here, the values 0 and 1 are just a convenient normalization, but it is an important assumption that the random variables are independent with finite variances. Given an initial value $x_0 \in \mathbb{R}$ and a fine discretization step size $h > 0$, we now define a stochastic process $(X_n^{(h)})$ in discrete time by $X_0^{(h)} = x_0$, and

$$X_{k+1}^{(h)} - X_k^{(h)} \quad = \quad b(X_k^{(h)}) \cdot h + \sigma(X_k^{(h)})\sqrt{h}\, \eta_{k+1}, \qquad \text{for } k = 0, 1, 2, \ldots \tag{2.2.5}$$

One should think of $X_k^{(h)}$ as an approximation for the value of the process (X_t) at time $t = k \cdot h$. The equation (2.2.5) can be rewritten as

$$X_n^{(h)} \;=\; x_0 + \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \cdot \sqrt{h} \cdot \eta_{k+1}. \tag{2.2.6}$$

To understand the scaling factors h and $\sqrt{h}$ we note first that if $\sigma \equiv 0$ then (2.2.5) respectively (2.2.6) is the Euler discretization of the ordinary differential equation (2.2.3). Furthermore, if $b \equiv 0$ and $\sigma \equiv 1$, then the *diffusive scaling* by a factor $\sqrt{h}$ in the second term ensures that the continuous time process $X_{\lfloor t/h \rfloor}^{(h)}$, $t \in [0, \infty)$, converges in distribution as $h \searrow 0$. Indeed, the functional central limit theorem (Donsker's invariance principle) states that the limit process in this case is a Brownian motion $(B_t)_{t\in[0,\infty)}$. In general, (2.2.6) is an Euler discretization of a stochastic differential equation of type

$$dX_t \quad = \quad b(X_t)\,dt + \sigma(X_t)\,dB_t$$

where $(B_t)_{t\geq 0}$ is a Brownian motion. Let $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$ denote the filtration generated by the random variables η_i. The following exercise summarizes basic properties of the process $X^{(h)}$ in the case of normally distributed increments.

Exercise. Suppose that the random variables η_i are standard normally distributed.

(1). Prove that the process $X^{(h)}$ is a time-homogeneous $(\mathcal{F}_n)$ Markov chain with transition kernel

$$p(x, \bullet) \quad = \quad N(x + b(x)h, \sigma(x)^2 h)[\,\bullet\,].$$

(2). Show that the Doob decomposition $X^{(h)} = M^{(h)} + A^{(h)}$ is given by

$$A_n^{(h)} \;=\; \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h, \quad M_n^{(h)} \;=\; x_0 + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \sqrt{h}\,\eta_{k+1}, \tag{2.2.7}$$

and the conditional variance process of the martingale part is

$$\langle M^{(h)} \rangle_n \quad = \quad \sum_{k=0}^{n-1} \sigma(X_k^{(h)})^2 \cdot h. \tag{2.2.8}$$

(3). Conclude that

$$E[(M_n^{(h)} - x_0)^2] \quad = \quad \sum_{k=0}^{n-1} E[\sigma(X_k^{(h)})^2] \cdot h. \tag{2.2.9}$$

The last equation can be used in combination with the maximal inequality for martingales to derive bounds for the processes $(X^{(h)})$ in an efficient way, cf. Section 2.4 below.

Remark (Quadratic variation). The quadratic variation of a square integrable martingale (M_n) is the process $[M]_n$ defined by

$$[M]_n \;=\; \sum_{k=1}^{n} (M_k - M_{k-1})^2, \qquad n \geq 0.$$

It is easy to verify that $M_n^2 - [M]_n$ is again a martingale. However, $[M]_n$ is not predictable. For continuous martingales in continuous time, the quadratic variation and the conditional variance process coincide. In discrete time or for discontinuous martingales they are usually different.

Martingale problem

For a Markov chain (X_n) we obtain a Doob decomposition

$$f(X_n) \;=\; M_n^{[f]} + A_n^{[f]} \tag{2.2.10}$$

for any function f on the state space such that $f(X_n)$ is integrable for each n. Computation of the predictable part leads to the following general result:

Theorem 2.5 (Martingale problem for time-homogeneuous Markov chains). *Let p be a stochastic kernel on a measurable space $(S, \mathcal{B})$. Then for an $(\mathcal{F}_n)$ adapted stochastic process $(X_n)_{n\geq 0}$ with state space $(S, \mathcal{B})$ the following statements are equivalent:*

(1). (X_n) is a time homogeneous $(\mathcal{F}_n)$ Markov chain with transition kernel p.

(2). (X_n) is a ***solution of the martingale problem for the operator*** *$\mathscr{L} = p - I$, i.e., there is a decomposition*

$$f(X_n) \;=\; M_n^{[f]} + \sum_{k=0}^{n-1} (\mathscr{L} f)(X_k), \qquad n \geq 0,$$

with an $(\mathcal{F}_n)$ martingale $(M_n^{[f]})$ for every function $f : S \to \mathbb{R}$ such that $f(X_n)$ is integrable for each n, or, equivalently, for every bounded function $f : S \to \mathbb{R}$.

In particular, we see once more that if $f(X_n)$ is integrable and f is harmonic ($\mathscr{L} f = 0$) then $f(X_n)$ is a martingale, and if f is superharmonic ($\mathscr{L} f \leq 0$), then $f(X_n)$ is a supermartingale. The theorem hence extends Theorem 2.3 above.

Proof. The implication "(i)⇒(ii)" is just the Doob decomposition for $f(X_n)$. In fact, by Theorem 2.4, the predictable part is given by

$$\begin{aligned} A_n^{[f]} &= \sum_{k=0}^{n-1} E[f(X_{k+1}) - f(X_k) \mid \mathcal{F}_k] \\ &= \sum_{k=0}^{n-1} (pf(X_k) - f(X_k)) \quad = \quad \sum_{k=0}^{n-1} (\mathscr{L}f)(X_k), \end{aligned}$$

and $M_n^{[f]} = f(X_n) - A_n^{[f]}$ is a martingale.
To prove the converse implication "(ii)⇒(i)" suppose that $M_n^{[f]}$ is a martingale for any bounded $f : S \to \mathbb{R}$. Then

$$\begin{aligned} 0 &= E[M_{n+1}^{[f]} - M_n^{[f]} \mid \mathcal{F}_n] \\ &= E[f(X_{n+1}) - f(X_n) \mid \mathcal{F}_n] - ((pf)(X_n) - f(X_n)) \\ &= E[f(X_{n+1}) \mid \mathcal{F}_n] - (pf)(X_n) \end{aligned}$$

almost surely for any bounded function f. Hence (X_n) is an $(\mathcal{F}_n)$ Markov chain with transition kernel p. □

Example (One dimensional Markov chains). Suppose that under P_x, the process (X_n) is a time homogeneous Markov chain with state space $S = \mathbb{R}$ or $S = \mathbb{Z}$, initial state $X_0 = x$, and transition kernel p. Assuming $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ for each n, we define the "*drift*" and the "*fluctuations*" of the process by

$$\begin{aligned} b(x) &:= E_x[X_1 - X_0] \\ a(x) &= \mathrm{Var}_x[X_1 - X_0]. \end{aligned}$$

We now compute the Doob decomposition of (X_n). Choosing $f(x) = x$ we have

$$(p - I)f(x) \quad = \quad \int y\, p(x, dy) - x \quad = \quad E_x[X_1 - X_0] \;=\; b(x).$$

Hence by Theorem 2.5,

$$X_n \quad = \quad M_n + \sum_{k=0}^{n-1} b(X_k) \tag{2.2.11}$$

with an $(\mathcal{F}_n)$ martingale (M_n). To obtain detailed information on M_n, we compute the variance process: By (2.2.11) and the Markov property, we obtain

$$\langle M\rangle_n \;=\; \sum_{k=0}^{n-1} \mathrm{Var}[M_{k+1}-M_k \mid \mathcal{F}_k] \;=\; \sum_{k=0}^{n-1} \mathrm{Var}[X_{k+1}-X_k \mid \mathcal{F}_k] \quad = \quad \sum_{k=0}^{n-1} a(X_k).$$

Therefore

$$M_n^2 \quad = \quad \widetilde{M}_n + \sum_{k=0}^{n-1} a(X_k) \tag{2.2.12}$$

with another $(\mathcal{F}_n)$ martingale $(\widetilde{M}_n)$. The functions $a(x)$ and $b(x)$ can now be used in connection with fundamental results for martingales as e.g. the maximal inequality (cf. 2.4 below) to derive bounds for Markov chains in an efficient way.

2.3 Gambling strategies and stopping times

Throughout this section, we fix a filtration $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$.

Martingale transforms

Suppose that $(M_n)_{n\geq 0}$ is a martingale w.r.t. $(\mathcal{F}_n)$, and $(C_n)_{n\in\mathbb{N}}$ is a predictable sequence of real-valued random variables. For example, we may think of C_n as the stake in the n-th round of a fair game, and of the martingale increment $M_n - M_{n-1}$ as the net gain (resp. loss) per unit stake. In this case, the capital I_n of a player with gambling strategy (C_n) after n rounds is given recursively by

$$I_n \quad = \quad I_{n-1} + C_n \cdot (M_n - M_{n-1}) \qquad \text{for any } n \in \mathbb{N},$$

i.e.,

$$I_n \quad = \quad I_0 + \sum_{k=1}^{n} C_k \cdot (M_k - M_{k-1}).$$

Definition (Martingale transform). *The stochastic process $C_\bullet M$ defined by*

$$(C_\bullet M)_n \;:=\; \sum_{k=1}^{n} C_k \cdot (M_k - M_{k-1}) \qquad \text{for any } n \geq 0,$$

*is called the **martingale transform** of the martingale $(M_n)_{n\geq 0}$ w.r.t. the predictable sequence $(C_n)_{n\geq 1}$, or the discrete stochastic integral of C w.r.t. M.*

We will see later that the process $C_\bullet M$ is a time-discrete version of the stochastic integral $\int C_s\, dM_s$ of a predictable continuous-time process C w.r.t. a continuous-time martingale M. To be precise, $(C_\bullet M)_n$ coincides with the Itô integral $\int_0^n C_{\lceil t\rceil}\, dM_{\lfloor t\rfloor}$ of the left continuous jump process $t \mapsto C_{\lceil t\rceil}$ w.r.t. the right continuous martingale $t \mapsto M_{\lfloor t\rfloor}$.

Example (Martingale strategy). One origin of the word “martingale” is the name of a well-known gambling strategy: In a standard coin-tossing game, the stake is doubled each time a loss occurs, and the player stops the game after the first time he wins. If the net gain in n rounds with unit stake is given by a standard Random Walk

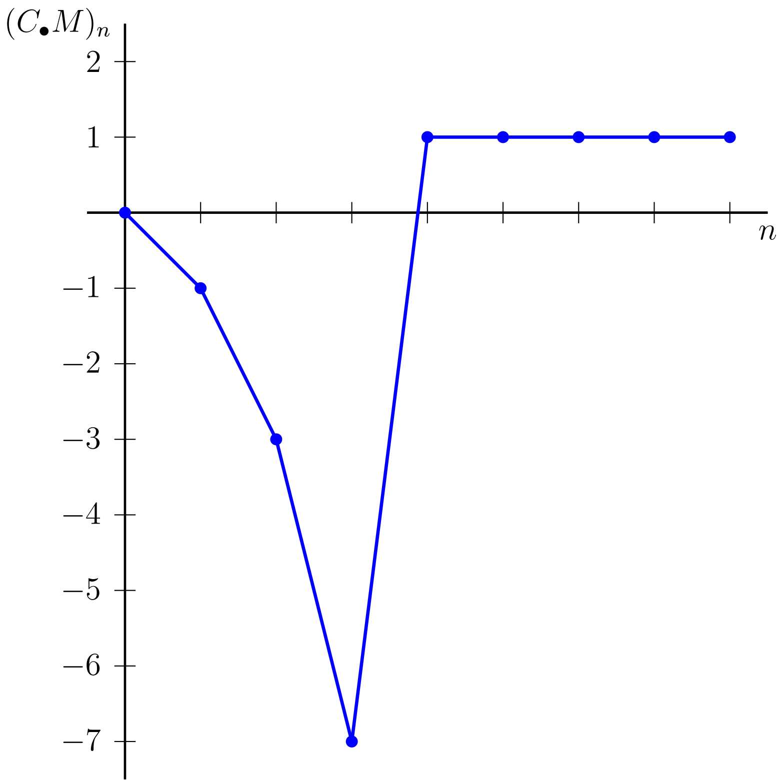
$$M_n \;=\; \eta_1 + \ldots + \eta_n, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = 1] = P[\eta_i = -1] = 1/2,$$

then the stake in the n-th round is

$$C_n \;=\; 2^{n-1} \quad \text{if } \eta_1 = \ldots = \eta_{n-1} = -1, \qquad \text{and} \qquad C_n \;=\; 0 \quad \text{otherwise.}$$

Clearly, with probability one, the game terminates in finite time, and at that time the player has always won one unit, i.e.,

$$P[(C_\bullet M)_n = 1 \quad \text{eventually}] \;=\; 1.$$



At first glance this looks like a safe winning strategy, but of course this would only be the case, if the player had unlimited capital and time available.

Theorem 2.6 (You can't beat the system!). *(1). If $(M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)$ martingale, and $(C_n)_{n\geq 1}$ is predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ for any $n \geq 1$, then $C_\bullet M$ is again an $(\mathcal{F}_n)$ martingale.*

(2). If (M_n) is an $(\mathcal{F}_n)$ supermartingale and $(C_n)_{n\geq 1}$ is non-negative and predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1$ for any n, then $C_\bullet M$ is again a supermartingale.

Proof. For $n \geq 1$ we have

$$\begin{aligned} E[(C_\bullet M)_n - (C_\bullet M)_{n-1} \mid \mathcal{F}_{n-1}] &= E[C_n \cdot (M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] \\ &= C_n \cdot E[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] \quad = \quad 0 \qquad P\text{-a.s.} \end{aligned}$$

This proves the first part of the claim. The proof of the second part is similar. □

The theorem shows that a fair game (a martingale) can not be transformed by choice of a clever gambling strategy into an unfair (or "superfair") game. In models of financial markets this fact is crucial to exclude the existence of arbitrage possibilities (riskless profit).

Example (Martingale strategy, cont.). For the classical martingale strategy, we obtain

$$E[(C_\bullet M)_n] \;=\; E[(C_\bullet M)_0] \;=\; 0 \qquad \text{for any } n \geq 0$$

by the martingale property, although

$$\lim_{n\to\infty} (C_\bullet M)_n \;=\; 1 \qquad P\text{-almost surely.}$$

This is a classical example showing that the assertion of the dominated convergence theorem may not hold if the assumptions are violated.

Remark. The integrability assumption in Theorem 2.6 is always satisfied if the random variables C_n are bounded, or if both C_n and M_n are square-integrable for any n.

Example (Financial market model with one risky asset). Suppose that during each time interval $(n-1, n)$, an investor is holding Φ_n units of an asset with price S_n per unit at time n. We assume that (S_n) is an adapted and (Φ_n) is a predictable stochastic process w.r.t. a filtration $(\mathcal{F}_n)$. If the investor always puts his remaining capital onto a bank account with guaranteed interest rate r ("riskless asset") then the change of his capital V_n during the time interval $(n-1, n)$ is given by

$$V_n \;=\; V_{n-1} + \Phi_n \cdot (S_n - S_{n-1}) + (V_{n-1} - \Phi_n \cdot S_{n-1}) \cdot r. \tag{2.3.1}$$

Considering the discounted quantity $\widetilde{V}_n = V_n/(1+r)^n$, we obtain the equivalent recursion

$$\widetilde{V}_n \;=\; \widetilde{V}_{n-1} + \Phi_n \cdot (\widetilde{S}_n - \widetilde{S}_{n-1}) \qquad \text{for any } n \geq 1. \tag{2.3.2}$$

In fact, (2.3.1) holds if and only if

$$V_n - (1+r)V_{n-1} \;=\; \Phi_n \cdot (S_n - (1+r)S_{n-1}),$$

which is equivalent to (2.3.2). Therefore, the discounted capital at time n is given by

$$\widetilde{V}_n \;=\; V_0 + (\Phi_\bullet \widetilde{S})_n.$$

By Theorem 2.6, we can conclude that, if the discounted price process $(\widetilde{S}_n)$ is an $(\mathcal{F}_n)$ martingale w.r.t. a given probability measure, then $(\widetilde{V}_n)$ is a martingale as well. In this case, assuming that V_0 is constant, we obtain in particular

$$E[\widetilde{V}_n] \;=\; V_0,$$

or, equivalently,

$$E[V_n] \;=\; (1+r)^n V_0 \qquad \text{for any } n \geq 0. \tag{2.3.3}$$

This fact, together with the existence of a martingale measure, can now be used for option pricing under a **no-arbitrage assumption**. To this end we assume that the payoff of an option at time N is given by an $(\mathcal{F}_N)$-measurable random variable F. For example, the payoff of a European call option with strike price K based on the asset with price process (S_n) is $S_N - K$ if the price S_n at maturity exceeds K, and 0 otherwise, i.e.,

$$F \;=\; (S_N - K)^+.$$

Suppose further that the option can be *replicated by a hedging strategy* (Φ_n), i.e., there exists an $\mathcal{F}_0$-measurable random variable V_0 and a predictable sequence of random variables $(\Phi_n)_{1\leq n\leq N}$ such that

$$F \;=\; V_N$$

is the value at time N of a portfolio with initial value V_0 w.r.t. the trading strategy (Φ_n). Then, assuming the non-existence of arbitrage possibilities, the option price at time 0 has to be V_0, since otherwise one could construct an arbitrage strategy by selling the option and investing money in the stock market with strategy (Φ_n), or conversely. Therefore, if a martingale measure exists (i.e., an underlying probability measure such that the discounted stock price $(\widetilde{S}_n)$ is a martingale), then the no-arbitrage price of the option at time 0 can be computed by (2.3.3) where the expectation is taken w.r.t. the martingale measure.

The following exercise shows how this works out in the Cox-Ross-Rubinstein binomial model:

Exercise (No-Arbitrage Pricing in the CRR model). Consider the CRR binomial model, i.e., $\Omega = \{1+a, 1+b\}^N$ with $-1 < a < r < b < \infty$, $X_i(\omega_1, \ldots, \omega_N) = \omega_i$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, and

$$S_n = S_0 \cdot \prod_{i=1}^{n} X_i, \quad n = 0, 1, \ldots, N,$$

where S_0 is a constant.

(1). *Completeness of the CRR model:* Prove that for any function $F : \Omega \to \mathbb{R}$ there exists a constant V_0 and a predictable sequence $(\Phi_n)_{1 \le n \le N}$ such that $F = V_N$ where $(V_n)_{1 \le n \le N}$ is defined by (2.3.1), or, equivalently,

$$\frac{F}{(1+r)^N} = \widetilde{V}_N = V_0 + (\Phi_\bullet \widetilde{S})_N.$$

Hence in the CRR model, any $\mathcal{F}_N$-measurable function F can be replicated by a predictable trading strategy. Market models with this property are called *complete*.

Hint: Prove inductively that for $n = N, N-1, \ldots, 0$, $\widetilde{F} = F/(1+r)^N$ can be represented as

$$\widetilde{F} = \widetilde{V}_n + \sum_{i=n+1}^{N} \Phi_i \cdot (\widetilde{S}_i - \widetilde{S}_{i-1})$$

with an $\mathcal{F}_n$-measurable function $\widetilde{V}_n$ and a predictable sequence $(\Phi_i)_{n+1 \le i \le N}$.

(2). *Option pricing:* Derive a general formula for the no-arbitrage price of an option with payoff function $F : \Omega \to \mathbb{R}$ in the CRR model. Compute the no-arbitrage price for a European call option with maturity N and strike K explicitly.

Stopped Martingales

One possible strategy for controlling a fair game is to terminate the game at a time depending on the previous development. Recall that a random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a **stopping time** w.r.t. the filtration $(\mathcal{F}_n)$ if and only if the event $\{T = n\}$ is contained in $\mathcal{F}_n$ for any $n \ge 0$, or equivalently, iff $\{T \le n\} \in \mathcal{F}_n$ for any $n \ge 0$.

Example (Hitting times). (1). The **first hitting time**

$$T_B \quad = \quad \min\{n \geq 0 : X_n \in B\} \qquad (\text{where } \min \emptyset := \infty)$$

and the **first passage or return time**

$$S_B \quad = \quad \min\{n \geq 1 : X_n \in B\}$$

to a measurable subset B of the state space by an $(\mathcal{F}_n)$ adapted stochastic process are $(\mathcal{F}_n)$ stopping times. For example, for $n \geq 0$,

$$\{T_B = n\} \quad = \quad \{X_1 \in B^C, \ldots, X_{n-1} \in B^C, X_n \in B\} \quad \in \quad \mathcal{F}_n.$$

If one decides to sell an asset as soon as the price S_n exceeds a given level $\lambda > 0$ then the selling time equals $T_{(\lambda,\infty)}$ and is hence a stopping time.

(2). On the other hand, the **last visit time**

$$L_B \quad := \quad \sup\{n \geq 0 : X_n \in B\} \qquad (\text{where } \sup \emptyset := 0)$$

is not a stopping time in general. Intuitively, to decide whether $L_B = n$, information on the future development of the process is required.

We consider an $(\mathcal{F}_n)$-adapted stochastic process $(M_n)_{n\geq 0}$, and an $(\mathcal{F}_n)$-stopping time T on the probability space $(\Omega, \mathcal{A}, P)$. The process stopped at time T is defined as $(M_{T\wedge n})_{n\geq 0}$ where

$$M_{T\wedge n}(\omega) \;=\; M_{T(\omega)\wedge n}(\omega) \;=\; \begin{cases} M_n(\omega) & \text{for } n \leq T(\omega), \\ M_{T(\omega)}(\omega) & \text{for } n \geq T(\omega). \end{cases}$$

For example, the process stopped at a hitting time T_B gets stuck at the first time it enters the set B.

Theorem 2.7 (Optional Stopping Theorem,Version 1). *If $(M_n)_{n\geq 0}$ is a martingale (resp. a supermartingale) w.r.t. $(\mathcal{F}_n)$, and T is an $(\mathcal{F}_n)$-stopping time, then the stopped process $(M_{T\wedge n})_{n\geq 0}$ is again an $(\mathcal{F}_n)$-martingale (resp. supermartingale). In particular, we have*

$$E[M_{T\wedge n}] \quad \overset{(\leq)}{=} \quad E[M_0] \qquad \text{for any } n \geq 0.$$

Proof. Consider the following strategy:

$$C_n \;=\; I_{\{T\geq n\}} \;=\; 1 - I_{\{T\leq n-1\}},$$

i.e., we put a unit stake in each round before time T and quit playing at time T. Since T is a stopping time, the sequence (C_n) is predictable. Moreover,

$$M_{T\wedge n} - M_0 \;=\; (C_\bullet M)_n \qquad \text{for any } n \geq 0. \tag{2.3.4}$$

In fact, for the increments of the stopped process we have

$$M_{T\wedge n} - M_{T\wedge(n-1)} \;=\; \left\{ \begin{array}{ll} M_n - M_{n-1} & \text{if } T \geq n \\ 0 & \text{if } T \leq n-1 \end{array} \right\} \;=\; C_n \cdot (M_n - M_{n-1}),$$

and (2.3.4) follows by summing over n. Since the sequence (C_n) is predictable, bounded and non-negative, the process $C_\bullet M$ is a martingale, supermartingale respectively, provided the same holds for M. □

Remark (IMPORTANT). (1). In general, it is NOT TRUE under the assumptions in Theorem 2.7 that

$$E[M_T] \;=\; E[M_0], \quad E[M_T] \;\leq\; E[M_0] \quad \text{respectively.} \tag{2.3.5}$$

Suppose for example that (M_n) is the classical Random Walk starting at 0 and $T = T_{\{1\}}$ is the first hitting time of the point 1. Then, by recurrence of the Random Walk, $T < \infty$ and $M_T = 1$ hold almost surely although $M_0 = 0$.

(2). If, on the other hand, T is a *bounded stopping time*, then there exists $n \in \mathbb{N}$ such that $T(\omega) \leq n$ for any ω. In this case, the optional stopping theorem implies

$$E[M_T] \;=\; E[M_{T\wedge n}] \;\overset{(\leq)}{=}\; E[M_0].$$

More general sufficient conditions for (2.3.5) are given in Theorems 2.8, 2.9 and 2.10 below.

Example (Classical ruin problem). Let $a, b, x \in \mathbb{Z}$ with $a < x < b$. We consider the classical Random Walk

$$X_n \;=\; x + \sum_{i=1}^{n} \eta_i, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = \pm 1] \;=\; \frac{1}{2},$$

with initial value $X_0 = x$. We now show how to apply the Optional Stopping Theorem to compute the distributions of the exit time

$$T(\omega) \;=\; \min\{n \geq 0 \,:\, X_n(\omega) \notin (a,b)\},$$

and the exit point X_T. These distributions can also be computed by more traditional methods (first step analysis, reflection principle), but martingales yield an elegant and general approach.

(1). *Ruin probability* $r(x) = P[X_T = a]$.
The process (X_n) is a martingale w.r.t. the filtration $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$, and $T < \infty$ almost surely holds by elementary arguments. As the stopped process $X_{T\wedge n}$ is bounded ($a \leq X_{T\wedge n} \leq b$), we obtain

$$x \;=\; E[X_0] \;=\; E[X_{T\wedge n}] \overset{n\to\infty}{\to} E[X_T] \;=\; a\cdot r(x) + b\cdot(1-r(x))$$

by the Optional Stopping Theorem and the Dominated Convergence Theorem. Hence

$$r(x) \quad = \quad \frac{b-x}{a-x}. \tag{2.3.6}$$

(2). *Mean exit time from* (a,b).
To compute the expectation $E[T]$, we apply the Optional Stopping Theorem to the $(\mathcal{F}_n)$ martingale

$$M_n \;:=\; X_n^2 - n.$$

By monotone and dominated convergence, we obtain

$$\begin{aligned} x^2 \quad &= \quad E[M_0] \;=\; E[M_{T\wedge n}] \;=\; E[X_{T\wedge n}^2] - E[T\wedge n] \\ &\overset{n\to\infty}{\longrightarrow} \quad E[X_T^2] - E[T]. \end{aligned}$$

Therefore, by (2.3.6),

$$\begin{aligned} E[T] \quad &= \quad E[X_T^2] - x^2 \;=\; a^2\cdot r(x) + b^2\cdot(1-r(x)) - x^2 \\ &= \quad (b-x)\cdot(x-a). \end{aligned} \tag{2.3.7}$$

(3). *Mean passage time of b.*

The first passage time $T_b = \min\{n \geq 0 \ : \ X_n = b\}$ is greater or equal than the exit time from the interval (a, b) for any $a < x$. Thus by (2.3.7), we have

$$E[T_b] \;\geq\; \lim_{a \to -\infty} (b - x) \cdot (x - a) \;=\; \infty,$$

i.e., T_b is **not integrable**! These and some other related passage times are important examples of random variables with a heavy-tailed distribution and infinite first moment.

(4). *Distribution of passage times.*

We now compute the distribution of the first passage time T_b explicitly in the case $x = 0$ and $b = 1$. Hence let $T = T_1$. As shown above, the process

$$M_n^\lambda \;:=\; e^{\lambda X_n}/(\cosh \lambda)^n, \qquad n \geq 0,$$

is a martingale for each $\lambda \in \mathbb{R}$. Now suppose $\lambda > 0$. By the Optional Stopping Theorem,

$$1 \;=\; E[M_0^\lambda] \;=\; E[M_{T \wedge n}^\lambda] \;=\; E[e^{\lambda X_{T \wedge n}}/(\cosh \lambda)^{T \wedge n}] \tag{2.3.8}$$

for any $n \in \mathbb{N}$. As $n \to \infty$, the integrands on the right hand side converge to $e^\lambda (\cosh \lambda)^{-T} \cdot I_{\{T < \infty\}}$. Moreover, they are uniformly bounded by e^λ, since $X_{T \wedge n} \leq 1$ for any n. Hence by the Dominated Convergence Theorem, the expectation on the right hand side of (2.3.8) converges to $E[e^\lambda/(\cosh \lambda)^T \ ; \ T < \infty]$, and we obtain the identity

$$E[(\cosh \lambda)^{-T} \ ; \ T < \infty] \;=\; e^{-\lambda} \qquad \text{for any } \lambda > 0. \tag{2.3.9}$$

Taking the limit as $\lambda \searrow 0$, we see that $P[T < \infty] = 1$. Taking this into account, and substituting $s = 1/\cosh \lambda$ in (2.3.9), we can now compute the generating function of T explicitly:

$$E[s^T] \;=\; e^{-\lambda} \;=\; (1 - \sqrt{1 - s^2})/s \qquad \text{for any } s \in (0, 1). \tag{2.3.10}$$

Developing both sides into a power series finally yields

$$\sum_{n=0}^{\infty} s^n \cdot P[T = n] \;=\; \sum_{m=1}^{\infty} (-1)^{m+1} \binom{1/2}{m} s^{2m-1}.$$

Therefore, the distribution of the first passage time of 1 is given by

$$P[T = 2m-1] = (-1)^{m+1}\binom{1/2}{m} = (-1)^{m+1}\cdot\frac{1}{2}\cdot\left(-\frac{1}{2}\right)\cdots\left(\frac{1}{2}-m+1\right)/m!$$

and $P[T = 2m] = 0$ for any $m\mathbb{N}$.

Optional Stopping Theorems

Stopping times occurring in applications are typically not bounded. Therefore, we need more general conditions guaranteeing that (2.3.5) holds nevertheless. A first general criterion is obtained by applying the Dominated Convergence Theorem:

Theorem 2.8 (Optional Stopping Theorem, Version 2). *Suppose that (M_n) is a martingale w.r.t. $(\mathcal{F}_n)$, T is an $(\mathcal{F}_n)$-stopping time with $P[T < \infty] = 1$, and there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ such that*

$$|M_{T\wedge n}| \leq Y \qquad P\text{-almost surely for any } n \in \mathbb{N}.$$

Then

$$E[M_T] = E[M_0].$$

Proof. Since $P[T < \infty] = 1$, we have

$$M_T = \lim_{n\to\infty} M_{T\wedge n} \qquad P\text{-almost surely.}$$

By Theorem 2.7, $E[M_0] = E[M_{T\wedge n}]$, and by the Dominated Convergence Theorem, $E[M_{T\wedge n}] \longrightarrow E[M_T]$ as $n \to \infty$. □

Remark (Weakening the assumptions). Instead of the existence of an integrable random variable Y dominating the random variables $M_{T\wedge n}$, $n \in \mathbb{N}$, it is enough to assume that these random variables are **uniformly integrable**, i.e.,

$$\sup_{n\in\mathbb{N}} E\big[|M_{T\wedge n}| \,;\, |M_{T\wedge n}| \geq c\big] \quad \to \quad 0 \qquad \text{as } c \to \infty.$$

A corresponding generalization of the Dominated Convergence Theorem is proven in Section 4.3 below.

For non-negative supermartingales, we can apply Fatou's Lemma instead of the Dominated Convergence Theorem to pass to the limit as $n \to \infty$ in the Stopping Theorem. The advantage is that no integrability assumption is required. Of course, the price to pay is that we only obtain an inequality:

Theorem 2.9 (Optional Stopping Theorem, Version 3). *If (M_n) is a non-negative supermartingale w.r.t. $(\mathcal{F}_n)$, then*

$$E[M_0] \;\geq\; E[M_T\,;\,T<\infty]$$

holds for any $(\mathcal{F}_n)$ stopping time T.

Proof. Since $M_T = \lim_{n\to\infty} M_{T\wedge n}$ on $\{T<\infty\}$, and $M_T \geq 0$, Theorem 2.7 combined with Fatou's Lemma implies

$$E[M_0] \;\geq\; \liminf_{n\to\infty} E[M_{T\wedge n}] \;\geq\; E\left[\liminf_{n\to\infty} M_{T\wedge n}\right] \;\geq\; E[M_T\,;\,T<\infty].$$

□

Example (Dirichlet problem for Markov chains). Suppose that w.r.t. the probability measure P_x, the process (X_n) is a time-homogeneous Markov chain with measurable state space $(S,\mathcal{B})$, transition kernel p, and start in x. Let $D \in \mathcal{B}$ be a measurable subset of the state space, and $f : D^C \to \mathbb{R}$ a measurable function (the given "boundary values"), and let

$$T \;=\; \min\{n\geq 0 \,:\, X_n \in D^C\}$$

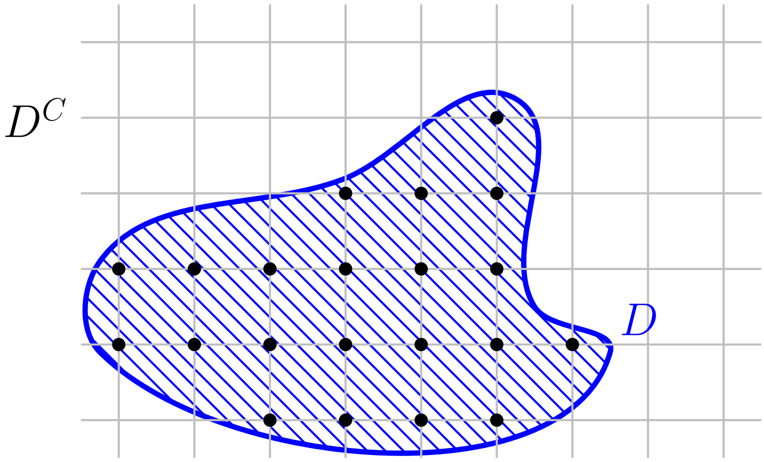
denote the first exit time of the Markov chain from D. By conditioning on the first step of the Markov chain, one can show that if f is non-negative or bounded, then the function

$$h(x) \;=\; E_x[f(X_T)\,;\,T<\infty], \qquad (x\in S),$$

is a solution of the *Dirichlet problem*

$$\begin{aligned} (ph)(x) &= h(x) \qquad \text{for } x\in D, \\ h(x) &= f(x) \qquad \text{for } x \in D^C, \end{aligned}$$

see [XXXStochastic Processes].



By considering the martingale $h(X_{T\wedge n})$ for a function h that is harmonic on D, we obtain a converse statement:

Exercise (Uniqueness of the Dirichlet problem). Suppose that $P_x[T < \infty] = 1$ for any $x \in S$.

(1). Prove that $h(X_{T\wedge n})$ is a martingale w.r.t. P_x for any bounded solution h of the Dirichlet problem and any $x \in S$.

(2). Conclude that if f is bounded, then

$$h(x) \;=\; E_x[f(X_T)] \tag{2.3.11}$$

is the unique bounded solution of the Dirichlet problem.

(3). Similarly, show that for any non-negative f, the function h defined by (2.3.11) is the minimal non-negative solution of the Dirichlet problem.

We finally state a version of the Optional Stopping Theorem that applies in particular to martingales with bounded increments:

Corollary 2.10 (Optional Stopping for martingales with bounded increments). *Suppose that (M_n) is an $(\mathcal{F}_n)$ martingale, and there exists a finite constant $K \in (0, \infty)$ such that*

$$E[|M_{n+1} - M_n| \,|\, \mathcal{F}_n] \;\leq\; K \qquad P\text{-almost surely for any } n \geq 0. \tag{2.3.12}$$

Then for any $(\mathcal{F}_n)$ stopping time T with $E[T] < \infty$, we have

$$E[M_T] \;=\; E[M_0].$$

Proof. For any $n \geq 0$,

$$|M_{T \wedge n}| \;\leq\; |M_0| + \sum_{i=0}^{\infty} |M_{i+1} - M_i| \cdot I_{\{T > i\}}.$$

Let Y denote the expression on the right hand side. We will show that Y is an integrable random variable – this implies the assertion by Theorem 2.8. To verify integrability of Y note that the event $\{T > i\}$ is contained in $\mathcal{F}_i$ for any $i \geq 0$ since T is a stopping time. Therefore and by (2.3.12),

$$E[|M_{i+1} - M_i|\,;\, T > i] \;=\; E[E[|M_{i+1} - M_i| \,|\, \mathcal{F}_i]\,;\, T > i] \;\leq\; k \cdot P[T > i].$$

Summing over i, we obtain

$$E[Y] \;\leq\; E[|M_0|] + k \cdot \sum_{i=0}^{\infty} P[T > i] \;=\; E[|M_0|] + k \cdot E[T] \;<\; \infty$$

by the assumptions. $\square$

Exercise (Integrability of stopping times). Prove that the expectation $E[T]$ of a stopping time T is finite if there exist constants $\varepsilon > 0$ and $k \in \mathbb{N}$ such that

$$P[T \leq n + k \,|\, \mathcal{F}_n] \;\geq\; \varepsilon \qquad P\text{-a.s. for any } n \in \mathbb{N}.$$

Wald's identity for random sums

We finally apply the Optional Stopping Theorem to sums of independent random variables with a random number T of summands. The point is that we do not assume that T is independent of the summands but only that it is a stopping time w.r.t. the filtration generated by the summands.

Let $S_n = \eta_1 + \ldots + \eta_n$ with i.i.d. random variables $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Denoting by m the expectations of the increments η_i, the process

$$M_n \;=\; S_n - n \cdot m$$

is a martingale w.r.t. $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$. By applying Corollary 2.10 to this martingale, we obtain:

Theorem 2.11 (Wald's identity). *Suppose that T is an $(\mathcal{F}_n)$ stopping time with $E[T] < \infty$. Then*

$$E[S_T] \;=\; m \cdot E[T].$$

Proof. For any $n \geq 0$, we have

$$E[|M_{n+1} - M_n| \mid \mathcal{F}_n] \;=\; E[|\eta_{n+1} - m| \,|\mathcal{F}_n] \;=\; E[|\eta_{n+1} - m|]$$

by the independence of the η_i. As the η_i are identically distributed and integrable, the right hand side is a finite constant. Hence Corollary 2.10 applies, and we obtain

$$0 \;=\; E[M_0] \;=\; E[M_T] \;=\; E[S_T] - m \cdot E[T].$$

□

2.4 Maximal inequalities

For a standard Random Walk $S_n = \eta_1 + \ldots + \eta_n$, η_i i.i.d. with $P[\eta_i = \pm 1] = 1/2$, the reflection principle implies the identity

$$\begin{aligned} P[\max(S_0, S_1, \ldots, S_n) \geq c] &= P[S_n \geq c] + P[S_n < c; \max(S_0, S_1, \ldots, S_n) \geq c] \\ &= P[|S_n| > c] + P[S_n > c] \end{aligned}$$

for any $c \in \mathbb{N}$. In combination with the Markov-Čebyšev inequality this can be used to control the running maximum of the Random Walk in terms of the moments of the last value S_n.

Maximal inequalities are corresponding estimates for $\max(M_0, M_1, \ldots, M_n)$ or $\sup\limits_{k\geq 0} M_k$ when (M_n) is a sub- or supermartingale respectively. These estimates are an important tool in stochastic analysis. They are a consequence of the Optional Stopping Theorem.

Doob's inequality

We first prove the basic version of maximal inequalities for sub- and supermartingales:

Theorem 2.12 (Doob).

(1). Suppose that $(M_n)_{n\geq 0}$ is a non-negative supermartingale. Then

$$P\left[\sup_{k\geq 0} M_k \geq c\right] \;\leq\; \frac{1}{c}\cdot E[M_0] \qquad \textit{for any } c > 0.$$

(2). Suppose that $(M_n)_{n\geq 0}$ is a non-negative submartingale. Then

$$P\left[\max_{0\leq k\leq n} M_k \geq c\right] \leq \frac{1}{c}\cdot E\left[M_n\,;\; \max_{0\leq k\leq n} M_k \geq c\right] \leq \frac{1}{c}\cdot E[M_n] \quad \textit{for any } c > 0.$$

Proof. (1). For $c > 0$ we consider the stopping time

$$T_c \;=\; \min\{k \geq 0 \;:\; M_k \geq c\}, \qquad \min\emptyset = \infty.$$

Note that $T_c < \infty$ whenever $\sup M_k > c$. Hence by the version of the Optional Stopping Theorem for non-negative supermartingales, we obtain

$$P[\sup M_k > c] \;\leq\; P[T_c < \infty] \;\leq\; \frac{1}{c}E[M_{T_c}\,;\, T_c < \infty] \;\leq\; \frac{1}{c}E[M_0].$$

Here we have used in the second and third step that (M_n) is non-negative. Replacing c by $c-\varepsilon$ and letting ε tend to zero we can conclude

$$P[\sup M_k \geq c] \;=\; \lim_{\varepsilon\searrow 0} P[\sup M_k > c-\varepsilon] \;\leq\; \liminf_{\varepsilon\searrow 0} \frac{1}{c-\varepsilon}E[M_0] \;=\; \frac{1}{c}\cdot E[M_0].$$

(2). For a non-negative supermartingale, we obtain

$$\begin{aligned} P\left[\max_{0\le k\le n} M_k \ge c\right] &= P[T_c \le n] \;\le\; \frac{1}{c} E[M_{T_c}\,;\, T_c \le n] \\ &= \frac{1}{c}\sum_{k=0}^{n} E[M_k\,;\, T_c = k] \;\le\; \frac{1}{c}\sum_{k=0}^{n} E[M_n\,;\, T_c = k] \\ &= \frac{1}{c}\cdot E[M_n\,;\, T_c \le n]. \end{aligned}$$

Here we have used in the second last step that $E[M_k\,;\, T_c = k] \le E[M_n\,;\, T_c = k]$ since (M_n) is a supermartingale and $\{T_c = k\}$ is in $\mathcal{F}_k$.

□

First consequences of Doob's maximal inequality for submartingales are extensions of the classical Markov- Čebyšev inequalities:

Corollary 2.13. *(1). Suppose that $(M_n)_{n\ge 0}$ is an arbitrary submartingale (not necessarily non-negative!). Then*

$$\begin{aligned} P\left[\max_{k\le n} M_k \ge c\right] &\le \frac{1}{c} E\left[M_n^+\,;\, \max_{k\le n} M_k \ge c\right] \qquad \text{for any } c > 0, \text{ and} \\ P\left[\max_{k\le n} M_k \ge c\right] &\le e^{-\lambda c} E\left[e^{\lambda M_n}\,;\, \max_{k\le n} M_k \ge c\right] \qquad \text{for any } \lambda, c > 0. \end{aligned}$$

(2). If (M_n) is a martingale then, moreover, the estimates

$$P\left[\max_{k\le n} |M_k| \ge c\right] \le \frac{1}{c^p} E\left[|M_n|^p\,;\, \max_{k\le n} |M_k| \ge c\right]$$

hold for any $c > 0$ and $p \in [1,\infty)$.

Proof. The corollary follows by applying the maximal inequality to the non-negative submartingales $M_n^+, \exp(\lambda M_n), |M_n|^p$ respectively. These processes are indeed submartingales, as the functions $x \mapsto x^+$ and $x \mapsto \exp(\lambda x)$ are convex and non-decreasing for any $\lambda > 0$, and the functions $x \mapsto |x|^p$ are convex for any $p \ge 1$. □

L^p inequalities

The last estimate in Corollary 2.13 can be used to bound the L^p norm of the running maximum of a martingale in terms of the L^p-norm of the last value. The resulting bound, known as Doob's L^p-inequality, is crucial for stochastic analysis. We first remark:

Lemma 2.14. *If $Y : \Omega \to \mathbb{R}_+$ is a non-negative random variable, and $G(y) = \int_0^y g(x)dx$ is the integral of a non-negative function $g : \mathbb{R}_+ \to \mathbb{R}_+$, then*

$$E[G(Y)] \;=\; \int_0^\infty g(c) \cdot P[Y \geq c]\, dc.$$

Proof. By Fubini's theorem we have

$$\begin{aligned} E[G(Y)] \;&=\; E\left[\int_0^Y g(c)\, dc\right] \;=\; E\left[\int_0^\infty I_{[0,Y]}(c) g(c)\, dc\right] \\ &=\; \int_0^\infty g(c) \cdot P[Y \geq c]\, dc. \end{aligned}$$

□

Theorem 2.15 (Doob's L^p inequality). *Suppose that $(M_n)_{n\geq 0}$ is a martingale, and let*

$$M_n^* \;:=\; \max_{k\leq n} |M_k|, \qquad \text{and} \qquad M^* \;:=\; \sup_k |M_k|.$$

Then, for any $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\|M_n^*\|_{L^p} \;\leq\; q \cdot \|M_n\|_{L^p}, \qquad \text{and} \qquad \|M^*\|_{L^p} \;\leq\; q \cdot \sup_n \|M_n\|_{L^p}.$$

In particular, if (M_n) is bounded in L^p then M^* is contained in L^p.

Proof. By Lemma 2.14, Corollary 2.13 applied to the martingales M_n and $(-M_n)$, and Fubini's theorem, we have

$$\begin{aligned} E[(M_n^*)^p] &\overset{2.14}{=} \int_0^\infty pc^{p-1} \cdot P[M_n^* \geq c]\, dc \\ &\overset{2.13}{\leq} \int_0^\infty pc^{p-2} E[|M_n|\,;\, M_n^* \geq c]\, dc \\ &\overset{\text{Fub.}}{=} E\left[|M_n| \cdot \int_0^{M_n^*} pc^{p-2}\, dp \right] \\ &= \frac{p}{p-1} E[|M_n| \cdot (M_n^*)^{p-1}] \end{aligned}$$

for any $n \geq 0$ and $p \in (1,\infty)$. Setting $q = \frac{p}{p-1}$ and applying Hölder's inequality to the right hand side, we obtain

$$E[(M_n^*)^p] \;\leq\; q \cdot \|M_n\|_{L^p} \cdot \|(M_n^*)^{p-1}\|_{L^q} \;=\; q \cdot \|M_n\|_{L^p} \cdot E[(M_n^*)^p]^{1/q},$$

i.e.,

$$\|M_n^*\|_{L^p} \;=\; E[(M_n^*)^p]^{1-1/q} \;\leq\; q \cdot \|M_n\|_{L^p}. \tag{2.4.1}$$

This proves the first inequality. The second inequality follows as $n \to \infty$, since

$$\|M^*\|_{L^p} \;=\; \left\| \lim_{n\to\infty} M_n^* \right\|_{L^p} \;=\; \liminf_{n\to\infty} \|M_n^*\|_{L^p} \;\leq\; q \cdot \sup_{n\in\mathbb{N}} \|M_n\|_{L^p}$$

by Fatou's Lemma. □

Hoeffding's inequality

For a standard Random Walk (S_n) starting at 0, the reflection principle combined with Bernstein's inequality implies the upper bound

$$P[\max(S_0,\ldots,S_n) \geq c] \;\leq\; 2 \cdot P[S_n \geq c] \;\leq\; 2 \cdot \exp(-2c^2/n)$$

for any $n \in \mathbb{N}$ and $c \in (0,\infty)$. A similar inequality holds for arbitrary martingales with bounded increments:

Theorem 2.16 (Azuma, Hoeffding). *Suppose that (M_n) is a martingale such that*

$$|M_n - M_{n-1}| \;\le\; a_n \qquad P\text{-almost surely}$$

for a sequence (a_n) of non-negative constants. Then

$$P\left[\max_{k\le n}(M_k - M_0) \ge c\right] \;\le\; \exp\left(-\frac{1}{2}c^2 \Big/ \sum_{i=1}^n a_i^2\right) \tag{2.4.2}$$

for any $n \in \mathbb{N}$ and $c \in (0,\infty)$.

Proof. W.l.o.g. we may assume $M_0 = 0$. Let $Y_n = M_n - M_{n-1}$ denote the martingale increments. We will apply the exponential form of the maximal inequality. For $\lambda > 0$ and $n \in \mathbb{N}$, we have,

$$E[e^{\lambda M_n}] \;=\; E\left[\prod_{i=1}^n e^{\lambda Y_i}\right] \;=\; E\left[e^{\lambda M_{n-1}} \cdot E\left[e^{\lambda Y_n} \mid \mathcal{F}_{n-1}\right]\right]. \tag{2.4.3}$$

To bound the conditional expectation, note that

$$e^{\lambda Y_n} \;\le\; \frac{1}{2}\frac{a_n - Y_n}{a_n}e^{-\lambda a_n} + \frac{1}{2}\frac{a_n + Y_n}{a_n}e^{\lambda a_n}$$

holds almost surely, since $x \mapsto \exp(\lambda x)$ is a convex function, and $-a_n \le Y_n \le a_n$. Indeed, the right hand side is the value at Y_n of the secant connecting the points $(-a_n, \exp(-\lambda a_n))$ and $(a_n, \exp(\lambda a_n))$. Since (M_n) is a martingale, we have

$$E[Y_n|\mathcal{F}_{n-1}] \;=\; 0,$$

and therefore

$$E[e^{\lambda Y_n} \mid \mathcal{F}_{n-1}] \;\le\; \left(e^{-\lambda a_n} + e^{\lambda a_n}\right)/2 \;=\; \cosh(\lambda a_n) \;\le\; e^{(\lambda a_n)^2/2}$$

almost surely. Now, by (2.4.3), we obtain

$$E[e^{\lambda Y_n}] \;\le\; E[e^{\lambda M_{n-1}}] \cdot e^{(\lambda a_n)^2/2}.$$

Hence, by induction on n,

$$E[e^{\lambda M_n}] \;\le\; \exp\left(\frac{1}{2}\lambda^2\sum_{i=1}^n a_i^2\right) \qquad \text{for any } n\in\mathbb{N}, \tag{2.4.4}$$

and, by the exponential maximal inequality (cf. Corollary 2.13),

$$P[\max_{k\le n} M_k \ge c] \;\le\; \exp\left(-\lambda c + \frac{1}{2}\lambda^2\sum_{i=1}^n a_i^2\right) \tag{2.4.5}$$

holds for any $n\in\mathbb{N}$ and $c,\lambda>0$.

For a given c and n, the expression on the right hand side of (2.4.5) is minimal for $\lambda = c/\sum_{i=1}^n a_i^2$. Choosing λ correspondingly, we finally obtain the upper bound (2.4.2). □

Hoeffding's concentration inequality has numerous applications, for example in the analysis of algorithms, cf. [Mitzenmacher, Upful: Probability and Computing]. Here, we just consider one simple example to illustrate the way it typically is applied:

Example (Pattern Matching). Suppose that $X_1, X_2, \dots, X_n$ is a sequence of independent, uniformly distributed random variables ("letters") taking value sin a finite set S (the underlying "alphabet"), and let

$$N \;=\; \sum_{i=0}^{n-l} I_{\{X_{i+1}=a_1, X_{i+2}=a_x, \dots, X_{i+l}=a_l\}} \tag{2.4.6}$$

denote the number of occurrences of a given "word" $a_1a_2\cdots a_l$ with l letters in the random text. In applications, the "word" could for example be a DNA sequence. We easily obtain

$$E[N] \;=\; \sum_{i=0}^{n-l} P[X_{i+k}=a_k \;\text{ for } k=1,\dots,l] \;=\; (n-l+1)/|S|^l. \tag{2.4.7}$$

To estimate the fluctuations of the random variable N around its mean value, we consider the martingale

$$M_i \;=\; E[N \mid \sigma(X_1,\dots,X_i)], \qquad (i=0,1,\dots,n)$$

with initial value $M_0 = E[N]$ and terminal value $M_n = N$. Since at most l of the summands in (2.4.6) are not independent of i, and each summand takes values 0 and 1 only, we have

$$|M_i - M_{i-1}| \;\leq\; l \qquad \text{for each } i = 0, 1, \ldots, n.$$

Therefore, by Hoeffding's inequality, applied in both directions, we obtain

$$P[|N - E[N]| \geq c] \;=\; P[|M_n - M_0| \geq c] \;\leq\; 2\exp(-c^2/(2nl^2))$$

for any $c > 0$, or equivalently,

$$P[|N - E[N]| \geq \varepsilon \cdot l\sqrt{n}] \;\leq\; 2 \cdot \exp(-\varepsilon^2/2) \qquad \text{for any } \varepsilon > 0. \tag{2.4.8}$$

The equation (2.4.7) and the bound (2.4.8) show that N is highly concentrated around its mean if l is small compared to $\sqrt{n}$.

Chapter 3

Martingales in continuous time

The notion of a martingale, sub- and supermartingale in continuous time can be defined similarly as in the discrete parameter case. Fundamental results such as the optional stopping theorem or the maximal inequality carry over from discrete parameter to continuous time martingales under additional regularity conditions as, for example, continuity of the sample paths. Similarly as for Markov chains in discrete time, martingale methods can be applied to derive explicit expressions and bounds for probabilities and expectations of Brownian motion in a clear and efficient way.

We start with the definition of martingales in continuous time. Let $(\Omega, \mathcal{A}, P)$ denote a probability space.

Definition. *(1). A continuous-time* ***filtration*** *on* $(\Omega, \mathcal{A})$ *is a family* $(\mathcal{F}_t)_{t\in[0,\infty)}$ *of* σ*-algebras* $\mathcal{F}_t \subseteq \mathcal{A}$ *such that* $\mathcal{F}_s \subseteq \mathcal{F}_t$ *for any* $0 \le s \le t$.

(2). A real-valued stochastic process $(M_t)_{t\in[0,\infty)}$ *on* $(\Omega, \mathcal{A}, P)$ *is called a* ***martingale*** *(or* ***super****-,* ***submartingale****) w.r.t. a filtration* $(\mathcal{F}_t)$ *if and only if*

(a) (M_t) *is adapted w.r.t.* $(\mathcal{F}_t)$*, i.e.,* M_t *is* $\mathcal{F}_t$ *measurable for any* $t \ge 0$.

(b) For any $t \ge 0$*, the random variable* M_t *(resp.* M_t^+, M_t^-*) is integrable.*

(c) $E[M_t \mid \mathcal{F}_s] \overset{(\le,\ge)}{=} M_s$ P*-almost surely for any* $0 \le s \le t$.

3.1 Some fundamental martingales of Brownian Motion

In this section, we identify some important martingales that are functions of Brownian motion. Let $(B_t)_{t\geq 0}$ denote a d-dimensional Brownian motion defined on $(\Omega, \mathcal{A}, P)$.

Filtrations generated by Brownian motion

Any stochastic process $(X_t)_{t\geq 0}$ in continuous time generates a filtration

$$\mathcal{F}_t^X \;=\; \sigma(X_s : 0 \leq s \leq t), \qquad t \geq 0.$$

However, not every hitting time that we are interested in is a stopping time w.r.t. this filtration. For example, for one-dimensional Brownian motion (B_t), the first hitting time $T \;=\; \inf\{t \geq 0 : B_t > c\}$ of the *open* interval (c, ∞) is not an $(\mathcal{F}_t^B)$ stopping time. An intuitive explanation for this fact is that for $t \geq 0$, the event $\{T \leq t\}$ is not contained in $\mathcal{F}_t^B$, since for a path with $B_s \leq c$ on $[0,t]$ and $B_t = c$, we can not decide at time t, if the path will enter the interval (c,∞) in the next instant. For this and other reasons, we also consider the right-continuous filtration

$$\mathcal{F}_t \;:=\; \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}^B, \qquad t \geq 0,$$

that takes into account "infinitesimal information on the future development."

Exercise (Hitting times as stopping times). Prove that the first hitting time $T_A = \inf\{t \geq 0 \;:\; B_t \in A\}$ of a set $A \subseteq \mathbb{R}^d$ is an $(\mathcal{F}_t^B)$ stopping time if A is closed, whereas T_A is an $(\mathcal{F}_t)$ stopping time, but not necessarily an $(\mathcal{F}_t^B)$ stopping time if A is open.

It is easy to verify that a d-dimensional Brownian motion (B_t) is also a Brownian motion w.r.t. the right-continuous filtration $(\mathcal{F}_t)$:

Lemma 3.1. *For any $0 \leq s < t$, the increment $B_t - B_s$ is independent of $\mathcal{F}_s$ with distribution $N(0, (t-s)\cdot I_d)$.*

Proof. Since $t \mapsto B_t$ is almost surely continuous, we have

$$B_t - B_s \;=\; \lim_{\substack{\varepsilon \searrow 0 \\ \varepsilon \in \mathbb{Q}}} (B_t - B_{s+\varepsilon}) \qquad P\text{-a.s.} \tag{3.1.1}$$

For small $\varepsilon > 0$ the increment $B_t - B_{s+\varepsilon}$ is independent of $\mathcal{F}^B_{s+\varepsilon}$, and hence independent of $\mathcal{F}_s$. Therefore, by (3.1.1), $B_t - B_s$ is independent of $\mathcal{F}_s$ as well. □

Another filtration of interest is the completed filtration $(\mathcal{F}^P_t)$. A σ-algebra $\mathcal{F}$ is called **complete** w.r.t. a probability measure P iff it contains all subsets of P-measure zero sets. The **completion** of a σ-algebra $\mathcal{A}$ w.r.t. a probability measure P on $(\Omega, \mathcal{A})$ is the complete σ-algebra

$$\mathcal{A}^P \;=\; \{A \subseteq \Omega : \exists A_1, A_2 \in \mathcal{A} \;:\; A_1 \subseteq A \subseteq A_2, P[A_2 \setminus A_1] = 0\}$$

generated by all sets in $\mathcal{A}$ and all subsets of P-measure zero sets in $\mathcal{A}$.

It can be shown that the completion $(\mathcal{F}^P_t)$ of the right-continuous filtration $(\mathcal{F}_t)$ is again right-continuous. The assertion of Lemma 3.1 obviously carries over to the completed filtration.

Remark (The "usual conditions"). Some textbooks on stochastic analysis consider only complete right-continuous filtrations. A filtration with these properties is said to **satisfy the usual conditions**. A disadvantage of completing the filtration, however, is that $(\mathcal{F}^P_t)$ depends on the underlying probability measure P (or, more precisely, on its null sets). This can cause problems when considering several non-equivalent probability measures at the same time.

Brownian Martingales

We now identify some basic martingales of Brownian motion:

Theorem 3.2 (Elementary martingales of Brownian motion). *For a d-dimensional Brownian motion (B_t) the following processes are martingales w.r.t. each of the filtrations $(\mathcal{F}^B_t), (\mathcal{F}_t)$ and $(\mathcal{F}^P_t)$:*

(1). The coordinate processes $B^{(i)}_t$, $1 \le i \le d$.

(2). $B^{(i)}_t B^{(j)}_t - t \cdot \delta_{ij}$ for any $1 \le i, j \le d$.

(3). $\exp(\alpha \cdot B_t - \frac{1}{2}|\alpha|^2 t)$ *for any* $\alpha \in \mathbb{R}^d$.

The processes $M_t^\alpha = \exp(\alpha \cdot B_t - \frac{1}{2}|\alpha|^2 t)$ *are called* ***exponential martingales***.

Proof. We only prove the second assertion for $d = 1$ and the right-continuous filtration $(\mathcal{F}_t)$. The verification of the remaining statements is left as an exercise.
For $d = 1$, since B_t is normally distributed, the $\mathcal{F}_t$-measurable random variable $B_t^2 - t$ is integrable for any t. Moreover, by Lemma 3.1,

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

almost surely. Hence

$$E[B_t^2 - t \mid \mathcal{F}_s] \;=\; B_s^2 - s \qquad P\text{-a.s. for any } 0 \le s \le t,$$

i.e., $B_t^2 - t$ is an $(\mathcal{F}_t)$ martingale. □

Remark (Doob decomposition, variance process of Brownian motion). For a one-dimensional Brownian motion (B_t), the theorem yields the Doob decomposition

$$B_t^2 \;=\; M_t + t$$

of the submartingale (B_t^2) into a martingale (M_t) and the continuous increasing adapted process $\langle B \rangle_t = t$.

A Doob decomposition of the process $f(B_t)$ for general functions $f \in C^2(\mathbb{R})$ will be obtained below as a consequence of Itô's celebrated formula. It states that

$$f(B_t) - f(B_0) \;=\; \int_0^t f'(B_s)\, dB_s + \frac{1}{2}\int_0^t f''(B_s)\, ds \tag{3.1.2}$$

where the first integral is an Itô stochastic integral, cf. Section 6.3. If, for example, f' is bounded, then the Itô integral is a martingale as a function of t. If f is convex then $f(B_t)$ is a submartingale and the second integral is a continuous increasing adapted process in

t. It is a consequence of (3.1.2) that Brownian motion solves the martingale problem for the operator $\mathscr{L}f = f''/2$ with domain $\mathrm{Dom}(\mathscr{L}) = \{f \in C^2(\mathbb{R}) : f' \text{ bounded}\}$.

Itô's formula (3.1.2) can also be extended to the multi-dimensional case, see Section 6.4 below. The second derivative is then replaced by the Laplacian $\Delta f = \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2}$. The multi-dimensional Itô formula implies that a sub- or superharmonic function of d-dimensional Brownian motion is a sub- or supermartingale respectively, if appropriate integrability conditions hold. We now give a direct proof of this fact by the mean value property:

Lemma 3.3 (Mean value property for harmonic function in $\mathbb{R}^d$). *Suppose that $h \in C^2(\mathbb{R}^d)$ is a* ***(super-)harmonic function****, i.e.,*

$$\Delta h(x) \stackrel{(\le)}{=} 0 \qquad \text{for any } x \in \mathbb{R}^d.$$

Then for any $x \in \mathbb{R}^d$ and any rotationally invariant probability measure μ on $\mathbb{R}^d$,

$$\int h(x+y)\,\mu(dy) \stackrel{(\le)}{=} h(x). \tag{3.1.3}$$

Proof. By the classical mean value property, $h(x)$ is equal to (resp. greater or equal than) the average value $\fint_{\partial B_r(x)} h$ of h on any sphere $\partial B_r(x)$ with center at x and radius $r > 0$, cf. e.g. [XXXKönigsberger: Analysis II]. Moreover, if μ is a rotationally invariant probability measure then the integral in (3.1.3) is an average of average values over spheres:

$$\int h(x+y)\,\mu(dy) = \int \fint_{\partial B_r(x)} h\,\mu_R(dr) \stackrel{(\le)}{=} h(x),$$

where μ_R is the distribution of $R(x) = |x|$ under μ. □

Theorem 3.4 (Superharmonic functions of Brownian motion are supermartingales). *If $h \in C^2(\mathbb{R}^d)$ is a (super-) harmonic function then $(h(B_t))$ is a (super-) martingale w.r.t. $(\mathcal{F}_t)$ provided $h(B_t)$ (resp. $h(B_t)^+$) is integrable for any $t \ge 0$.*

Proof. By Lemma 3.1 and the mean value property, we obtain

$$\begin{aligned} E[h(B_t) \mid \mathcal{F}_s](\omega) &= E[h(B_s + B_t - B_s) \mid \mathcal{F}_s](\omega) \\ &= E[h(B_s(\omega) + B_t - B_s)] \\ &= \int h(B_s(\omega) + y)\ N(0, (t-s)\, I)(dy) \\ &\overset{(\leq)}{=} h(B_s(\omega)) \end{aligned}$$

for any $0 \leq s \leq t$ and P-almost every ω. □

3.2 Optional Sampling and Optional Stopping

The Optional Sampling Theorem

The optional stopping theorem can be easily extended to continuous time martingales with continuous sample paths. We directly prove a generalization:

Theorem 3.5 (Optional Sampling Theorem). *Suppose that $(M_t)_{t \in [0,\infty]}$ is a martingale w.r.t. an arbitrary filtration $(\mathcal{F}_t)$ such that $t \mapsto M_t(\omega)$ is continuous for P-almost every ω. Then*

$$E[M_T \mid \mathcal{F}_S] \;=\; M_S \qquad P\text{-almost surely} \tag{3.2.1}$$

for any bounded $(\mathcal{F}_t)$ stopping times S and T with $S \leq T$.

We point out that an additional assumption on the filtration (e.g. right-continuity) is not required in the theorem. Stopping times and the σ-algebra $\mathcal{F}_S$ are defined for arbitrary filtrations in complete analogy to the definitions for the filtration $(\mathcal{F}_t^B)$ in Section 1.5.

Remark (Optional Stopping). By taking expectations in the Optional Sampling Theorem, we obtain

$$E[M_T] \;=\; E[E[M_T \mid \mathcal{F}_0]] \;=\; E[M_0]$$

for any bounded stopping time T. For unbounded stopping times,

$$E[M_T] \;=\; E[M_0]$$

holds by dominated convergence provided $T < \infty$ almost surely, and the random variables $M_{T\wedge n}, n \in \mathbb{N}$, are uniformly integrable.

Proof of Theorem 3.5. We verify the defining properties of the conditional expectation in (3.4) by approximating the stopping times by discrete random variables:

(1). *M_S has an $\mathcal{F}_S$-measurable modification:* For $n \in \mathbb{N}$ let $\widetilde{S}_n = 2^{-n}\lfloor 2^n S\rfloor$, i.e.,

$$\widetilde{S}_n \;=\; k\cdot 2^{-n} \qquad \text{on} \quad \{k\cdot 2^{-n} \le S < (k+1)2^{-n}\} \text{ for any } k = 0,1,2,\ldots.$$

We point out that in general, $\widetilde{S}_n$ is **not** a stopping time w.r.t. $(\mathcal{F}_t)$. Clearly, the sequence $(\widetilde{S}_n)_{n\in\mathbb{N}}$ is *increasing* with $S = \lim S_n$. By almost sure continuity

$$M_S \;=\; \lim_{n\to\infty} M_{\widetilde{S}_n} \qquad P\text{-almost surely.} \tag{3.2.2}$$

On the other hand, each of the random variables $M_{\widetilde{S}_n}$ is $\mathcal{F}_S$-measurable. In fact,

$$M_{\widetilde{S}_n}\cdot I_{\{S\le t\}} \;=\; \sum_{k:k\cdot 2^{-n}\le t} M_{k\cdot 2^{-n}}\cdot I_{\{k2^{-n}\le S<(k+1)2^{-n} \text{ and } S\le t\}}$$

is $\mathcal{F}_t$-measurable for any $t \ge 0$ since S is an $(\mathcal{F}_t)$ stopping time. Therefore, by (3.2.2), the random variable $\widetilde{M_S} := \limsup_{n\to\infty} M_{\widetilde{S}_n}$ is an $\mathcal{F}_S$-measurable modification of M_S.

(2). $E[M_T\,;\,A] \;=\; E[M_S\,;\,A]$ *for any* $A \in \mathcal{F}_S$*:* For $n \in \mathbb{N}$, the discrete random variables $T_n = 2^{-n}\cdot\lceil 2^n T\rceil$ and $S_n = 2^{-n}\cdot\lceil 2^n S\rceil$ are $(\mathcal{F}_t)$ stopping times satisfying $T_n \ge S_n \ge S$, cf. the proof of Theorem 1.26. In particular, $\mathcal{F}_S \subseteq \mathcal{F}_{S_n} \subseteq \mathcal{F}_{T_n}$. Furthermore, (T_n) and (S_n) are *decreasing* sequences with $T = \lim T_n$ and $S = \lim S_n$. As T and S are bounded random variables by assumption, the sequences (T_n) and (S_n) are *uniformly bounded* by a finite constant $c \in (0,\infty)$. Therefore, we obtain

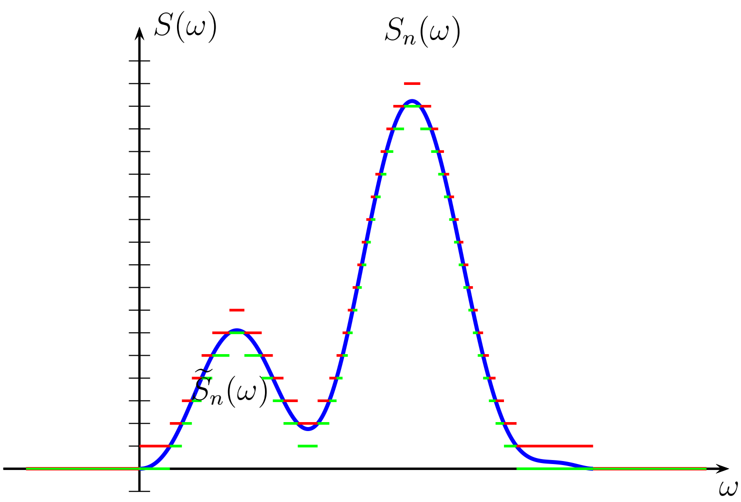


Figure 3.1: Two ways to approximate a continuous stopping time.

$$\begin{aligned} E[M_{T_n}\,;\,A] &= \sum_{k:k\cdot 2^{-n}\le c} E[M_{k\cdot 2^{-n}}\,;\,A\cap\{T_n=k\cdot 2^{-n}\}] \\ &= \sum_{k:k\cdot 2^{-n}\le c} E[M_c\,;\,A\cap\{T_n=k\cdot 2^{-n}\}] \qquad (3.2.3)\\ &= E[M_c\,;\,A] \qquad \text{for any } A\in\mathcal{F}_{T_n}, \end{aligned}$$

and similarly

$$E[M_{S_n}\,;\,A] \quad = \quad E[M_c\,;\,A] \qquad \text{for any } A\in\mathcal{F}_{S_n}. \qquad (3.2.4)$$

In (3.2.3) we have used that (M_t) is an $(\mathcal{F}_t)$ martingale, and $A\cap\{T_n=k\cdot 2^{-n}\}\in\mathcal{F}_{k\cdot 2^{-n}}$. A set $A\in\mathcal{F}_S$ is contained both in $\mathcal{F}_{T_n}$ and $\mathcal{F}_{S_n}$. Thus by (3.2.3) and (3.2.4),

$$E[M_{T_n}\,;\,A] = E[M_{S_n}\,;\,A] \qquad \text{for any } n\in\mathbb{N} \text{ and any } A\in\mathcal{F}_S. \qquad (3.2.5)$$

As $n\to\infty$, $M_{T_n}\to M_T$ and $M_{S_n}\to M_S$ almost surely by continuity. It remains to show that the expectations in (3.2.5) converge as well. To this end note that by (3.2.3) and (3.2.4),

$$M_{T_n} = E[M_c\,|\,\mathcal{F}_{T_n}] \quad \text{and} \quad M_{S_n} = E[M_c\,|\,\mathcal{F}_{S_n}] \qquad P\text{-almost surely.}$$

We will prove in Section 4.3 that any family of conditional expectations of a given random variable w.r.t. different σ-algebras is uniformly integrable, and that for uniformly integrable random variables a generalized Dominated Convergence Theorem holds, cf. Theorem 4.13. Therefore, we finally obtain

$$\begin{aligned} E[M_T\,;\,A] &= E[\lim M_{T_n}\,;\,A] = \lim E[M_{T_n}\,;\,A] \\ &= \lim E[M_{S_n}\,;\,A] = E[\lim M_{S_n}\,;\,A] = E[M_S\,;\,A], \end{aligned}$$

completing the proof of the theorem.

□

Remark (Measurability and completion). In general, the random variable M_S is not necessarily $\mathcal{F}_S$-measurable. However, we have shown in the proof that M_S always has an $\mathcal{F}_S$-measurable modification $\widetilde{M}_S$. If the filtration contains all measure zero sets, then this implies that M_S itself is $\mathcal{F}_S$-measurable and hence a version of $E[M_T \mid \mathcal{F}_S]$.

Ruin probabilities and passage times revisited

Similarly as for random walks, the Optional Sampling Theorem can be applied to compute distributions of passage times and hitting probabilities for Brownian motion. For a one-dimensional Brownian motion (B_t) starting at 0, and $a, b > 0$, let

$$T = \inf\{t \geq 0 \,:\, B_t \notin (-b, a)\} \qquad \text{and} \qquad T_a = \inf\{t \geq 0 \,:\, B_t = a\}$$

denote the first exit time from the interval $(-b, a)$ and the first passage time to the point a, respectively. In Section 1.5 we have computed the distribution of T_a by the reflection principle. This and other results can be recovered by applying optional stopping to the basic martingales of Brownian motion. The advantage of this approach is that it carries over to other diffusion processes.

Exercise (Exit and passage times of Brownian motion). Prove by optional stopping:

(1). *Law of the exit point:* $P[B_T = a] = b/(a+b), \quad P[B_T = -b] = a/(a+b)$,

(2). *Mean exit time:* $E[T] = a \cdot b$ and $E[T_a] = \infty$, (3). *Laplace transform of passage times:* For any $s > 0$,

$$E[\exp(-sT_a)] \;=\; \exp(-a\sqrt{2s}).$$

Conclude that the distribution of T_a on $(0, \infty)$ is absolutely continuous with density

$$f_{T_a}(t) \;=\; a \cdot (2\pi t^3)^{-1/2} \cdot \exp(-a^2/2t).$$

Exit laws and Dirichlet problem

Applying optional stopping to harmonic functions of a multidimensional Brownian motion yields a generalization of the mean value property and a stochastic representation for solutions of the Dirichlet problem. This will be exploited in full generality in Chapter 7. Here, we only sketch the basic idea.

Suppose that $h \in C^2(\mathbb{R}^d)$ is a harmonic function and that $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion starting at x w.r.t. the probability measure P_x. Assuming that

$$E_x[h(B_t)] \;<\; \infty \qquad \text{for any } t \geq 0,$$

the mean value property for harmonic functions implies that $h(B_t)$ is a martingale under P_x, cf. Theorem 3.4. The first hitting time $T = \inf\{t \geq 0 : B_t \in \mathbb{R}^d \setminus D\}$ of the complement of an open set $D \subseteq \mathbb{R}^d$ is a stopping time w.r.t. the filtration $(\mathcal{F}_t^B)$. Therefore, by Theorem 3.5 and the remark below, we obtain

$$E_x[h(B_{T\wedge n})] \;=\; E_x[h(B_0)] \;=\; h(x) \qquad \text{for any } n \in \mathbb{N}. \tag{3.2.6}$$

Now let us assume in addition that the set D is bounded. Then T is almost surely finite, and the sequence of random variables $h(B_{T\wedge n})$ $(n \in \mathbb{N})$ is uniformly bounded because $B_{T\wedge n}$ takes values in the closure $\overline{D}$ for any $n \in \mathbb{N}$. Applying the Dominated Convergence Theorem to (3.2.6), we obtain the integral representation

$$h(x) \;=\; E_x[h(B_T)] \;=\; \int_{\partial D} h(y)\mu_x(dy) \tag{3.2.7}$$

where $\mu_x = P_x \circ B_T^{-1}$ denotes the exit law from D for Brownian motion starting at x. In Chapter 7, we show that the representation (3.2.7) still holds true if h is a continuous function defined on $\overline{D}$ that is C^2 and harmonic on D. The proof requires localization techniques that will be developed below in the context of stochastic calculus. For the moment we note that the representation (3.2.7) has several important aspects and applications:

Generalized mean value property for harmonic functions. For any bounded domain $D \subseteq \mathbb{R}^d$ and any $x \in D$, $h(x)$ is the average of the boundary values of h on ∂D w.r.t. the measure μ_x.

Stochastic representation for solutions of the Dirichlet problem. A solution $h \in C^2(D) \cap C(\overline{D})$ of the Dirichlet problem

$$\begin{aligned} \Delta h(x) &= 0 \qquad \text{for } x \in D, \\ h(x) &= f(x) \qquad \text{for } x \in \partial D, \end{aligned} \tag{3.2.8}$$

has a stochastic representation

$$h(x) = E_x[f(B_T)] \qquad \text{for any } x \in \overline{D}. \tag{3.2.9}$$

Monte Carlo solution of the Dirichlet problem. The stochastic representation (3.2.9) can be used as the basis of a Monte Carlo method for computing the harmonic function $h(x)$ approximately by simulating a large number n of sample paths of Brownian motion starting at x, and estimating the expectation by the corresponding empirical average. Although in many cases classical numerical methods are more efficient, the Monte Carlo method is useful in high dimensional cases. Furthermore, it carries over to far more general situations.

Computation of exit law. Conversely, if the Dirichlet problem (3.2.8) has a unique solution h, then computation of h (for example by standard numerical methods) enables us to obtain the expectations in (3.2.8). In particular, the probability $h(x) = P_x[B_T \in A]$ for Brownian motion exiting the domain on a subset $A \subseteq \partial D$ is informally given as the solution of the Dirichlet problem

$$\Delta h = 0 \quad \text{on } D, \qquad h = I_A \quad \text{on } \partial D.$$

This can be made rigorous under regularity assumptions. The full exit law is the *harmonic measure*, i.e., the probability measure μ_x such that the representation (3.2.7) holds for any function $h \in C^2(D) \cap C(\overline{D})$ with $\Delta h = 0$ on D. For simple domains such as half-spaces, balls and cylinders, this harmonic measure can be computed explicitly.

Example (Exit laws from balls). For $d \geq 2$, the exit law from the unit ball $D = \{y \in \mathbb{R}^d : |y| < 1\}$ for Brownian motion starting at a point $x \in \mathbb{R}^d$ with $|x| < 1$ is given by

$$\mu_x(dy) = \frac{1 - |x|^2}{|y - x|^d} \nu(dy)$$

where ν denotes the normalized surface measure on the unit sphere $S^{d-1} = \{y \in \mathbb{R}^d : |y| = 1\}$. Indeed, the classical Poisson integral formula states that for any $f \in C(S^{d-1})$, the function

$$h(x) = \int f(y)\, \mu_x(dy)$$

solves the Dirichlet problem on D with boundary values $\lim_{x \to z} h(x) = f(z)$ for any $z \in S^{d-1}$, cf. e.g. [XXX Karatzas/Shreve, Ch. 4]. Hence by (3.2.9),

$$E_x[f(B_T)] = \int f(y)\, \frac{1 - |x|^2}{|y - x|^d} \nu(dy)$$

holds for any $f \in C(S^{d-1})$, and thus by a standard approximation argument, for any indicator function of a measurable subset of S^{d-1}.

3.3 Maximal inequalities and the Law of the Iterated Logarithm

The extension of Doob's maximal inequality to the continuous time case is straightforward. As a first application, we give a proof for the upper bound in the law of the iterated logarithm.

Maximal inequalities in continuous time

Theorem 3.6 (Doob's L^p inequality in continuous time). *Suppose that $(M_t)_{t\in[0,\infty)}$ is a martingale with almost surely right continuous sample paths $t \mapsto M_t(\omega)$. Then the following estimates hold for any $a \in [0,\infty)$, $p \in [1,\infty)$, $q \in (1,\infty]$ with $\frac{1}{p}+\frac{1}{q}=1$, and $c > 0$:*

(1). $$P\left[\sup_{t\in[0,a]} |M_t| \ \geq\ c\right] \quad\leq\quad c^{-p}\cdot E[|M_a|^p],$$

(2). $$\left\|\sup_{t\in[0,a]} |M_t|\right\|_{L^p} \quad\leq\quad q\cdot \|M_a\|_{L^p}.$$

Remark. The same estimates hold for non-negative submartingales.

Proof. Let (π_n) denote an increasing sequence of partitions of the interval $[0,a]$ such that the mesh size of π_n goes to 0 as $n\to\infty$. By Corollary 2.13 applied to the discrete time martingale $(M_t)_{t\in\pi_n}$, we obtain

$$P\left[\max_{t\in\pi_n}|M_t| \geq c\right] \ \leq\ E[|M_a|^p]/c^p \qquad \text{for any } n\in\mathbb{N}.$$

Moreover, as $n\to\infty$,

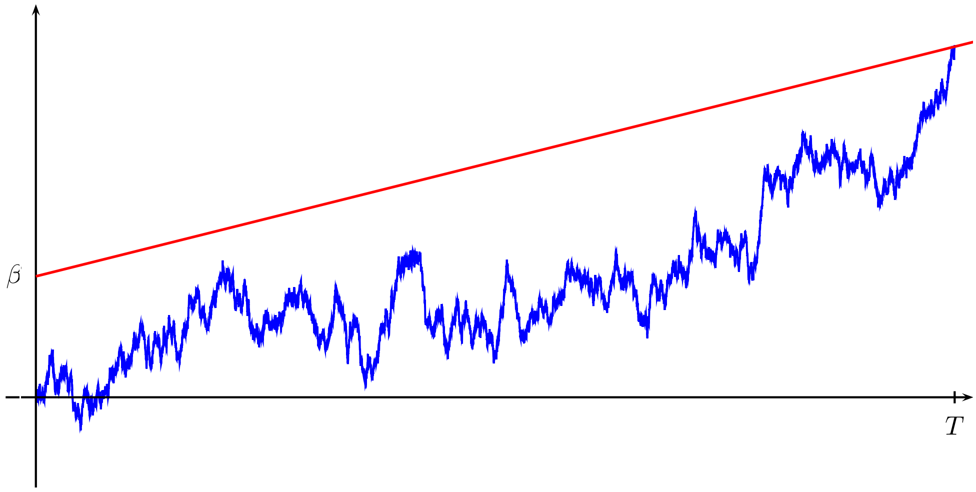
$$\max_{t\in\pi_n}|M_t| \quad\nearrow\quad \sup_{t\in[0,a]}|M_t| \qquad \text{almost surely}$$

by right continuity of the sample paths. Hence

$$\begin{aligned} P\left[\sup_{t\in[0,a]}|M_t| > c\right] &= P\left[\bigcup_n \left\{\max_{t\in\tau_n}|M_t| > c\right\}\right] \\ &= \lim_{n\to\infty} P\left[\max_{t\in\tau_n}|M_t| > c\right] \quad\leq\quad E[|M_a|^p]/c^p. \end{aligned}$$

The first assertion now follows by replacing c by $c-\varepsilon$ and letting ε tend to 0. The second assertion follows similarly from Theorem 2.15. □

As a first application of the maximal inequality to Brownian motion, we derive an upper bound for the probability that the graph of one-dimensional Brownian motion passes a line in $\mathbb{R}^2$:



Lemma 3.7 (Passage probabilities for lines). *For a one-dimensional Brownian motion (B_t) starting at 0 we have*

$$P[B_t \geq \beta + \alpha t/2 \quad \textit{for some } t \geq 0] \;\leq\; \exp(-\alpha\beta) \qquad \textit{for any } \alpha, \beta > 0.$$

Proof. Applying the maximal inequality to the exponential martingale

$$M_t^\alpha \;=\; \exp(\alpha B_t - \alpha^2 t/2)$$

yields

$$\begin{aligned} P[B_t \geq \beta + \alpha t/2 \quad \text{for some } t \in [0,a]] \;&=\; P\left[\sup_{t\in[0,a]} (B_t - \alpha t/2) \geq \beta\right] \\ =\; P\left[\sup_{t\in[0,a]} M_t^\alpha \geq \exp(\alpha\beta)\right] \;&\leq\; \exp(-\alpha\beta) \cdot E[M_a^\alpha] \;=\; \exp(-\alpha\beta) \end{aligned}$$

for any $a > 0$. The assertion follows in the limit as $a \to \infty$. □

With slightly more effort, it is possible to compute the passage probability and the distribution of the first passage time of a line explicitly, cf. **??** below.

Application to LIL

A remarkable consequence of Lemma 3.7 is a simplified proof for the upper bound half of the Law of the Iterated Logarithm:

Theorem 3.8 (LIL, upper bound). *For a one-dimensional Brownian motion (B_t) starting at 0,*

$$\limsup_{t \searrow 0} \frac{B_t}{\sqrt{2t \log \log t^{-1}}} \quad \leq \quad +1 \qquad P\text{-almost surely.} \tag{3.3.1}$$

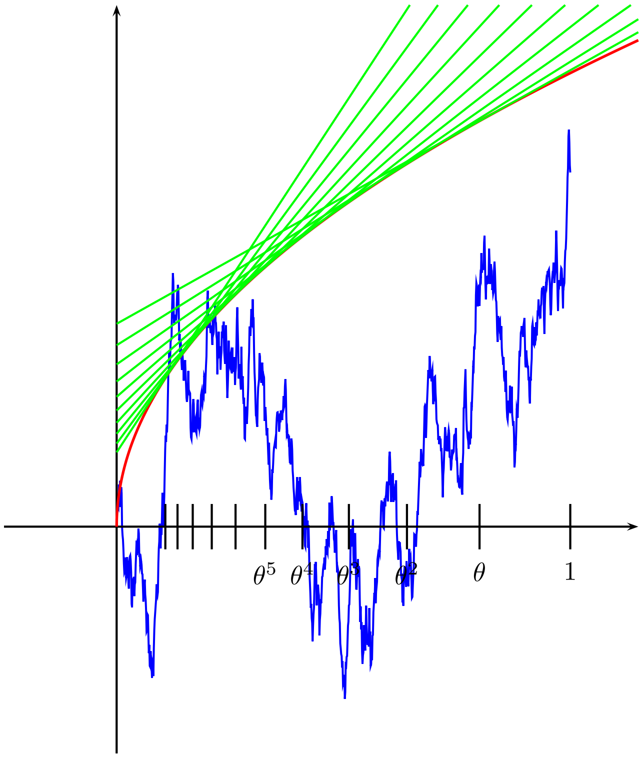
Proof. Let $\delta > 0$. We would like to show that almost surely,

$$B_t \;\leq\; (1+\delta)h(t) \qquad \text{for sufficiently small } t > 0,$$

where $h(t) := \sqrt{2t \log \log t^{-1}}$. Fix $\theta \in (0,1)$. The idea is to approximate the function $h(t)$ by affine functions

$$l_n(t) \;=\; \beta_n + \alpha_n t/2$$

on each of the intervals $[\theta^n, \theta^{n-1}]$, and to apply the upper bounds for the passage probabilities from the lemma.

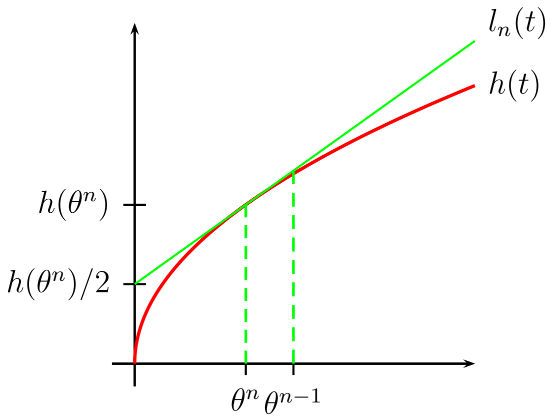


We choose α_n and β_n in a such way that $l_n(\theta^n) = h(\theta^n)$ and $l_n(0) = h(\theta^n)/2$, i.e.,

$$\beta_n \;=\; h(\theta^n)/2 \quad \text{and} \quad \alpha_n \;=\; h(\theta^n)/\theta^n.$$

For this choice we have $l_n(\theta^n) \geq \theta \cdot l_n(\theta^{n-1})$, and hence

$$\begin{aligned} l_n(t) \;&\leq\; l_n(\theta^{n-1}) \quad \leq \quad \frac{l_n(\theta^n)}{\theta} \\ &=\; \frac{h(\theta^n)}{\theta} \quad \leq \quad \frac{h(t)}{\theta} \qquad \text{for any } t \in [\theta^n, \theta^{n-1}]. \end{aligned} \tag{3.3.2}$$



We now want to apply the Borel-Cantelli lemma to show that with probability one, $B_t \le (1+\delta)l_n(t)$ for large n. By Lemma 3.7,

$$\begin{aligned} P[B_t \ge (1+\delta)l_n(t) \quad \text{for some } t \ge 0] &\le \exp(-\alpha_n\beta_n \cdot (1+\delta)^2) \\ &= \exp\left(-\frac{h(\theta^n)^2}{2\theta^n} \cdot (1+\delta)^2\right). \end{aligned}$$

Choosing $h(t) = \sqrt{2t \log\log t^{-1}}$, the right hand side is equal to a constant multiple of $n^{-(1+\delta)^2}$, which is a summable sequence. Note that we do not have to know the precise form of $h(t)$ in advance to carry out the proof – we just choose $h(t)$ in such a way that the probabilities become summable!

Now, by Borel-Cantelli, for P-almost every ω there exists $N(\omega) \in \mathbb{N}$ such that

$$B_t(\omega) \quad \le \quad (1+\delta)l_n(t) \qquad \text{for any } t \in [0,1] \text{ and } n \ge N(\omega). \tag{3.3.3}$$

By (3.3.2), the right hand side of (3.3.3) is dominated by $(1+\delta)h(t)/\theta$ for $t \in [\theta^n, \theta^{n-1}]$. Hence

$$B_t \le \frac{1+\delta}{\theta} h(t) \qquad \text{for any } t \in \bigcup_{n \ge N} [\theta^n, \theta^{n-1}],$$

i.e., for any $t \in (0, \theta^{N-1})$, and therefore,

$$\limsup_{t \searrow 0} \frac{B_t}{h(t)} \le \frac{1+\delta}{\theta} \qquad P\text{-almost surely.}$$

The assertion then follows in the limit as $\theta \nearrow 1$ and $\delta \searrow 0$. $\square$

Since $(-B_t)$ is again a Brownian motion starting at 0, the upper bound (3.3.1) also implies

$$\liminf_{t\searrow 0} \frac{B_t}{\sqrt{2t\log\log t^{-1}}} \;\geq\; -1 \qquad P\text{-almost surely.} \tag{3.3.4}$$

The converse bounds are actually easier to prove since we can use the independence of the increments and apply the second Borel-Cantelli Lemma. We only mention the key steps and leave the details as an exercise:

Exercise (Complete proof of LIL). Prove the Law of the Iterated Logarithm:

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \;=\; +1 \qquad \text{and} \qquad \liminf_{t\searrow 0} \frac{B_t}{h(t)} \;=\; -1$$

where $h(t) = \sqrt{2t\log\log t^{-1}}$. Proceed in the following way:

(1). Let $\theta \in (0,1)$ and consider the increments $Z_n = B_{\theta^n} - B_{\theta^{n+1}}, n \in \mathbb{N}$. Show that if $\varepsilon > 0$, then

$$P[Z_n > (1-\varepsilon)h(\theta^n) \quad \text{infinitely often}] \;=\; 1.$$

(Hint: $\int_x^\infty \exp(-z^2/2)dz \;\leq\; x^{-1}\exp(-x^2/2)$)

(2). Conclude that by (3.3.4),

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \;\geq\; 1-\varepsilon \qquad P\text{-almost surely for any } \varepsilon > 0,$$

and complete the proof of the LIL by deriving the lower bounds

$$\limsup_{t\searrow 0} \frac{B_t}{h(t)} \;\geq\; 1 \quad \text{and} \quad \liminf_{t\searrow 0} \frac{B_t}{h(t)} \;\leq\; -1 \qquad P\text{-almost surely.} \tag{3.3.5}$$

Chapter 4

Martingale Convergence Theorems

The strength of martingale theory is partially due to powerful general convergence theorems that hold for martingales, sub- and supermartingales. In this chapter, we study convergence theorems with different types of convergence including almost sure, L^2 and L^1 convergence, and consider first applications.

At first, we will again focus on discrete-parameter martingales – the results can then be easily extended to continuous martingales.

4.1 Convergence in L^2

Already when proving the Law of Large Numbers, L^2 convergence is much easier to show than, for example, almost sure convergence. The situation is similar for martingales: A necessary and sufficient condition for convergence in the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ can be obtained by elementary methods.

Martingales in L^2

Consider a discrete-parameter martingale $(M_n)_{n\geq 0}$ w.r.t. a filtration $(\mathcal{F}_n)$ on a probability space $(\Omega, \mathcal{A}, P)$. Throughout this section we assume:

Assumption (Square integrability). $E[M_n^2] < \infty$ *for any* $n \geq 0$.

We start with an important remark:

Lemma 4.1. *The increments $Y_n = M_n - M_{n-1}$ of a square-integrable martingale are centered and orthogonal in $L^2(\Omega, \mathcal{A}, P)$ (i.e. uncorrelated).*

Proof. By definition of a martingale, $E[Y_n \,|\, \mathcal{F}_{n-1}] = 0$ for any $n \geq 0$. Hence $E[Y_n] = 0$ and $E[Y_m Y_n] = E[Y_m \cdot E[Y_n \mid \mathcal{F}_{n-1}]] = 0$ for $0 \leq m < n$. □

Since the increments are also orthogonal to M_0 by an analogue argument, a square integrable martingale sequence consists of partial sums of a sequence of uncorrelated random variables:

$$M_n \;=\; M_0 + \sum_{k=1}^{n} Y_k \qquad \text{for any } n \geq 0.$$

The Convergence Theorem

The central result of this section shows that an L^2-bounded martingale (M_n) can **always** be extended to $n \in \{0, 1, 2, \ldots\} \cup \{\infty\}$:

Theorem 4.2 ($\mathbf{L^2}$ Martingale Convergence Theorem). *The martingale sequence (M_n) converges in $L^2(\Omega, \mathcal{A}, P)$ as $n \to \infty$ if and only if it is bounded in L^2 in the sense that*

$$\sup_{n\geq 0} E[M_n^2] < \infty. \tag{4.1.1}$$

In this case, the representation

$$M_n \;=\; E[M_\infty \mid \mathcal{F}_n]$$

holds almost surely for any $n \geq 0$, where M_∞ denotes the limit of M_n in $L^2(\Omega, \mathcal{A}, P)$.

We will prove in the next section that (M_n) does also converge almost surely to M_∞. An analogue result to Theorem 4.2 holds with L^2 replaced by L^p for any $p \in (1, \infty)$ but not for $p = 1$, cf. Section 4.3 below.

Proof. (1). Let us first note that

$$E[(M_n - M_m)^2] \;=\; E[M_n^2] - E[M_m^2] \qquad \text{for } 0 \leq m \leq n. \tag{4.1.2}$$

Indeed,

$$\begin{aligned} E[M_n^2]-E[M_m^2] &= E[(M_n-M_m)(M_n+M_m)] \\ &= E[(M_n-M_m)^2]+2E[M_m\cdot(M_n-M_m)], \end{aligned}$$

and the last term vanishes since the increment $M_n - M_m$ is orthogonal to M_m in L^2.

(2). To prove that (4.1.1) is sufficient for L^2 convergence, note that the sequence $(E[M_n^2])_{n\geq 0}$ is increasing by (4.1.2). If (4.1.1) holds then this sequence is bounded, and hence a Cauchy sequence. Therefore, by (4.1.2), (M_n) is a Cauchy sequence in L^2. Convergence now follows by completeness of $L^2(\Omega,\mathcal{A},P)$.

(3). Conversely, if (M_n) converges in L^2 to a limit M_∞, then the L^2 norms are bounded. Moreover, by Jensen's inequality, for each fixed $k\geq 0$,

$$E[M_n \mid \mathcal{F}_k] \longrightarrow E[M_\infty \mid \mathcal{F}_k] \qquad \text{in } L^2(\Omega,\mathcal{A},P) \text{ as } n\to\infty.$$

As (M_n) is a martingale, we have $E[M_n \mid \mathcal{F}_k] = M_k$ for $n \geq k$, and hence

$$M_k = E[M_\infty \mid \mathcal{F}_k] \qquad P\text{-almost surely.}$$

□

Remark (Functional analytic interpretation of L^2 convergence theorem). The assertion of the L^2 martingale convergence theorem can be rephrased as a purely functional analytic statement:

An infinite sum $\sum\limits_{k=1}^{\infty} Y_k$ of orthogonal vectors Y_k in the Hilbert space $L^2(\Omega,\mathcal{A},P)$ is convergent if and only if the sequence of partial sums $\sum\limits_{k=1}^{n} Y_k$ is bounded.

How can boundedness in L^2 be verified for martingales? Writing the martingale (M_n) as the sequence of partial sums of its increments $Y_n = M_n - M_{n-1}$, we have

$$E[M_n^2] = \left(M_0 + \sum_{k=1}^{n} Y_k, M_0 + \sum_{k=1}^{n} Y_k \right)_{L^2} = E[M_0^2] + \sum_{k=1}^{n} E[Y_k^2]$$

by orthogonality of the increments and M_0. Hence

$$\sup_{n\geq 0} E[M_n^2] \;=\; E[M_0^2] + \sum_{k=1}^{\infty} E[Y_k^2].$$

Alternatively, we have $E[M_n^2] = E[M_0^2] + E[\langle M\rangle_n]$. Hence by monotone convergence

$$\sup_{n\geq 0} E[M_n^2] \;=\; E[M_0^2] + E[\langle M\rangle_\infty]$$

where $\langle M\rangle_\infty \;=\; \sup \langle M\rangle_n$.

Summability of sequences with random signs

As a first application we study the convergence of series with coefficients with random signs. In an introductory analysis course it is shown as an application of the integral and Leibniz criterion for convergence of series that

$$\begin{aligned} \sum_{n=1}^{\infty} n^{-\alpha} \text{ converges} \quad &\Longleftrightarrow \quad \alpha > 1 \quad \text{, whereas} \\ \sum_{n=1}^{\infty} (-1)^n n^{-\alpha} \text{ converges} \quad &\Longleftrightarrow \quad \alpha > 0. \end{aligned}$$

Therefore, it seems interesting to see what happens if the signs are chosen randomly. The L^2 martingale convergence theorem yields:

Corollary 4.3. *Let (a_n) be a real sequence. If (ε_n) is a sequence of independent random variables on $(\Omega, \mathcal{A}, P)$ with $P[\varepsilon_n = +1] = P[\varepsilon_n = -1] = 1/2$, then*

$$\sum_{n=1}^{\infty} \varepsilon_n a_n \quad \textit{converges in } L^2(\Omega, \mathcal{A}, P) \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} a_n^2 < \infty.$$

Proof. The sequence $M_n = \sum_{k=1}^{n} \varepsilon_k a_k$ of partial sums is a martingale with

$$\sup_{n\geq 0} E[M_n^2] \;=\; \sum_{k=1}^{\infty} E[\varepsilon_k^2 a_k^2] \;=\; \sum_{k=1}^{\infty} a_k^2.$$

□

Example. The series $\sum_{n=1}^{\infty} \varepsilon_n \cdot n^{-\alpha}$ converges in L^2 if and only if $\alpha > \frac{1}{2}$.

Remark (Almost sure asymptotics). By the Supermartingale Convergence Theorem (cf. Theorem 4.5 below), the series $\sum \varepsilon_n a_n$ also converges almost surely if $\sum a_n^2 < \infty$. On the other hand, if $\sum a_n^2 = \infty$ then the series of partial sums has almost surely unbounded oscillations:

Exercise. Suppose that $\sum a_n = \infty$, and let $M_n = \sum_{k=1}^{n} \varepsilon_k a_k$.

(1). Compute the conditional variance process $\langle M \rangle_n$.

(2). For $c > 0$ let $T_c = \inf\{n \geq 0 \ : \ |M_n| \geq c\}$. Apply the Optional Stopping Theorem to the martingale in the Doob decomposition of (M_n^2), and conclude that $P[T_c = \infty] = 0$.

(3). Prove that (M_n) has almost surely unbounded oscillations.

$\mathbf{L^2}$ convergence in continuous time

The L^2 convergence theorem directly extends to the continuous-parameter case.

Theorem 4.4 ($\mathbf{L^2}$ Martingale Convergence Theorem in continuous time). *Let $a \in (0, \infty]$. If $(M_t)_{t \in [0,a)}$ is a martingale w.r.t. a filtration $(\mathcal{F}_t)_{t \in [0,a)}$ such that*

$$\sup_{t \in [0,u)} E[M_t^2] \ < \ \infty$$

then $M_u = \lim_{t \nearrow u} M_t$ exists in $L^2(\Omega, \mathcal{A}, P)$ and $(M_t)_{t \in [0,u]}$ is again a square-integrable martingale.

Proof. Choose any increasing sequence $t_n \in [0, u)$ such that $t_n \to u$. Then (M_{t_n}) is an L^2-bounded discrete-parameter martingale. Hence the limit $M_u = \lim M_{t_n}$ exists in L^2, and

$$M_{t_n} \ = \ E[M_u \mid \mathcal{F}_{t_n}] \qquad \text{for any } n \in \mathbb{N}. \tag{4.1.3}$$

For an arbitrary $t \in [0, u)$, there exists $n \in \mathbb{N}$ with $t_n \in (t, u)$. Hence

$$M_t \;=\; E[M_{t_n} \mid \mathcal{F}_t] \;=\; E[M_u \mid \mathcal{F}_t]$$

by (4.1.3) and the tower property. In particular, $(M_t)_{t\in[0,u]}$ is a square-integrable martingale. By orthogonality of the increments,

$$E[(M_u - M_{t_n})^2] \;=\; E[(M_u - M_t)^2] + E[(M_t - M_{t_n})^2] \;\geq\; E[(M_u - M_t)^2]$$

whenever $t_n \leq t \leq u$. Since $M_{t_n} \to M_u$ in L^2, we obtain

$$\lim_{t\nearrow u} E[(M_u - M_t)^2] \;=\; 0.$$

□

Remark. (1). Note that in the proof it is enough to consider a fixed sequence $t_n \nearrow u$.

(2). To obtain almost sure convergence, an additional regularity condition on the sample paths (e.g. right-continuity) is required, cf. below. This assumption is not needed for L^2 convergence.

4.2 Almost sure convergence of supermartingales

Let $(Z_n)_{n\geq 0}$ be a discrete-parameter supermartingale w.r.t. a filtration $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$. The following theorem yields a stochastic counterpart to the fact that any lower bounded decreasing sequence of reals converges to a finite limit:

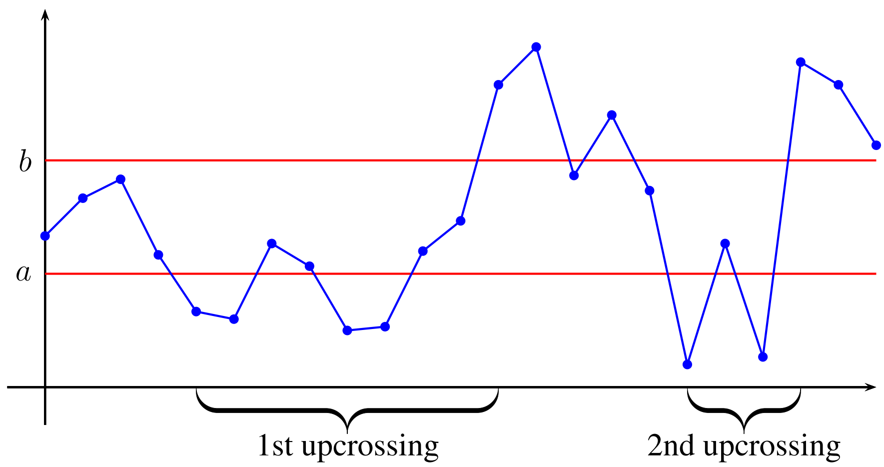
Theorem 4.5 (Supermartingale Convergence Theorem, Doob). *If* $\sup_{n\geq 0} E[Z_n^-] < \infty$ *then* (Z_n) *converges almost surely to a random variable* $Z_\infty \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$.
In particular, supermartingales that are uniformly bounded from below converge almost surely to an integrable random variable.

Remark ($\mathbf{L^1}$ boundedness vs. $\mathbf{L^1}$ convergence). (1). The condition $\sup E[Z_n^-] < \infty$ holds if and only if (Z_n) is bounded in L^1. Indeed, as $E[Z_n^+] < \infty$ by our definition of a supermartingale, we have

$$E[\,|Z_n|\,] \;=\; E[Z_n] + 2E[Z_n^-] \;\leq\; E[Z_0] + 2E[Z_n^-] \qquad \text{for any } n \geq 0.$$

(2). Although (Z_n) is bounded in L^1 and the limit is integrable, L^1 convergence does **not** hold in general, cf. the examples below.

For proving the Supermartingale Convergence Theorem, we introduce the number $U^{(a,b)}(\omega)$ of upcrossings of an interval (a,b) by the sequence $Z_n(\omega)$, cf. below for the exact definition.



Note that if $U^{(a,b)}(\omega)$ is finite for every non-empty bounded interval $[a,b]$ then $\limsup Z_n(\omega)$ and $\liminf Z_n(\omega)$ coincide, i.e., the sequence $(Z_n(\omega))$ converges. Therefore, to show almost sure convergence of (Z_n), we derive an upper bound for $U^{(a,b)}$. We first prove this key estimate and then complete the proof of the theorem.

Doob's upcrossing inequality

For $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $a < b$, we define the number $U_n^{(a,b)}$ of upcrossings of the interval $[a,b]$ before time n by

$$U_n^{(a,b)} \;=\; \max\big\{k \geq 0 : \exists\, 0 \leq s_1 < t_1 < s_2 < t_2 \ldots < s_k < t_k \leq n \;: \\ Z_{s_i}(\omega) \leq a,\, Z_{t_i}(\omega) \geq b\big\}.$$

Lemma 4.6 (Doob). *If (Z_n) is a supermartingale then*

$$(b-a)\cdot E[U_n^{(a,b)}] \;\leq\; E[(Z_n-a)^-] \qquad \textit{for any } a<b \textit{ and } n\geq 0.$$

Proof. We may assume $E[Z_n^-]<\infty$ since otherwise there is nothing to prove. The key idea is to set up a predictable gambling strategy that increases our capital by $(b-a)$ for each completed upcrossing. Since the net gain with this strategy should again be a supermartingale this yields an upper bound for the average number of upcrossings. Here is the strategy:

repeat

- Wait until $Z_k \leq a$.
- Then play unit stakes until $Z_k \geq b$.
-

The stake C_k in round k is

$$C_1 \;=\; \begin{cases} 1 & \text{if } Z_0 \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

and for $k\geq 2$,

$$C_k \;=\; \begin{cases} 1 & \text{if } (C_{k-1}=1 \text{ and } Z_{k-1}\leq b) \text{ or } (C_{k-1}=0 \text{ and } Z_{k-1}\leq a), \\ 0 & \text{otherwise} \end{cases}.$$

Clearly, (C_k) is a predictable, bounded and non-negative sequence of random variables. Moreover, $C_k\cdot(Z_k-Z_{k-1})$ is integrable for any $k\leq n$, because C_k is bounded and

$$E\big[\,|Z_k|\,\big] \;=\; 2E[Z_k^+]-E[Z_k] \;\leq\; 2E[Z_k^+]-E[Z_n] \leq 2E[Z_k^+]-E[Z_n^-]$$

for $k\leq n$. Therefore, by Theorem 2.6 and the remark below, the process

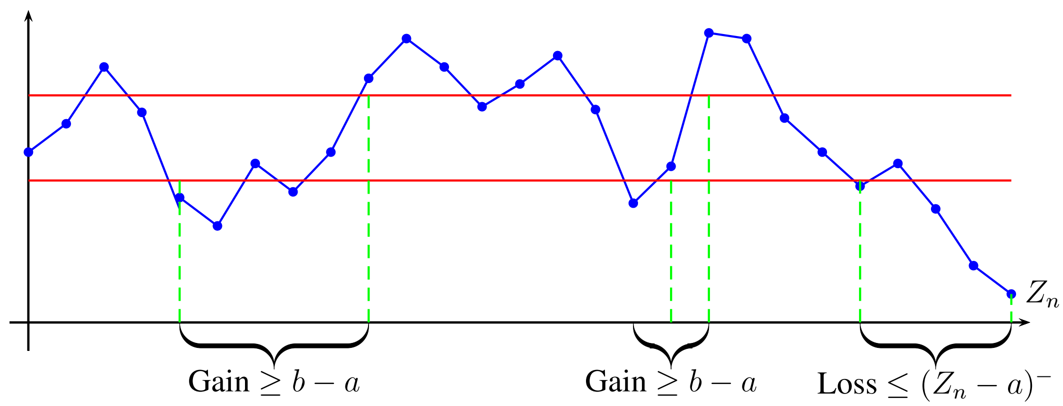
$$(C_\bullet Z)_k \;=\; \sum_{i=1}^{k} C_i\cdot(Z_i-Z_{i-1}), \qquad 0\leq k\leq n,$$

is again a supermartingale.

Clearly, the value of the process $C_\bullet Z$ increases by at least $(b-a)$ units during each completed upcrossing. Between upcrossing periods, the value of $(C_\bullet Z)_k$ is constant. Finally, if the final time n is contained in an upcrossing period, then the process can decrease by at most $(Z_n - a)^-$ units during that last period (since Z_k might decrease before the next upcrossing is completed). Therefore, we have

$$(C_\bullet Z)_n \quad \geq \quad (b-a) \cdot U_n^{(a,b)} - (Z_n - a)^-, \qquad \text{i.e.,}$$

$$(b-a) \cdot U_n^{(a,b)} \quad \leq \quad (C_\bullet Z)_n + (Z_n - a)^-.$$



Since $C_\bullet Z$ is a supermartingale with initial value 0, we obtain the upper bound

$$(b-a)E[U_n^{(a,b)}] \ \leq \ E[(C_\bullet Z)_n] + E[(Z_n - a)^-] \ \leq \ E[(Z_n - a)^-].$$

□

Proof of Doob's Convergence Theorem

We can now complete the proof of Theorem 4.5.

Proof. Let

$$U^{(a,b)} \ = \ \sup_{n \in \mathbb{N}} U_n^{(a,b)}$$

denote the total number of upcrossings of the supermartingale (Z_n) over an interval (a,b) with $-\infty < a < b < \infty$. By the upcrossing inequality and monotone convergence,

$$E[U^{(a,b)}] \ = \ \lim_{n\to\infty} E[U_n^{(a,b)}] \ \leq \ \frac{1}{b-a} \cdot \sup_{n\in\mathbb{N}} E[(Z_n - a)^-]. \qquad (4.2.1)$$

Assuming $\sup E[Z_n^-] < \infty$, the right hand side of (4.2.1) is finite since $(Z_n - a)^- \leq |a| + Z_n^-$. Therefore,

$$U^{(a,b)} \ < \ \infty \qquad P\text{-almost surely,}$$

and hence the event

$$\{\liminf Z_n \ \neq \ \limsup Z_n\} \ = \ \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{U^{(a,b)} = \infty\}$$

has probability zero. This proves almost sure convergence.

It remains to show that the almost sure limit $Z_\infty = \lim Z_n$ is an integrable random variable (in particular, it is finite almost surely). This holds true as, by the remark below Theorem 4.5, $\sup E[Z_n^-] < \infty$ implies that (Z_n) is bounded in L^1, and therefore

$$E[\,|Z_\infty|\,] \ = \ E[\liminf \ |Z_n|\,] \ \leq \ \liminf E[\,|Z_n|\,] \ < \ \infty$$

by Fatou's lemma. □

Examples and first applications

We now consider a few prototypic applications of the almost sure convergence theorem:

Example (1. Sums of i.i.d. random variables). Consider a Random Walk

$$S_n = \sum_{i=1}^{n} \eta_i$$

on $\mathbb{R}$ with centered and bounded increments

$$\eta_i \ \text{ i.i.d. with } |\eta_i| \leq c \text{ and } E[\eta_i] = 0, \qquad c \in \mathbb{R}.$$

Suppose that $P[\eta_i \neq 0] > 0$. Then there exists $\varepsilon > 0$ such that $P[|\eta_i| \geq \varepsilon] > 0$. As the increments are i.i.d., the event $\{|\eta_i| \geq \varepsilon\}$ occurs infinitely often with probability one. Therefore, almost surely, the martingale (S_n) does not converge as $n \to \infty$.

Now let $a \in \mathbb{R}$. We consider the first hitting time

$$T_a \ = \ \inf\{t \geq 0 \ : \ S_n \geq a\}$$

of the interval $[a, \infty)$. By the Optional Stopping Theorem, the stopped Random Walk $(S_{T_a \wedge n})_{n \geq 0}$ is again a martingale. Moreover, as $S_k < a$ for any $k < T_a$ and the increments are bounded by c, we obtain the upper bound

$$S_{T_a \wedge n} \; < \; a + c \qquad \text{for any } n \in \mathbb{N}.$$

Therefore, the stopped Random Walk converges almost surely by the Supermartingale Convergence Theorem. As (S_n) does not converge, we can conclude that

$$P[T_a < \infty] \; = \; 1 \qquad \text{for any } a > 0, \text{ i.e., } \limsup S_n \; = \; \infty \qquad \text{almost surely.}$$

Since (S_n) is also a submartingale, we obtain

$$\liminf S_n \; = \; -\infty \qquad \text{almost surely}$$

by an analogue argument. A generalization of this result is given in Theorem 4.7 below.

Remark (Almost sure vs. $\mathbf{L^p}$ convergence). In the last example, the stopped process does not converge in L^p for any $p \in [1, \infty)$. In fact,

$$\lim_{n \to \infty} E[S_{T_a \wedge n}] \; = \; E[S_{T_a}] \; \geq \; a \qquad \text{whereas} \quad E[S_{T_a \wedge n}] \; = \; E[S_0] \; = \; 0 \text{ for all } n.$$

Example (2. Products of non-negative i.i.d. random variables). Consider a growth process

$$Z_n \quad = \quad \prod_{i=1}^{n} Y_i$$

with i.i.d. factors $Y_i \geq 0$ with finite expectation $\alpha \in (0, \infty)$. Then

$$M_n \quad = \quad Z_n / \alpha^n$$

is a martingale. By the almost sure convergence theorem, a finite limit M_∞ exists almost surely, because $M_n \geq 0$ for all n. For the almost sure asymptotics of (Z_n), we distinguish three different cases:

(1). $\alpha < 1$: In this case,

$$Z_n \quad = \quad M_n \cdot \alpha^n$$

converges to 0 exponentially fast with probability one.

(2). $\alpha = 1$*:* Here (Z_n) is a martingale and converges almost surely to a finite limit. If $P[Y_i \neq 1] > 0$ then there exists $\varepsilon > 0$ such that $Y_i \geq 1 + \varepsilon$ infinitely often with probability one. This is consistent with convergence of (Z_n) only if the limit is zero. Hence, if (Z_n) is not almost surely constant, then also in the critical case, $Z_n \to 0$ almost surely.

(3). $\alpha > 1$ *(supercritical):* In this case, on the set $\{M_\infty > 0\}$,

$$Z_n \;=\; M_n \cdot \alpha^n \quad \sim \quad M_\infty \cdot \alpha^n,$$

i.e., (Z_n) grows exponentially fast. The asymptotics on the set $\{M_\infty = 0\}$ is not evident and requires separate considerations depending on the model.

Although most of the conclusions in the last example could have been obtained without martingale methods (e.g. by taking logarithms), the martingale approach has the advantage of carrying over to far more general model classes. These include for example branching processes or exponentials of continuous time processes.

Example (3. Boundary behaviors of harmonic functions). Let $D \subseteq \mathbb{R}^d$ be a bounded open domain, and let $h : D \to \mathbb{R}$ be a harmonic function on D that is bounded from below:

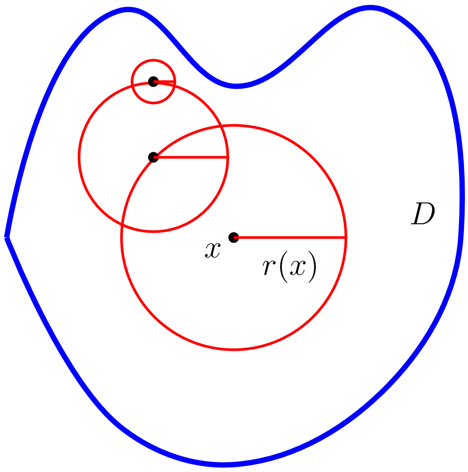
$$\Delta h(x) \;=\; 0 \quad \text{for any } x \in D, \qquad \inf_{x \in D} h(x) > -\infty. \tag{4.2.2}$$

To study the asymptotic behavior of $h(x)$ as x approaches the boundary ∂D, we construct a Markov chain (X_n) such that $h(X_n)$ is a martingale: Let $r : D \to (0, \infty)$ be a continuous function such that

$$0 \;<\; r(x) \;<\; \operatorname{dist}(x, \partial D) \qquad \text{for any } x \in D, \tag{4.2.3}$$

and let (X_n) w.r.t P_x denote the canonical time-homogeneous Markov chain with state space D, initial value x, and transition probabilities

$$p(x, dy) \;=\; \text{Uniform distribution on } \{y \in \mathbb{R}^d \;:\; |y - x| = r(x)\}.$$



By (4.2.3), the function h is integrable w.r.t. $p(x, dy)$, and, by the mean value property,

$$(ph)(x) \;=\; h(x) \qquad \text{for any } x \in D.$$

Therefore, the process $h(X_n)$ is a martingale w.r.t. P_x for each $x \in D$. As $h(X_n)$ is lower bounded by (4.2.2), the limit as $n \to \infty$ exists P_x-almost surely by the Supermartingale Convergence Theorem. In particular, since the coordinate functions $x \mapsto x_i$ are also harmonic and lower bounded on $\overline{D}$, the limit $X_\infty = \lim_{n\to\infty} X_n$ exists P_x-almost surely. Moreover, X_∞ is in ∂D, because r is bounded from below by a strictly positive constant on any compact subset of D.

Summarizing we have shown:

(1). *Boundary regularity:* If h is harmonic and bounded from below on D then the limit $\lim_{n\to\infty} h(X_n)$ exists along almost every trajectory X_n to the boundary ∂D.

(2). *Representation of h in terms of boundary values:* If h is continuous on $\overline{D}$, then $h(X_n) \to h(X_\infty)$ P_x-almost surely and hence

$$h(x) \;=\; \lim_{n\to\infty} E_x[h(X_n)] \;=\; E[h(X_\infty)],$$

i.e., the law of X_∞ w.r.t. P_x is the harmonic measure on ∂D.

Note that, in contrast to classical results from analysis, the first statement holds without any smoothness condition on the boundary ∂D. Thus, although boundary values of h may not exist in the classical sense, they do exist along almost every trajectory of the Markov chain!

Generalized Borel-Cantelli Lemma

Another application of the almost sure convergence theorem is a generalization of the Borel-Cantelli lemmas. We first prove a dichotomy for the asymptotic behavior of martingales with L^1-bounded increments:

Theorem 4.7 (Asymptotics of martingales with L^1 bounded increments). *Suppose that (M_n) is a martingale, and there exists an integrable random variable Y such that*

$$|M_n - M_{n-1}| \leq Y \qquad \text{for any } n \in \mathbb{N}.$$

Then for P-almost every ω, the following dichotomy holds:

Either: *The limit $\lim\limits_{n\to\infty} M_n(\omega)$ exists in $\mathbb{R}$,*

or: $\limsup\limits_{n\to\infty} M_n(\omega) = +\infty$ *and* $\liminf\limits_{n\to\infty} M_n(\omega) = -\infty$.

The theorem and its proof are a generalization of Example 1 above.

Proof. For $a \in (-\infty, 0)$ let $T_a = \min\{n \geq 0 \,:\, M_n \geq a\}$. By the Optional Stopping Theorem, $(M_{T_a \wedge n})$ is a martingale. Moreover,

$$M_{T_a \wedge n} \geq \min(M_0, a - Y) \qquad \text{for any } n \geq 0,$$

and the right hand side is an integrable random variable. Therefore, (M_n) converges almost surely on $\{T_a = \infty\}$. Since this holds for every $a < 0$, we obtain almost sure convergence on the set

$$\{\liminf M_n > -\infty\} = \bigcup_{\substack{a<0 \\ a\in\mathbb{Q}}} \{T_a = \infty\}.$$

Similarly, almost sure convergence follows on the set $\{\limsup M_n < \infty\}$. □

Now let $(\mathcal{F}_n)_{n\geq 0}$ be an arbitrary filtration. As a consequence of Theorem 4.7 we obtain:

Corollary 4.8 (Generalized Borel-Cantelli Lemma). *If (A_n) is a sequence of events with $A_n \in \mathcal{F}_n$ for any n, then the equivalence*

$$\omega \in A_n \quad \text{infinitely often} \quad \iff \quad \sum_{n=1}^{\infty} P[A_n \mid \mathcal{F}_{n-1}](\omega) \;=\; \infty$$

holds for almost every $\omega \in \Omega$.

Proof. Let $S_n = \sum_{k=1}^{n} I_{A_k}$ and $T_n = \sum_{k=1}^{n} E[I_{A_k} \mid \mathcal{F}_{k-1}]$. Then S_n and T_n are almost surely increasing sequences. Let $S_\infty = \sup S_n$ and $T_\infty = \sup T_n$ denote the limits on $[0, \infty]$. The claim is that almost surely,

$$S_\infty \;=\; \infty \quad \iff \quad T_\infty \;=\; \infty. \tag{4.2.4}$$

To prove (4.2.4) we note that $S_n - T_n$ is a martingale with bounded increments. Therefore, almost surely, $S_n - T_n$ converges to a finite limit, or ($\limsup(S_n - T_n) = \infty$ and $\liminf(S_n - T_n) = -\infty$). In the first case, (4.2.4) holds. In the second case, $S_\infty = \infty$ and $T_\infty = \infty$, so (4.2.4) holds, too. □

The assertion of Corollary 4.8 generalizes both classical Borel-Cantelli Lemmas: If (A_n) is an arbitrary sequence of events in a probability space $(\Omega, \mathcal{A}, P)$ then we can consider the filtration $\mathcal{F}_n = \sigma(A_1, \ldots, A_n)$. By Corollary 4.8 we obtain:

1^{st} ***Borel-Cantelli Lemma:*** If $\sum P[A_n] < \infty$ then $\sum P[A_n \mid \mathcal{F}_{n-1}] < \infty$ almost surely, and therefore

$$P[A_n \text{ infinitely often}] \;=\; 0.$$

2^{nd} ***Borel-Cantelli Lemma:*** If $\sum P[A_n] = \infty$ and the A_n are independent then $\sum P[A_n \mid \mathcal{F}_{n-1}] \;=\; \sum P[A_n] \;=\; \infty$ almost surely, and therefore

$$P[A_n \text{ infinitely often}] \;=\; 1.$$

Upcrossing inequality and convergence theorem in continuous time

The upcrossing inequality and the supermartingale convergence theorem carry over immediately to the continuous time case if we assume right continuity (or left continuity) of the sample paths. Let $u \in (0, \infty]$, and let $(Z_s)_{s\in[0,u)}$ be a supermartingale in continuous time w.r.t. a filtration $(\mathcal{F}_s)$. We define the number of upcrossings of (Z_s) over an interval (a, b) before time t as the supremum of the number of upcrossings over all time discretizations $(Z_s)_{s\in\pi}$ where π is a partition of the interval $[0, t]$:

$$U_t^{(a,b)}[Z] := \sup_{\substack{\pi\subset[0,t] \\ \text{finite}}} U^{(a,b)}[(Z_s)_{s\in\pi}].$$

Note that if (Z_s) has right-continuous sample paths and (π_n) is a sequence of partitions of $[0, t]$ such that $0, t \in \pi_0$, $\pi_n \subset \pi_{n+1}$ and $\text{mesh}(\pi_n) \to 0$ then

$$U_t^{(a,b)}[Z] = \lim_{n\to\infty} U^{(a,b)}[(Z_s)_{s\in\pi_n}].$$

Theorem 4.9 (Supermatingale Convergence Theorem in continuous time). *Suppose that $(Z_s)_{s\in[0,u)}$ is a right continuous supermartingale.*

(1). Upcrossing inequality: For any $t \in [0, u)$ and $a < b$,

$$E[U_t^{(a,b)}] \le \frac{1}{b-a} E[(Z_t - a)^-].$$

(2). Convergence Theorem: If $\sup_{s\in[0,u)} E[Z_s^-] < \infty$, then the limit $Z_{u-} = \lim_{s\nearrow u} Z_s$ exists almost surely, and Z_{u-} is an integrable random variable.

Proof. (1). By the upcrossing inequality in discrete time,

$$E[U^{(a,b)}[(Z_s)_{s\in\pi_n}]] \le E[(Z_t - a)^-] \qquad \text{for any } n \in \mathbb{N},$$

where (π_n) is a sequence of partitions as above. The assertion now follows by the Monotone Convergence Theorem.

(2). The almost sure convergence can now be proven in the same way as in the discrete time case.

□

More generally than stated above, the upcrossing inequality also implies that for a right-continuous supermartingale $(Z_s)_{s\in[0,u)}$ all the left limits $\lim_{s\nearrow t} Z_s, t \in [0,u)$, exist *simultaneously* with probability one. Thus almost every sample path is *càdlàg* (continue à droite, limites a gauche, i.e., right continuous with left limits). By similar arguments, the existence of a modification with right continuous (and hence càdlàg) sample paths can be proven for *any* supermartingale (Z_s) provided the filtration is right continuous and complete, and $s \mapsto E[Z_s]$ is right continuous, cf. e.g. [XXXRevuz/Yor, Ch.II,§2].

4.3 Uniform integrability and $\mathbf{L^1}$ convergence

The Supermartingale Convergence Theorem shows that every supermartingale (Z_n) that is bounded in L^1 converges almost surely to an integrable limit Z_∞. However, L^1 convergence does not necessarily hold:

Example. (1). Suppose that $Z_n = \prod_{i=1}^n Y_i$ where the Y_i are i.i.d. with $E[Y_i] = 1$, $P[Y_i \neq 1] > 0$. Then, $Z_n \to 0$ almost surely, cf. Example 2 in Section 4.2. On the other hand, L^1 convergence does not hold as $E[Z_n] = 1$ for any n.

(2). Similarly, the exponential martingale $M_t = \exp(B_t - t/2)$ of a Brownian motion converges to 0 almost surely, but $E[M_t] = 1$ for any t.

L^1 convergence of martingales is of interest because it implies that a martingale sequence (M_n) can be extended to $n = \infty$, and the random variables M_n are given as conditional expectations of the limit M_∞. Therefore, we now prove a generalization of the Dominated Convergence Theorem that leads to a necessary and sufficient condition for L^1 convergence.

Uniform integrability

Let $(\Omega, \mathcal{A}, P)$ be a probability space. The key condition required to deduce L^1 convergence from convergence in probability is uniform integrability. To motivate the definition we first recall two characterizations of integrable random variables:

Lemma 4.10. *If $X : \Omega \to \mathbb{R}$ is an integrable random variable on $(\Omega, \mathcal{A}, P)$, then*

(1). $\lim\limits_{c\to\infty} E[|X|\,;\,|X| \geq c] = 0,$ *and*

(2). for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$E[|X|\,;\,A] \;<\; \varepsilon \quad \textit{for any } A \in \mathcal{A} \textit{ with } P[A] < \delta.$$

The second statement says that the positive measure

$$Q(A) \;=\; E[|X|\,;\,A], \qquad A \in \mathcal{A},$$

with relative density $|X|$ w.r.t. P is **absolutely continuous** w.r.t. P in the following sense: *For any $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$P[A] \;<\; \delta \quad \Rightarrow \quad Q(A) \;<\; \varepsilon.$$

Proof. (1). For an integrable random variable X the first assertion holds by the Monotone Convergence Theorem, since $|X| \cdot I_{\{|X|\geq c\}} \searrow 0$ as $c \nearrow \infty$.

(2). Let $\varepsilon > 0$. By (1),

$$\begin{aligned} E[|X|\,;\,A] \;&=\; E[|X|\,;\,A \cap \{|X| \geq c\}] + E[|X|\,;\,A \cap \{|X| < c\}] \\ &\leq\; E[|X|\,;\,|X| \geq c] + c \cdot P[A] \\ &<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon \end{aligned}$$

provided $c \in (0, \infty)$ is chosen appropriately and $P[A] < \varepsilon/2c$.

$\square$

Uniform integrability means that properties (1) and (2) hold uniformly for a family of random variables:

Definition (Uniform integrability). *A family $\{X_i : i \in I\}$ of random variables on $(\Omega, \mathcal{A}, P)$ is called* ***uniformly integrable*** *if and only if*

$$\sup_{i\in I} E[|X_i| \,;\, |X_i| \geq c] \quad \longrightarrow \quad 0 \qquad \text{as } c \to \infty.$$

Exercise (Equivalent characterization of uniform integrability). Prove that $\{X_i : i \in I\}$ is uniformly integrable if and only if $\sup E[|X_i| \,;\, A] < \infty$, and the measures $Q_i(A) = E[|X_i| \,;\, A]$ are **uniformly absolutely continuous**, i.e., for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$P[A] \,<\, \delta \quad \Rightarrow \quad \sup_{i\in I} E[|X_i| \,;\, A] \,<\, \varepsilon.$$

We will prove below that convergence in probability plus uniform integrability is equivalent to L^1 convergence. Before, we state two lemmas giving sufficient conditions for uniform integrability (and hence for L^1 convergence) that can often be verified in applications:

Lemma 4.11 (Sufficient conditions for uniform integrability). *A family $\{X_i : i \in I\}$ of random variables is uniformly integrable if one of the following conditions holds:*

(1). There exists an integrable random variable Y such that

$$|X_i| \,\leq\, Y \qquad \text{for any } i \in I.$$

(2). There exists a measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{x\to\infty} \frac{g(x)}{x} \,=\, \infty \qquad \text{and} \qquad \sup_{i\in I} E[g(|X_i|)] \,<\, \infty.$$

Proof. (1). If $|X_i| \leq Y$ then

$$\sup_{i\in I} E[|X_i| \,;\, |X_i| \geq c] \,\leq\, E[Y \,;\, Y \geq c].$$

The right hand side converges to 0 as $c \to \infty$ if Y is integrable.

(2). The second condition implies uniform integrability, because

$$\sup_{i\in I} E[|X_i|\,;\,|X_i|\geq c] \;\leq\; \sup_{y\geq c}\frac{y}{g(y)}\cdot\sup_{i\in I}E[g(|X_i|)].$$

□

The first condition in Lemma 4.11 is the classical assumption in the Dominated Convergence Theorem. The second condition holds in particular if

$$\sup_{i\in I} E[|X_i|^p] \;<\; \infty \qquad \text{for some } p>1 \qquad (\mathbf{L^p\ boundedness}),$$

or, if

$$\sup_{i\in I} E[|X_i|(\log|X_i|)^+] \;<\; \infty \qquad (\textbf{Entropy condition})$$

is satisfied. Boundedness in L^1, however, does not imply uniform integrability, cf. the examples at the beginning of this section.

The next observation is crucial for the application of uniform integrability to martingales:

Lemma 4.12 (**Conditional expectations are uniformly integrable**). *If X is an integrable random variable on $(\Omega,\mathcal{A},P)$ then the family*

$$\{E[X\mid\mathcal{F}]\;:\;\mathcal{F}\subseteq\mathcal{A}\quad \sigma\text{-algebra}\}$$

of all conditional expectations of X given sub-σ-algebras of $\mathcal{A}$ is uniformly integrable.

Proof. By Lemma 4.10, for any $\varepsilon>0$ there exists $\delta>0$ such that

$$\begin{aligned} E[|E[X\mid\mathcal{F}]|\,;\,|E[X\mid\mathcal{F}]|\geq c] &\leq E[E[|X|\mid\mathcal{F}]\,;\,|E[X\mid\mathcal{F}]|\geq c] \qquad (4.3.1)\\ &= E[|X|\,;\,|E[X\mid\mathcal{F}]|\geq c] \;<\; \varepsilon \end{aligned}$$

holds for $c>0$ with $P[|E[X\mid\mathcal{F}]|\geq c]<\delta$. Since

$$P[|E[X\mid\mathcal{F}]|\geq c] \;\leq\; \frac{1}{c}E[|E[X\mid\mathcal{F}]|] \;\leq\; \frac{1}{c}E[\,|X|\,],$$

(4.3.1) holds simultaneously for all σ-algebras $\mathcal{F}\subseteq\mathcal{A}$ if c is sufficiently large. □

Definitive version of Lebesgue's Dominated Convergence Theorem

Theorem 4.13. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of integrable random variables. Then (X_n) converges to a random variable X w.r.t. the L^1 norm if and only if X_n converges to X in probability and the family $\{X_n \,:\, n \in \mathbb{N}\}$ is uniformly integrable.*

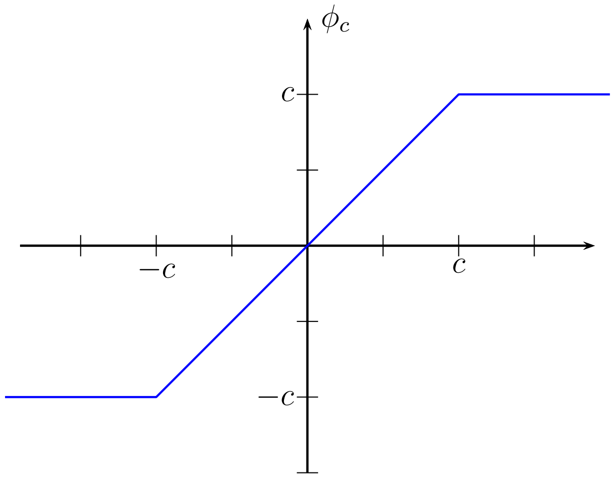
Proof. (1). We first prove the "if" part of the assertion under the additional assumption that the random variables $|X_n|$ are uniformly bounded by a finite constant c: For $\varepsilon > 0$,

$$\begin{aligned} E[\,|X_n - X|\,] \;&=\; E[\,|X_n - X|\,;\,|X_n - X| > \varepsilon] + E[\,|X_n - X|\,;\,|X_n - X| \leq \varepsilon] \\ &\leq\; 2c \cdot P[\,|X_n - X| > \varepsilon] \,+\, \varepsilon. \qquad (4.3.2) \end{aligned}$$

Here we have used that $|X_n| \leq c$ and hence $|X| \leq c$ with probability one, because a subsequence of (X_n) converges almost surely to X. For sufficiently large n, the right hand side of (4.3.2) is smaller than 2ε. Therefore, $E[\,|X_n - X|\,] \to 0$ as $n \to \infty$.

(2). To prove the "if" part under the uniform integrability condition, we consider the cut-off-functions

$$\phi_c(x) \;=\; (x \wedge c) \vee (-c)$$



For $c \in (0, \infty)$, the function $\phi_c : \mathbb{R} \to \mathbb{R}$ is a contraction. Therefore,

$$|\phi_c(X_n) - \phi_c(X)| \;\leq\; |X_n - X| \qquad \text{for any } n \in \mathbb{N}.$$

If $X_n \to X$ in probability then $\phi_c(X_n) \to \phi_c(X)$ in probability. Hence by (1),

$$E[\,|\phi_c(X_n) - \phi_c(X)|\,] \;\longrightarrow\; 0 \qquad \text{for any } c > 0. \tag{4.3.3}$$

We would like to conclude that $E[\,|X_n - X|\,] \to 0$ as well. Since (X_n) is uniformly integrable, and a subsequence converges to X almost surely, we have $E[\,|X|\,] \leq \liminf E[\,|X_n|\,] < \infty$ by Fatou's Lemma. We now estimate

$$\begin{aligned}
&E[\,|X_n - X|\,] \\
&\leq\; E[\,|X_n - \phi_c(X_n)|\,] + E[\,|\phi_c(X_n) - \phi_c(X)|\,] + E[\,|\phi_c(X) - X|\,] \\
&\leq\; E[\,|X_n|\,;\,|X_n| \geq c] + E[\,|\phi_c(X_n) - \phi_c(X)|\,] + E[\,|X|\,;\,|X| \geq c].
\end{aligned}$$

Let $\varepsilon > 0$ be given. Choosing c large enough, the first and the last summand on the right hand side are smaller than $\varepsilon/3$ for all n by uniform integrability of $\{X_n : n \in \mathbb{N}\}$ and integrability of X. Moreover, by (4.3.3), there exists $n_0(c)$ such that the middle term is smaller than $\varepsilon/3$ for $n \geq n_0(c)$. Hence $E[\,|X_n - X|\,] < \varepsilon$ for $n \geq n_0$, and thus $X_n \to X$ in L^1.

(3). Now suppose conversely that $X_n \to X$ in L^1. Then $X_n \to X$ in probability by Markov's inequality. To prove uniform integrability, we observe that

$$E[\,|X_n|\,;\,A] \;\leq\; E[\,|X|\,;\,A] + E[\,|X - X_n|\,] \qquad \text{for any } n \in \mathbb{N} \text{ and } A \in \mathcal{A}.$$

For $\varepsilon > 0$, there exist $n_0 \in \mathbb{N}$ and $\delta > 0$ such that

$$\begin{aligned}
E[\,|X - X_n|\,] &\;<\; \varepsilon/2 \qquad \text{for any } n > n_0, \text{ and} \\
E[\,|X|\,;\,A] &\;<\; \varepsilon/2 \qquad \text{whenever } P[A] < \delta,
\end{aligned}$$

cf. Lemma 4.10. Hence, if $P[A] < \delta$ then $\sup_{n \geq n_0} E[\,|X_n|\,;\,A] < \varepsilon$.
Moreover, again by Lemma 4.10, there exist $\delta_1, \ldots, \delta_{n_0} > 0$ such that for $n \leq n_0$,

$$E[\,|X_n|\,;\,A] \;<\; \varepsilon \qquad \text{if} \quad P[A] < \delta_n.$$

Choosing $\widetilde{\delta} = \min(\delta, \delta_1, \delta_2, \ldots, \delta_{n_0})$, we obtain

$$\sup_{n\in\mathbb{N}} E[\,|X_n|\,;\,A] < \varepsilon \qquad \text{whenever} \qquad P[A] \,<\, \widetilde{\delta}.$$

Therefore, $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable by the exercise below the definition of uniform integrability on page 140.

□

$\mathbf{L^1}$ convergence of martingales

If X is an integrable random variable and $(\mathcal{F}_n)$ is a filtration then $M_n = E[X \mid \mathcal{F}_n]$ is a martingale w.r.t. $(\mathcal{F}_n)$. The next result shows that an arbitrary martingale can be represented in this way if and only if it is uniformly integrable:

Theorem 4.14 ($\mathbf{L^1}$ Martingale Convergence Theorem). *Suppose that $(M_n)_{n\geq 0}$ is a martingale w.r.t. a filtration $(\mathcal{F}_n)$. Then the following statements are equivalent:*

(1). $\{M_n \,:\, n \geq 0\}$ is uniformly integrable.

(2). The sequence (M_n) converges w.r.t. the L^1 norm.

(3). There exists an integrable random variable X such that

$$M_n \;=\; E[X \mid \mathcal{F}_n] \qquad \textit{for any } n \geq 0.$$

Proof.

(3) $\Rightarrow$ (1) holds by Lemma 4.12.

(1) $\Rightarrow$ (2): If the sequence (M_n) is uniformly integrable then it is bounded in L^1 because

$$\sup_n E[\,|M_n|\,] \;\leq\; \sup_n E[\,|M_n|\,;\,|M_n| \geq c] + c \;\leq\; 1 + c$$

for $c \in (0, \infty)$ sufficiently large. Therefore, the limit $M_\infty = \lim M_n$ exists almost surely and in probability by the almost sure convergence theorem. Uniform integrability then implies

$$M_n \rightarrow M_\infty \quad \text{in } L^1$$

by Theorem 4.13.

(2) ⇒ (3): If M_n converges to a limit M_∞ in L^1 then

$$M_n = E[M_\infty \mid \mathcal{F}_n] \qquad \text{for any } n \geq 0.$$

Indeed, M_n is a version of the conditional expectation since it is $\mathcal{F}_n$-measurable and

$$E[M_\infty \,;\, A] \;=\; \lim_{k\to\infty} E[M_k \,;\, A] \;=\; E[M_n \,;\, A] \qquad \text{for any } A \in \mathcal{F}_n \tag{4.3.4}$$

by the martingale property.

□

A first consequence of the L^1 convergence theorem is a limit theorem for conditional expectations:

Corollary 4.15. *If X is an integrable random variable and $(\mathcal{F}_n)$ is a filtration then*

$$E[X \mid \mathcal{F}_n] \;\rightarrow\; E[X \mid \mathcal{F}_\infty] \qquad \textit{almost surely and in } L^1,$$

where $\mathcal{F}_\infty := \sigma(\bigcup \mathcal{F}_n)$.

Proof. Let $M_n := E[X \mid \mathcal{F}_n]$. By the almost sure and the L^1 martingale convergence theorem, the limit $M_\infty = \lim M_n$ exists almost surely and in L^1. To obtain a measurable function that is defined everywhere, we set $M_\infty := \limsup M_n$. It remains to verify, that M_∞ is a version of the conditional expectation $E[X \mid \mathcal{F}_\infty]$. Clearly, M_∞ is measurable w.r.t. $\mathcal{F}_\infty$. Moreover, for $n \geq 0$ and $A \in \mathcal{F}_n$,

$$E[M_\infty \,;\, A] \;=\; E[M_n \,;\, A] \;=\; E[X \,;\, A]$$

by (4.3.4). Since $\bigcup \mathcal{F}_n$ is stable under finite intersections,

$$E[M_\infty \,;\, A] \;=\; E[X \,;\, A]$$

holds for all $A \in \sigma(\bigcup \mathcal{F}_n)$ as well. □

Example (Existence of conditional expectations). The common existence proof for conditional expectations relies either on the Radon-Nikodym Theorem or on the existence of orthogonal projections onto closed subspaces of the Hilbert space L^2. Martingale convergence can be used to give an alternative existence proof. Suppose that X is an integrable random variable on a probability space $(\Omega, \mathcal{A}, P)$ and $\mathcal{F}$ is a **separable** sub-σ-algebra of $\mathcal{A}$, i.e., there exists a countable collection $(A_i)_{i\in\mathbb{N}}$ of events $A_i \in \mathcal{A}$ such that $\mathcal{F} = \sigma(A_i : i \in \mathbb{N})$. Let

$$\mathcal{F}_n \;=\; \sigma(A_1, \ldots, A_n), \qquad n \geq 0.$$

Note that for each $n \geq 0$, there exist finitely many atoms $B_1, \ldots, B_k \in \mathcal{A}$ (i.e. disjoint events with $\bigcup B_i = \Omega$) such that $\mathcal{F}_n = \sigma(B_1, \ldots, B_k)$. Therefore, the conditional expectation given $\mathcal{F}_n$ can be defined in an elementary way:

$$E[X \mid \mathcal{F}_n] \;:=\; \sum_{i\,:\,P[B_i]\neq 0} E[X \mid B_i] \cdot I_{B_i}.$$

Moreover, by Corollary 4.15, the limit $M_\infty = \lim E[X \mid \mathcal{F}_n]$ exists almost surely and in L^1, and M_∞ is a version of the conditional expectation $E[X \mid \mathcal{F}]$.

You might (and should) object that the proofs of the martingale convergence theorems require the existence of conditional expectations. Although this is true, it is possible to state the necessary results by using only elementary conditional expectations, and thus to obtain a more constructive proof for existence of conditional expectations given separable σ-algebras.

Another immediate consequence of Corollary 4.15 is an extension of Kolmogorov's 0-1 law:

Corollary 4.16 (0-1 Law of P.Lévy). *If $(\mathcal{F}_n)$ is a filtration on $(\Omega, \mathcal{A}, P)$ then for any event $A \in \sigma(\bigcup \mathcal{F}_n)$,*

$$P[A \mid \mathcal{F}_n] \longrightarrow I_A \qquad P\text{-almost surely.} \tag{4.3.5}$$

Example (Kolmogorov's 0-1 Law). Suppose that $\mathcal{F}_n = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_n)$ with independent σ-algebras $\mathcal{A}_i \subseteq \mathcal{A}$. If A is a **tail event**, i.e., A is in $\sigma(\mathcal{A}_{n+1}, \mathcal{A}_{n+2}, \dots)$ for every $n \in \mathbb{N}$, then A is independent of $\mathcal{F}_n$ for any n. Therefore, the corollary implies that $P[A] = I_A$ P-almost surely, i.e.,

$$P[A] \in \{0, 1\} \qquad \text{for any tail event } A.$$

The L^1 Martingale Convergence Theorem also implies that any martingale that is L^p bounded for some $p \in (1, \infty)$ converges in L^p:

Exercise ($\mathbf{L^p}$ Martingale Convergence Theorem). Let (M_n) be an $(\mathcal{F}_n)$ martingale with $\sup E[\,|M_n|^p\,] < \infty$ for some $p \in (1, \infty)$.

(1). Prove that (M_n) converges almost surely and in L^1, and $M_n = E[M_\infty \mid \mathcal{F}_n]$ for any $n \geq 0$.

(2). Conclude that $|M_n - M_\infty|^p$ is uniformly integrable, and $M_n \to M_\infty$ in L^p.

Note that uniform integrability of $|M_n|^p$ holds automatically and has not to be assumed !

Backward Martingale Convergence

We finally remark that Doob's upcrossing inequality can also be used to prove that the conditional expectations $E[X \mid \mathcal{F}_n]$ of an integrable random variable given a *decreasing* sequence $(\mathcal{F}_n)$ of σ-algebras converge almost surely to $E[X \mid \bigcap \mathcal{F}_n]$. For the proof one considers the martingale $M_{-n} = E[X \mid \mathcal{F}_n]$ indexed by the negative integers:

Exercise (Backward Martingale Convergence Theorem and LLN). Let $(\mathcal{F}_n)_{n \geq 0}$ be a *decreasing* sequence of sub-σ-algebras on a probability space $(\Omega, \mathcal{A}, P)$.

(1). Prove that for any random variable $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, the limit $M_{-\infty}$ of the sequence $M_{-n} := E[X \mid \mathcal{F}_n]$ as $n \to -\infty$ exists almost surely and in L^1, and

$$M_{-\infty} = E[X \mid \bigcap \mathcal{F}_n] \qquad \text{almost surely.}$$

(2). Now let (X_n) be a sequence of i.i.d. random variables in $\mathcal{L}^1(\Omega, \mathcal{A}, P)$, and let $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots)$ where $S_n = X_1 + \ldots + X_n$. Prove that

$$E[X_1 \mid \mathcal{F}_n] = \frac{S_n}{n},$$

and conclude that the strong Law of Large Numbers holds:

$$\frac{S_n}{n} \longrightarrow E[X_1] \qquad \text{almost surely.}$$

Chapter 5

Stochastic Integration w.r.t. Continuous Martingales

Suppose that we are interested in a continuous-time scaling limit of a stochastic dynamics of type $X_0^{(h)} = x_0$,

$$X_{k+1}^{(h)} - X_k^{(h)} \;=\; \sigma(X_k^{(h)}) \cdot \sqrt{h} \cdot \eta_{k+1}, \qquad k = 0, 1, 2, \ldots, \tag{5.0.1}$$

with i.i.d. random variables $\eta_i \in \mathcal{L}^2$ such that $E[\eta_i] = 0$ and $\mathrm{Var}[\eta_i] = 1$, a continuous function $\sigma : \mathbb{R} \to \mathbb{R}$, and a scale factor $h > 0$. Equivalently,

$$X_n^{(h)} \;=\; X_0^{(h)} + \sqrt{h} \cdot \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \cdot \eta_{k+1}, \qquad n = 0, 1, 2, \ldots. \tag{5.0.2}$$

If σ is constant then as $h \searrow 0$, the rescaled process $(X_{\lfloor t/h \rfloor}^{(h)})_{t \geq 0}$ converges in distribution to $(\sigma \cdot B_t)$ where (B_t) is a Brownian motion. We are interested in the scaling limit for general σ. One can prove that the rescaled process again converges in distribution, and the limit process is a solution of a stochastic integral equation

$$X_t \;=\; X_0 + \int_0^t \sigma(X_s)\, dB_s, \qquad t \geq 0. \tag{5.0.3}$$

Here the integral is an Itô stochastic integral w.r.t. a Brownian motion (B_t). Usually the equation (5.0.3) is written briefly as

$$dX_t \;=\; \sigma(X_t)\, dB_t, \tag{5.0.4}$$

and interpreted as a stochastic differential equation. Stochastic differential equations occur more generally when considering scaling limits of appropriately rescaled Markov chains on $\mathbb{R}^d$ with finite second moments. The goal of this section is to give a meaning to the stochastic integral, and hence to the equations (5.0.3), (5.0.4) respectively.

Example (Stock prices, geometric Brownian motion). A simple discrete time model for stock prices is given by

$$X_{k+1} - X_k \;=\; X_k \cdot \eta_{k+1}, \qquad \eta_i \text{ i.i.d.}$$

To set up a corresponding continuous time model we consider the rescaled equation (5.0.1) as $h \searrow 0$. The limit in distribution is a solution of a stochastic differential equation

$$dX_t \;=\; X_t \, dB_t \tag{5.0.5}$$

w.r.t. a Brownian motion (B_t). Although with probability one, the sample paths of Brownian motion are nowhere differentiable, we can give a meaning to this equation by rewriting it in the form (5.0.3) with an Itô stochastic integral.
A naive guess would be that the solution of (5.0.5) with initial condition $X_0 = 1$ is $X_t = \exp B_t$. However, more careful considerations show that this can not be true! In fact, the discrete time approximations satisfy

$$X_{k+1}^{(h)} \;=\; (1 + \sqrt{h}\eta_{k+1}) \cdot X_k^{(h)} \qquad \text{for } k \geq 0.$$

Hence $(X_k^{(h)})$ is a product martingale:

$$X_n^{(h)} \;=\; \prod_{k=1}^{n} (1 + \sqrt{h}\eta_k) \qquad \text{for any } n \geq 0.$$

In particular, $E[X_n^{(h)}] = 1$. We would expect similar properties for the scaling limit (X_t), but $\exp B_t$ is not a martingale and $E[\exp(B_t)] = \exp(t/2)$.
It turns out that in fact, the unique solution of (5.0.5) with $X_0 = 1$ is not $\exp(B_t)$ but the exponential martingale

$$X_t \;=\; \exp(B_t - t/2),$$

which is also called a geometric Brownian motion. The reason is that the irregularity of Brownian paths enforces a correction term in the chain rule for stochastic differentials leading to Itô's famous formula, which is the fundament of stochastic calculus.

5.1 Defining stochastic integrals: A first attempt

Let us first fix some notation that will be used constantly below: By a **partition** π of $\mathbb{R}_+$ we mean an increasing sequence $0 = t_0 < t_1 < t_2 < \ldots$ such that $\sup t_n = \infty$. The *mesh size* of the partition is

$$\mathrm{mesh}(\pi) \ = \ \sup\{|t_i - t_{i-1}| \ : \ i \in \mathbb{N}\}.$$

We are interested in defining integrals of type

$$I_t \ = \ \int_0^t H_s \, dX_s, \qquad t \geq 0, \tag{5.1.1}$$

for continuous functions and, respectively, continuous adapted processes (H_s) and (X_s).

For a given $t \geq 0$ and a given partition π of $\mathbb{R}_+$, we define the increments of (X_s) up to time t by

$$\delta X_s \ := \ X_{s' \wedge t} - X_{s \wedge t} \qquad \text{for any } s \in \pi,$$

where $s' := \min\{u \in \pi \ : \ u > s\}$ denotes the next partition point after s. Note that the increments δX_s vanish for $s \geq t$. In particular, only finitely many of the increments are not equal to zero. A nearby approach for defining the integral I_t in (5.1.1) would be Riemann sum approximations:

Riemann sum approximations

There are various possibilities to define approximating Riemann sums w.r.t. a given sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, for example:

Variant 1 (non-anticipative): $I_t^n \ = \ \sum\limits_{s \in \pi_n} H_s \delta X_s$,

Variant 2 (anticipative): $\hat{I}_t^n \ = \ \sum\limits_{s \in \pi_n} H_{s'} \delta X_s$,

Variant 3 (anticipative): $\overset{\circ}{I}{}_t^n \ = \ \sum\limits_{s \in \pi_n} \frac{1}{2}(H_s + H_{s'}) \delta X_s$.

Note that for finite t, in each of the sums, only finitely many summands do not vanish. For example,

$$I_t^n \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} H_s \delta X_s \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} H_s \cdot (X_{s'\wedge t} - X_s).$$

Now let us consider at first the case where $H_s = X_s$ and $t = 1$, i.e., we would like to define the integral $I = \int\limits_0^1 X_s\, dX_s$. Suppose first that $X : [0,1] \to \mathbb{R}$ is a continuous function of finite variation, i.e.,

$$V^{(1)}(X) \;=\; \sup\left\{ \sum_{s\in\pi} |\delta X_s| \;:\; \pi \text{ partition of } \mathbb{R}_+ \right\} < \infty.$$

Then for $H = X$ and $t = 1$ all the approximations above converge to the same limit as $n \to \infty$. For example,

$$\|\hat{I}_1^n - I_1^n\| \;=\; \sum_{s\in\pi_n} (\delta X_s)^2 \;\le\; V^{(1)}(X) \cdot \sup_{s\in\pi_n} |\delta X_s|,$$

and the right-hand side converges to 0 by uniform continuity of X on $[0,1]$. In this case the limit of the Riemann sums is a Riemann-Stieltjes integral

$$\lim_{n\to\infty} I_1^n \;=\; \lim_{n\to\infty} \hat{I}_1^n \;=\; \int\limits_0^1 X_s\, dX_s,$$

which is well-defined whenever the integrand is continuous and the integrator is of finite variation or conversely. The sample paths of Brownian motion, however, are almost surely not of finite variation. Therefore, the reasoning above does not apply, and in fact if $X_t = B_t$ is a one-dimensional Brownian motion and $H_t = X_t$ then

$$E[\,|\hat{I}_1^n - I_1^n|\,] \;=\; \sum_{s\in\pi_n} E[(\delta B_s)^2] \;=\; \sum_{s\in\pi_n} \delta s \;=\; 1,$$

i.e., the L^1-limits of the random sequence (I_1^n) and $(\hat{I}_1^n)$ are different if they exist. Below we will see that indeed the limits of the sequences (I_1^n), $(\hat{I}_1^n)$ and $(\overset{\circ}{I}{}_1^n)$ do exist in L^2, and all the limits are different. The limit of the non-anticipative Riemann sums I_1^n is the *Itô stochastic integral* $\int_0^1 B_s\, dB_s$, the limit of $(\hat{I}_1^n)$ is the *backward Itô integral* $\int_0^1 B_s\, \hat{d}B_s$, and the limit of I_n° is the *Stratonovich integral* $\int_0^1 B_s \circ dB_s$. All three notions of stochastic integrals are relevant. The most important one is the Itô integral because the non-anticipating Riemann sum approximations imply that the Itô integral $\int_0^t H_s \, dB_s$ is a continuous time martingale transform of Brownian motion if the process (H_s) is adapted.

Itô integrals for continuous bounded integrands

We now give a first existence proof for Itô integrals w.r.t. Brownian motion. We start with a provisional definition that will be made more precise later:

Preliminary Definition. *For continuous functions or continuous stochastic processes (H_s) and (X_s) and a given sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, the* ***Itô integral of*** H **w.r.t.** X *is defined by*

$$\int_0^t H_s \, dX_s \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} H_s \delta X_s$$

whenever the limit exists in a sense to be specified.

Note that the definition is vague since the mode of convergence is not specified. Moreover, the Itô integral might depend on the sequence (π_n). In the following sections we will see which kind of convergence holds in different circumstances, and in which sense the limit is independent of (π_n).

To get started let us consider the convergence of Riemann sum approximations for the Itô integrals $\int_0^t H_s \, dB_s$ of a bounded continuous $(\mathcal{F}_s)$ adapted process $(H_s)_{s\geq 0}$ w.r.t. an $(\mathcal{F}_s)$ Brownian motion (B_s). Let (π_n) be a fixed sequence of partitions with $\pi_n \subseteq \pi_{n+1}$ and $\mathrm{mesh}(\pi_n) \to 0$. Then for the Riemann-Itô sums

$$I_t^n \;=\; \sum_{s\in\pi_n} H_s \, \delta B_s \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} H_s (B_{s'\wedge t} - B_s)$$

we have

$$I_t^n - I_t^m \;=\; \sum_{\substack{s\in\pi_n \\ s<t}} (H_s - H_{\lfloor s \rfloor_m}) \, \delta B_s \qquad \text{for any } m \leq n,$$

where $\lfloor s \rfloor_m = \max\{r \in \pi_m : r \leq s\}$ denotes the next partition point in π_m below s. Since Brownian motion is a martingale, we have $E[\delta B_s \mid \mathcal{F}_s] = 0$ for any $s \in \pi_n$. Moreover, $E[(\delta B_s)^2 \mid \mathcal{F}_s] = \delta s$. Therefore, we obtain by conditioning on $\mathcal{F}_s, \mathcal{F}_r$ respectively:

$$\begin{aligned} E[(I_t^n - I_t^m)^2] &= \sum_{\substack{s\in\pi_n \\ s<t}} \sum_{\substack{r\in\pi_n \\ r<t}} E[(H_s - H_{\lfloor s \rfloor_m})(H_r - H_{\lfloor r \rfloor_m})\delta B_s \delta B_r] \\ &= \sum_{\substack{s\in\pi_n \\ s<t}} E[(H_s - H_{\lfloor s \rfloor_m})^2 \delta s] \;\leq\; E[V_m] \cdot \sum_{\substack{s\in\pi_n \\ s<t}} \delta s \;=\; E[V_m] \cdot t, \end{aligned}$$

where

$$V_m := \sup_{|s-r| \leq \mathrm{mesh}(\pi_m)} (H_s - H_r)^2 \quad \longrightarrow \quad 0 \qquad \text{as } m \to \infty$$

by uniform continuity of (H_s) on $[0, t]$. Since H is bounded, $E[V_m] \to 0$ as $m \to \infty$, and hence (I_t^n) is a Cauchy sequence in $L^2(\Omega, \mathcal{A}, P)$ for any given $t \geq 0$. Thus we obtain:

Theorem 5.1 (Itô integrals for bounded continuous integrands, Variant 1). *Suppose that $(H_s)_{s\geq 0}$ is a bounded continuous $(\mathcal{F}_s)$ adapted process, and $(B_s)_{s\geq 0}$ is an $(\mathcal{F}_s)$ Brownian motion. Then for any fixed t ≥ 0, the Itô integral*

$$\int_0^t H_s \, dB_s = \lim_{n\to\infty} I_t^n \tag{5.1.2}$$

exists as a limit in $L^2(\Omega, \mathcal{A}, P)$. Moreover, the limit does not depend on the choice of a sequence of partitions (π_n) with $\mathrm{mesh}\,(\pi_n) \to 0$.

Proof. An analogue argument as above shows that for any partitions π and $\widetilde{\pi}$ such that $\pi \supseteq \widetilde{\pi}$, the L^2 distance of the corresponding Riemann sum approximations I_t^π and $I_t^{\widetilde{\pi}}$ is bounded by a constant $C(\mathrm{mesh}(\widetilde{\pi}))$ that only depends on the maximal mesh size of the two partitions. Moreover, the constant goes to 0 as the mesh sizes go to 0. By choosing a joint refinement and applying the triangle inequality, we see that

$$\|I_t^\pi - I_t^{\widetilde{\pi}}\|_{L^2(P)} \;\leq\; 2\,C(\Delta)$$

holds for arbitrary partitions $\pi, \widetilde{\pi}$ such that $\max(\mathrm{mesh}(\pi)), \mathrm{mesh}(\widetilde{\pi})) \leq \Delta$. The assertion now follows by completeness of $L^2(P)$. □

The definition of the Itô integral suggested by Theorem 5.1 has two obvious drawbacks:

Drawback 1: The integral $\int_0^t H_s dB_s$ is only defined as an equivalence class in $L^2(\Omega, \mathcal{A}, P)$, i.e., uniquely up to modification on P-measure zero sets. In particular, we do not have a *pathwise definition* of $\int_0^t H_s(\omega)\, dB_s(\omega)$ for a given Brownian sample path $s \mapsto B_s(\omega)$.

Drawback 2: Even worse, the construction above works only for a fixed integration interval $[0, t]$. The exceptional sets may depend on t and therefore, the process $t \mapsto \int_0^t H_s\, dB_s$ does not have a meaning yet. In particular, we do not know yet if there exists a version of this process that is almost surely continuous.

The first drawback is essential: In certain cases it is indeed possible to define stochastic integrals pathwise, cf. Chapter 6 below. In general, however, pathwise stochastic integrals cannot be defined. The extra impact needed is the Lévy area process, cf. the rough paths theory developed by T. Lyons and others [XXXLyons, Friz and Victoir, Friz and Hairer].

Fortunately, the second drawback can be overcome easily. By extending the Itô isometry to an isometry into the space M_c^2 of continuous L^2 bounded martingales, we can construct the complete process $t \mapsto \int_0^t H_s\, dB_s$ simultaneously as a continuous martingale. The key observation is that by the maximal inequality, continuous L^2 bounded martingales can be controlled uniformly in t by the L^2 norm of their final value.

The Hilbert space M_c^2

Fix $u \in (0, \infty]$ and suppose that for $t \in [0, u]$, (I_t^n) is a sequence of Riemann sum approximations for $\int_0^t H_s\, dB_s$ as considered above. It is not difficult to check that for each fixed $n \in \mathbb{N}$, the stochastic process $t \mapsto I_t^n$ is a continuous martingale. Our aim is to prove convergence of these continuous martingales to a further continuous martingale $I_t = \int_0^t H_s\, dB_s$. Since the convergence holds only almost surely, the limit process will not necessarily be $(\mathcal{F}_t)$ adapted. To ensure adaptedness, we have to consider the **completed filtration**

$$\mathcal{F}_t^P \;=\; \{A \in \mathcal{A} : P[A \bigtriangleup B] = 0 \text{ for some } B \in \mathcal{F}_t\}, \qquad t \geq 0,$$

where $A \bigtriangleup B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference of the sets A and B. Note that the conditional expectations given $\mathcal{F}_t$ and $\mathcal{F}_t^P$ agree P-almost surely. Hence, if (B_t) is a Brownian motion resp. a martingale w.r.t. the filtration $(\mathcal{F}_t)$ then it is also a Brownian motion or a martingale w.r.t. $(\mathcal{F}_t^P)$.

Let $\mathcal{M}^2([0,u])$ denote the space of all L^2-bounded $(\mathcal{F}_t^P)$ martingales $(M_t)_{0\leq t\leq u}$ on $(\Omega, \mathcal{A}, P)$. By $\mathcal{M}_c^2([0,u])$ and $\mathcal{M}_d^2([0,u])$ we denote the subspaces consisting of all continuous (respectively right continuous) martingales $M \in \mathcal{M}^2([0,u])$. Recall that by the L^2 martingale convergence theorem, any (right) continuous L^2-bounded martingale (M_t) defined for $t \in [0,u)$ can be extended to a (right) continuous martingale in $\mathcal{M}^2([0,u])$.

Two martingales $M, \widetilde{M} \in \mathcal{M}^2([0,u])$ are called **modifications** of each other if

$$P[M_t = \widetilde{M}_t] \;=\; 1 \qquad \text{for any } t \in [0,u].$$

If the martingales are right-continuous then two modifications agree almost surely, i.e.,

$$P[M_t = \widetilde{M}_t \; \forall t \in [0,u]] \;=\; 1.$$

In order to obtain norms and not just semi-norms, we consider the spaces

$$M^2([0,u]) \;:=\; \mathcal{M}^2([0,u])/\sim \qquad \text{and} \qquad M_c^2([0,u]) \;:=\; \mathcal{M}_c^2([0,u])/\sim$$

of equivalence classes of martingales that are modifications of each other. We will frequently identify equivalence classes and their representatives.

We endow the space $M^2([0,u])$ with the inner product

$$(M,N)_{M^2([0,u])} \;=\; (M_u, N_u)_{L^2} \;=\; E[M_u N_u].$$

As the process (M_t^2) is a submartingale for any $M \in M^2([0,u])$, the norm corresponding to the inner product is given by

$$\|M\|^2_{M^2([0,u])} \;=\; E[M_u^2] \;=\; \sup_{0\leq t\leq u} E[M_t^2].$$

Moreover, if (M_t) is right continuous then by **Doob's $\mathbf{L^2}$-maximal inequality**,

$$\left\| \sup_{0\le t\le u} |M_t| \right\|_{L^2(\Omega,\mathcal{A},P)} \;\le\; 2\cdot \sup_{0\le t\le u} \|M_t\|_{L^2(\Omega,\mathcal{A},P)} \;=\; 2\|M\|_{M^2([0,u])}. \tag{5.1.3}$$

This crucial estimate shows that on the subspaces M_c^2 and M_d^2, the M^2 norm is equivalent to the L^2 norm of the supremum of the martingale. Therefore, *the M^2 norm can be used to control (right) continuous martingales uniformly in t!*

Lemma 5.2 (Completeness). *(1). The space $M^2([0,u])$ is a Hilbert space, and the linear map $M \mapsto M_u$ from $M^2([0,u])$ to $L^2(\Omega,\mathcal{F}_u,P)$ is onto and isometric.*

(2). The spaces $M_c^2([0,u])$ and $M_d^2([0,u])$ are closed subspaces of $M^2([0,u])$, i.e., if (M^n) is a Cauchy sequence in $M_c^2([0,u])$, or in $M_d^2([0,u])$ respectively, then there exists a (right) continuous martingale $M \in M^2([0,u])$ such that

$$\sup_{t\in[0,u]} |M_t^n - M_t| \;\longrightarrow\; 0 \qquad \textit{in } L^2(\Omega,\mathcal{A},P).$$

Proof. (1). By definition of the inner product on $M^2([0,u])$, the map $M \mapsto M_u$ is an isometry. Moreover, for $X \in L^2(\Omega,\mathcal{F}_u,P)$, the process $M_t = E[X \mid \mathcal{F}_u]$ is in $M^2([0,u])$ with $M_u = X$. Hence, the range of the isometry is the whole space $L^2(\Omega,\mathcal{F}_u,P)$. Since $L^2(\Omega,\mathcal{F}_u,P)$ is complete w.r.t. the L^2 norm, the space $M^2([0,u])$ is complete w.r.t. the M^2 norm.

(2). If (M^n) is a Cauchy sequence in $M_c^2([0,u])$ or in $M_d^2([0,u])$ respectively, then by (5.1.3),

$$\|M^n - M^m\|_{\sup} \;:=\; \sup_{0\le t\le u} |M_t^n - M_t^m| \;\longrightarrow\; 0 \qquad \text{in } L^2(\Omega,\mathcal{A},P).$$

In particular, we can choose a subsequence (M^{n_k}) such that

$$P[\, \|M^{n_{k+1}} - M^{n_k}\|_{\sup} \ge 2^{-k} \,] \;\le\; 2^{-k} \qquad \text{for all } k \in \mathbb{N}.$$

Hence, by the Borel-Cantelli Lemma,

$$P[\, \|M^{n_{k+1}} - M^{n_k}\|_{\sup} < 2^{-k} \quad \text{eventually} \,] \;=\; 1,$$

and therefore $M_t^{n_k}$ converges almost surely uniformly in t as $k \to \infty$. The limit of the sequence (M^n) in $M^2([0,u])$ exists by (1), and the process M defined by

$$M_t \;:=\; \begin{cases} \lim M_t^{n_k} & \text{if } (M^{n_k}) \text{ converges uniformly,} \\ 0 & \text{otherwise,} \end{cases} \tag{5.1.4}$$

is a continuous (respectively right continuous) representative of the limit. Indeed, by Fatou's Lemma,

$$\begin{aligned} \|M^{n_k} - M\|^2_{M^2([0,u])} &\leq E[\,\|M^{n_k} - M\|^2_{\sup}\,] \;=\; E[\lim_{l\to\infty} \|M^{n_k} - M^{n_l}\|^2_{\sup}\,] \\ &\leq \liminf_{l\to\infty} E[\,\|M^{n_k} - M^{n_l}\|^2_{\sup}\,], \end{aligned}$$

and the right hand side converges to 0 as $k \to \infty$. Finally, M is a martingale w.r.t. $(\mathcal{F}_t^P)$, and hence an element in $M_c^2([0,u])$ or in $M_d^2([0,u])$ respectively.

□

Remark. We point out that the (right) continuous representative (M_t) defined by (5.1.4) is a martingale w.r.t. the complete filtration $(\mathcal{F}_t^P)$, but it is not necessarily adapted w.r.t. $(\mathcal{F}_t)$.

Definition of Itô integral in $\mathrm{M_c^2}$

Let $u \in \mathbb{R}^+$. For any bounded continuous $(\mathcal{F}_t)$ adapted process (H_t) and any sequence (π_n) of partitions of $\mathbb{R}_+$, the processes

$$I_t^n \;=\; \sum_{s\in\pi_n} H_s\,(B_{s'\wedge t} - B_{s\wedge t}), \qquad t \in [0,u],$$

are continuous L^2 bounded martingales on $[0,u]$. We can therefore restate Theorem 5.1 in the following way:

Corollary 5.3 (Itô integrals for bounded continuous integrands, Variant 2). *Suppose that $(H_s)_{s\geq 0}$ is a bounded continuous $(\mathcal{F}_s)$ adapted process. Then for any fixed $u \geq 0$, the Itô integral*

$$\int_0^\bullet H_s\,dB_s \;=\; \lim_{n\to\infty} (I_t^n)_{t\in[0,u]} \tag{5.1.5}$$

exists as a limit in $M_c^2([0,u])$*. Moreover, the limit does not depend on the choice of a sequence of partitions* (π_n) *with* $\mathrm{mesh}\,(\pi_n) \to 0$*.*

Proof. The assertion is an immediate consequence of the definition of the M^2 norm, Theorem 5.1 and Lemma 5.2. □

Similar arguments as above apply if Brownian motion is replaced by a bounded martingale with continuous sample paths. In the rest of this chapter we will work out the construction of the Itô integral w.r.t. Brownian motion and more general continuous martingales more systematically and for a broader class of integrands.

5.2 Itô's isometry

Let $(M_t)_{t\geq 0}$ be a continuous (or, more generally, right continuous) martingale w.r.t. a filtration $(\mathcal{F}_t)$ on a probability space $(\Omega, \mathcal{A}, P)$. We now develop a more systematic approach for defining stochastic integrals $\int_0^t H_s\, dM_s$ of adapted processes (H_t) w.r.t. (M_t).

Predictable step functions

In a first step, we define the integrals for predictable step functions (H_t) of type

$$H_t(\omega) \;=\; \sum_{i=0}^{n-1} A_i(\omega) I_{(t_i,t_{i+1}]}(t)$$

with $n \in \mathbb{N}, 0 \leq t_0 < t_1 < t_2 < \ldots < t_n$, and bounded $\mathcal{F}_{t_i}$-measurable random variables $A_i, i = 0, 1, \ldots, n-1$. Let $\mathscr{E}$ denote the vector space consisting of all stochastic processes of this form.

Definition (Itô integral for predictable step functions). *For stochastic processes $H \in \mathscr{E}$ and $t \geq 0$ we define*

$$\int_0^t H_s\, dM_s \;:=\; \sum_{i=0}^{n-1} A_i \cdot (M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) \;=\; \sum_{i\,:\,t_i<t} A_i \cdot (M_{t_{i+1}\wedge t} - M_{t_i}).$$

The stochastic process $H_\bullet M$ given by

$$(H_\bullet M)_t \;:=\; \int_0^t H_s\, dM_s \qquad \text{for } t \in [0,\infty]$$

is called the ***Itô integral*** *of H w.r.t. M.*

Note that the map $(H, M) \mapsto H_\bullet M$ is bilinear. The process $H_\bullet M$ is a continuous time **martingale transform** of M w.r.t. H. It models for example the net gain up to time t if we hold A_i units of an asset with price process (M_t) during each of the time intervals $(t_i, t_{i+1}]$.

Lemma 5.4. *For any $H \in \mathscr{E}$, the process $H_\bullet M$ is a continuous $(\mathcal{F}_t)$ martingale up to time $t = \infty$.*

Similarly to the discrete time case, the fact that A_i is predictable, i.e., $\mathcal{F}_{t_i}$-measurable, is essential for the martingale property:

Proof. By definition, $H_\bullet M$ is continuous and $(\mathcal{F}_t)$ adapted. It remains to verify that

$$E[(H_\bullet M)_t \mid \mathcal{F}_s] \;=\; (H_\bullet M)_s \qquad \text{for any } 0 \leq s \leq t. \tag{5.2.1}$$

We do this in three steps:

(1). At first we note that (5.2.1) holds for $s, t \in \{t_0, t_1, \ldots, t_n\}$. Indeed, since A_i is $\mathcal{F}_{t_i}$-measurable, the process

$$(H_\bullet M)_{t_j} \;=\; \sum_{i=0}^{j-1} A_i \cdot (M_{t_{i+1}} - M_{t_i}), \qquad j = 0, 1, \ldots, n,$$

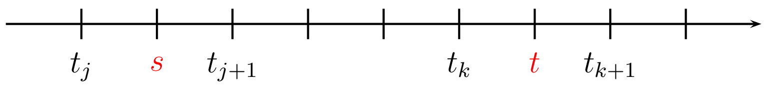
is a martingale transform of the discrete time martingale (M_{t_i}), and hence again a martingale.

(2). Secondly, suppose $s, t \in [t_j, t_{j+1}]$ for some $j \in \{0, 1, 2, \ldots, n-1\}$. Then

$$E[(H_\bullet M)_t - (H_\bullet M)_s \,|\, \mathcal{F}_s] \;=\; E[A_j \cdot (M_t - M_s) \,|\, \mathcal{F}_s] \;=\; A_j \cdot E[M_t - M_s \,|\, \mathcal{F}_s] \;=\; 0$$

because A_j is $\mathcal{F}_{t_j}$-measurable and hence $\mathcal{F}_s$-measurable, and (M_t) is a martingale.

(3). Finally, suppose that $s \in [t_j, t_{j+1}]$ and $t \in [t_k, t_{k+1}]$ with $j < k$.



Then by the tower property for conditional expectations and by (1) and (2),

$$\begin{aligned} E[(H_\bullet M)_t \mid \mathcal{F}_s] &= E[E[E[(H_\bullet M)_t \mid \mathcal{F}_{t_k}] \mid \mathcal{F}_{t_{j+1}}] \mid \mathcal{F}_s] \\ &\overset{(2)}{=} E[E[(H_\bullet M)_{t_k} \mid \mathcal{F}_{t_{j+1}}] \mid \mathcal{F}_s] \overset{(1)}{=} E[(H_\bullet M)_{t_{j+1}} \mid \mathcal{F}_s] \\ &\overset{(2)}{=} (H_\bullet M)_s. \end{aligned}$$

□

Remark (Riemann sum approximations). Non-anticipative Riemann sum approximations of stochastic integrals are Itô integrals of predictable step functions: If (H_t) is an adapted stochastic process and $\pi = \{t_0, t_1, \ldots, t_n\}$ is a partition then

$$\sum_{i=0}^{n-1} H_{t_i} \cdot (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) \;=\; \int_0^t H_s^\pi \, dM_s \tag{5.2.2}$$

where $H^\pi := \sum_{i=0}^{n-1} H_{t_i} \cdot I_{(t_i, t_{i+1}]}$ is a process in $\mathscr{E}$.

Itô's isometry for Brownian motion

Recall that our goal is to prove that non-anticipative Riemann sum approximations for a stochastic integral converge. Let (π_n) be a sequence of partitions of $[0,t]$ with $\mathrm{mesh}(\pi_n) \to 0$. By the remark above, the corresponding Riemann-Itô sums I^{π_n} defined by (5.2.2) are integrals of predictable step functions H^{π_n}. Hence in order to prove that the sequence (I^{π_n}) converges in the Hilbert space M_c^2 it suffices to show that

(1). (H^{π_n}) is a *Cauchy sequence w.r.t. an appropriate norm* on the vector space $\mathscr{E}$, and

(2). the "**Itô map**" $\mathcal{J} : \mathscr{E} \to M_c^2$ defined by

$$\mathcal{J}(H) = H_\bullet M = \int_0^\bullet H_s \, dM_s$$

is *continuous w.r.t. this norm.*

It turns out that we can even identify explicitly a simple norm on $\mathscr{E}$ such that the Itô map is an isometry. We first consider the case where (M_t) is a Brownian motion:

Theorem 5.5 (Itô's isometry for Brownian motion). *If (B_t) is an $(\mathcal{F}_t)$ Brownian motion on $(\Omega, \mathcal{A}, P)$ then for any $u \in [0, \infty]$, and for any process $H \in \mathscr{E}$,*

$$\|H_\bullet B\|^2_{M^2([0,u])} = E\left[\left(\int_0^u H_s \, dB_s\right)^2\right] = E\left[\int_0^u H_s^2 \, ds\right] = \|H\|^2_{L^2(P \otimes \lambda_{(0,u)})} \tag{5.2.3}$$

Proof. Suppose that $H = \sum_{i=0}^{n-1} A_i \cdot I_{(t_i, t_{i+1}]}$ with $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \ldots < t_n$ and A_i bounded and $\mathcal{F}_{t_i}$-measurable. With the notation $\delta_i B := B_{t_{i+1} \wedge u} - B_{t_i \wedge u}$, we obtain

$$E\left[\left(\int_0^u H_s \, dB_s\right)^2\right] = E\left[\left(\sum_{i=0}^{n-1} A_i \delta_i B\right)^2\right] = \sum_{i,k=0}^{n-1} E\left[A_i A_k \cdot \delta_i B \delta_k B\right]. \tag{5.2.4}$$

By the martingale property, the summands on the right hand side vanish for $i \neq k$. Indeed, if, for instance, $i < k$ then

$$E[A_i A_k \delta_i B \delta_k B] \;=\; E[A_i A_k \delta_i B \cdot E[\delta_k B \mid \mathcal{F}_{t_k}]] \;=\; 0.$$

Here we have used in an essential way, that A_k is $\mathcal{F}_{t_k}$-measurable. Similarly,

$$E[A_i^2 \cdot (\delta_i B)^2] \;=\; E[A_i^2 E[(\delta_i B)^2 \mid \mathcal{F}_{t_i}]] \;=\; E[A_i^2 \cdot \delta_i t]$$

by the independence of the increments of Brownian motion. Therefore, by (5.2.4) we obtain

$$E\left[\left(\int_0^u H_s\,dB_s\right)^2\right] \;=\; \sum_{i=0}^{n-1} E[A_i^2 \cdot (t_{i+1} \wedge u - t_i \wedge u)] \;=\; E\left[\int_0^u H_s^2\,ds\right].$$

The assertion now follows by definition of the M^2 norm. □

Theorem 5.5 shows that the linear map

$$\mathcal{J} : \mathscr{E} \to \mathcal{M}_c^2([0,u]), \quad \mathcal{J}(H) \;=\; \left(\int_0^r H_s\,dB_s\right)_{r \in [0,u]},$$

is an isometry if the space $\mathscr{E}$ of simple predictable processes $(s,\omega) \mapsto H_s(\omega)$ is endowed with the L^2 norm

$$\|H\|_{L^2(P \otimes \lambda_{(0,u)})} \;=\; E\left[\int_0^u H_s^2\,ds\right]^{1/2}$$

on the product space $\Omega \times (0,u)$. In particular, $\mathcal{J}$ respects $P \otimes \lambda$ classes, i.e., if $H_s(\omega) = \widetilde{H}_s(\omega)$ for $P \otimes \lambda$-almost every (ω, s) then $\int_0^\bullet H\,dB = \int_0^\bullet \widetilde{H}\,dB$ P-almost surely. Hence $\mathcal{J}$ also induces a linear map between the corresponding spaces of equivalence classes. As usual, we do not always differentiate between equivalence classes and functions, and so we denote the linear map on equivalence classes again by $\mathcal{J}$:

$$\begin{aligned} \mathcal{J} \;:\; \mathscr{E} \subset L^2(P \otimes \lambda_{(0,u)}) &\;\to\; M_c^2([0,u]), \\ \|\mathcal{J}(H)\|_{M^2([0,u])} &\;=\; \|H\|_{L^2(P \otimes \lambda_{(0,u)})}. \end{aligned} \tag{5.2.5}$$

Itô's isometry for martingales

An Itô isometry also holds if Brownian motion is replaced by a continuous square-integrable martingale (M_t). More generally, suppose that $(M_t)_{t\geq 0}$ is a right continuous square integrable $(\mathcal{F}_t)$ martingale satisfying the following assumption:

Assumption A. There exists a non-decreasing adapted continuous process $t \mapsto \langle M\rangle_t$ such that $\langle M\rangle_0 = 0$ and $M_t^2 - \langle M\rangle_t$ is a martingale.

For continuous square integrable martingales, the assumption is always satisfied. Indeed, assuming continuity, the "angle bracket process" $\langle M\rangle_t$ coincides almost surely with the quadratic variation process $[M]_t$ of M, cf. Section 6.3 below. For Brownian motion, Assumption A holds with

$$\langle B\rangle_t \;=\; t.$$

Note that for any $0 \leq s \leq t$, Assumption A implies

$$E\left[(M_t - M_s)^2 \,|\, \mathcal{F}_s\right] \;=\; E\left[M_t^2 - M_s^2 \,|\, \mathcal{F}_s\right] \;=\; E\left[\langle M\rangle_t - \langle M\rangle_s \,|\, \mathcal{F}_s\right]. \tag{5.2.6}$$

Since $t \mapsto \langle M\rangle_t(\omega)$ is continuous and non-decreasing for a given ω, it is the distribution function of a unique positive measure $\langle M\rangle(\omega, dt)$ on $\mathbb{R}_+$.

Theorem 5.6 (Itô's isometry for martingales). *Suppose that $(M_t)_{t\geq 0}$ is a right continuous $(\mathcal{F}_t)$ martingale with angle bracket process $\langle M\rangle$ satisfying Assumption A. Then for any $u \in [0,\infty]$, and for any process $H \in \mathscr{E}$,*

$$\|H_\bullet M\|^2_{M^2([0,u])} \;=\; E\left[\left(\int_0^u H_s\,dM_s\right)^2\right] \;=\; E\left[\int_0^u H_s^2\,d\langle M\rangle_s\right] \tag{5.2.7}$$

where $d\langle M\rangle$ denotes integration w.r.t. the positive measure with distribution function $F(t) = \langle M\rangle_t$.

For Brownian motion $\langle B\rangle_t = t$, so (5.2.7) reduces to (5.2.3).

Proof. The proof is similar to the proof of Theorem 5.5 above. Suppose again that $H = \sum_{i=0}^{n-1} A_i \cdot I_{(t_i,t_{i+1}]}$ with $n \in \mathbb{N}$, $0 \le t_0 < t_1 < \ldots < t_n$ and A_i bounded and $\mathcal{F}_{t_i}$-measurable. With the same notation as in the proof above, we obtain by the martingale properties of M and $M^2 - \langle M\rangle$,

$$E[A_i A_k \, \delta_i M \, \delta_k M] \;=\; 0 \qquad \text{for } i \neq k, \qquad \text{and}$$

$$E[A_i^2 \cdot (\delta_i M)^2] \;=\; E[A_i^2 E[(\delta_i M)^2 \mid \mathcal{F}_{t_i}]] \;=\; E[A_i^2 E[\delta_i \langle M\rangle \mid \mathcal{F}_{t_i}]] \;=\; E[A_i^2 \cdot \delta_i \langle M\rangle].$$

cf. (5.2.6). Therefore,

$$\begin{aligned} E\left[\left(\int_0^u H_s \, dM_s\right)^2\right] \;&=\; E\left[\left(\sum_{i=0}^{n-1} A_i \delta_i M\right)^2\right] = \sum_{i,k=0}^{n-1} E\left[A_i A_k \, \delta_i M \, \delta_k M\right] \\ &=\; \sum_{i=0}^{n-1} E[A_i^2 \, \delta_i \langle M\rangle] \;=\; E\left[\int_0^u H_s^2 \, d\langle M\rangle_s\right]. \end{aligned}$$

□

For a continuous square integrable martingale, Theorem 5.6 implies that the linear map

$$\mathcal{J} : \mathscr{E} \to \mathcal{M}_c^2([0,u]), \quad \mathcal{J}(H) \;=\; \left(\int_0^r H_s \, dM_s\right)_{r \in [0,u]},$$

is an isometry if the space $\mathscr{E}$ of simple predictable processes $(s,\omega) \mapsto H_s(\omega)$ is endowed with the L^2 norm

$$\|H\|_{L^2(\Omega \times (0,u), P_{\langle M\rangle})} \;=\; E\left[\int_0^u H_s^2 \, d\langle M\rangle_s\right]^{1/2}$$

on the product space $\Omega \times (0,u)$ endowed with the positive measure

$$P_{\langle M\rangle}(d\omega \, dt) \;=\; P(d\omega) \, \langle M\rangle(\omega, dt). \tag{5.2.8}$$

Again, we denote the corresponding linear map induced on equivalence classes by the same letter $\mathcal{J}$.

Definition of Itô integrals for square-integrable integrands

From now on we assume that (M_t) is a *continuous* square integrable $(\mathcal{F}_t)$ martingale with angle bracket process $\langle M\rangle_t$. We fix $u \in [0,\infty]$ and consider the isometry

$$\begin{aligned} \mathcal{J} \;:\; \mathscr{E} \subset L^2(\Omega\times(0,u), P_{\langle M\rangle}) \;&\to\; M_c^2([0,u]),. \\ H \;&\mapsto\; H_\bullet M \end{aligned} \tag{5.2.9}$$

mapping an elementary predictable process H to the continuous martingale

$$(H_\bullet M)_t \;=\; \int_0^t H_s\,dM_s.$$

More precisely, we consider the induced map on equivalence classes.

Let $\overline{\mathscr{E}_u}$ denote the closure of the space $\mathscr{E}$ in $L^2(\Omega\times(0,u), P_{\langle M\rangle})$. Since $\mathcal{J}$ is linear with

$$\|\mathcal{J}(H)\|_{M^2([0,u])} \;=\; \|H\|_{L^2(\Omega\times(0,u),P_{\langle M\rangle})} \qquad \text{for any } H\in\mathcal{E},$$

there is a unique extension to a continuous (and even isometric) linear map

$$\overline{\mathcal{J}} \;:\; \overline{\mathscr{E}_u} \quad\subseteq\quad L^2(\Omega\times(0,u),P_{\langle M\rangle}) \quad\to\quad M_c^2([0,u]).$$

This can be used to define the Itô integral for any process in $\overline{\mathscr{E}_u}$, i.e., for any process that can be approximated by predictable step functions w.r.t. the $L^2(P_{\langle M\rangle})$ norm:

$$H_\bullet B \;:=\; \overline{\mathcal{J}}(H), \qquad \int_0^t H_s\,dB_s \;:=\; (H_\bullet B)_t.$$

Explicitly, we obtain the following definition of stochastic integrals for integrands in $\overline{\mathscr{E}_u}$:

Definition (Itô integral). *For $H\in\overline{\mathscr{E}_u}$ the process $H_\bullet M = (\int_0^t H_s\,dM_s)_{t\in[0,u]}$ is the up to modifications unique continuous martingale on $[0,u]$ satisfying*

$$(H_\bullet M)_t \;=\; \lim_{n\to\infty}(H^n_\bullet M)_t \qquad \text{in } L^2(P) \qquad \text{for any } t\in[0,u]$$

whenever (H^n) is a sequence of elementary predictable processes such that $H^n\to H$ in $L^2(\Omega\times(0,u),P_{\langle M\rangle})$.

Remark. (1). By construction, the map $H \mapsto H_\bullet M$ is an isometry from $\overline{\mathscr{E}_u}$ endowed with the $L^2(P_{\langle M\rangle})$ norm to $M_c^2([0,u])$. If $t \mapsto \langle M\rangle_t$ is absolutely continuous, then the closure $\overline{\mathscr{E}_u}$ of the elementary processes actually contains any $(\mathcal{F}_t^P)$ adapted process $(\omega,t) \mapsto H_t(\omega)$ that is square-integrable w.r.t. $P_{\langle M\rangle}$, see XXX below.

(2). The definition above is consistent in the following sense: If $H_\bullet M$ is the stochastic integral defined on the time interval $[0,v]$ and $u \le v$, then the restriction of $H_\bullet M$ to $[0,u]$ coincides with the stochastic integral on $[0,u]$.

For $0 \le s \le t$ we define

$$\int_s^t H_r\,dM_r \;:=\; (H_\bullet M)_t - (H_\bullet M)_s.$$

Exercise. Verify that for any $H \in \overline{\mathscr{E}_t}$,

$$\int_s^t H_r\,dM_r \;=\; \int_0^t H_r\,dB_r - \int_0^t I_{(0,s)}(r)H_r\,dM_r \;=\; \int_0^t I_{(s,t)}(r)H_r\,dM_r.$$

Having defined the Itô integral, we now show that bounded adapted processes with left-continuous sample paths are contained in the closure of the simple predictable processes, and the corresponding stochastic integrals are limits of predictable Riemann sum approximations. As above, we consider a sequence (π_n) of partitions of $\mathbb{R}_+$ such that $\mathrm{mesh}(\pi_n) \to 0$.

Theorem 5.7 (Approximation by Riemann-Itô sums). *Let $u \in (0,\infty)$, and suppose that $(H_t)_{t\in[0,u)}$ is an $(\mathcal{F}_t^P)$ adapted stochastic process on $(\Omega,\mathcal{A},P)$ such that $(t,\omega) \mapsto H_t(\omega)$ is product-measurable and bounded. If $t \mapsto H_t$ is P-almost surely left continuous then H is in $\overline{\mathscr{E}_u}$, and*

$$\int_0^t H_s\,dM_s \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} H_s(M_{s'\wedge t} - M_{s\wedge t}), \qquad t \in [0,u], \tag{5.2.10}$$

w.r.t. convergence uniformly in t in the $L^2(P)$ sense.

Remark. (1). In particular, a subsequence of the predictable Riemann sum approximations converges uniformly in t with probability one.

(2). The assertion also holds if H is unbounded with $\sup_{s\leq u}|H_s| \in \mathcal{L}^2(P)$.

Proof. For any $t \in [0,u]$, the Riemann sums on the right hand side of (5.2.10) are the stochastic integrals $\int_0^t H_s^n\,dM_s$ of the predictable step functions

$$H_t^n \;:=\; \sum_{s\in\pi_n, s<u} H_s \cdot I_{(s,s']}(t), \qquad n \in \mathbb{N}.$$

By left-continuity, $H_t^n \to H_t$ as $n \to \infty$ for any $t \in [0,u]$, P-almost surely. Therefore, $H^n \to H$ $P_{\langle M\rangle}$-almost surely, and, by dominated convergence,

$$H^n \;\to\; H \quad \text{ in } L^2(P_{\langle M\rangle}).$$

Here we have used that the sequence (H^n) is uniformly bounded since H is bounded by assumption. Now, by Itô's isometry,

$$\int_0^\bullet H_s\,dM_s \;=\; \lim_{n\to\infty}\int_0^\bullet H_s^n\,dM_s \qquad \text{in } M_c^2([0,u]).$$

□

Identification of admissible integrands

Let $u \in (0,\infty]$. We have already shown that if $u < \infty$ then any product-measurable adapted bounded process with left-continuous sample paths is in $\overline{\mathscr{E}_u}$. More generally, we will prove now that if $M_t = B_t$ is a Brownian motion then any adapted process in $\mathcal{L}^2(P\otimes\lambda_{[0,u)})$ is contained in $\overline{\mathscr{E}_u}$, and hence "integrable" w.r.t. (B_t). Let $\mathcal{L}_a^2(0,u)$ denote the linear space of all product-measurable, $(\mathcal{F}_t^P)$ adapted stochastic processes $(\omega,t)\mapsto H_t(\omega)$ defined on $\Omega\times(0,u)$ such that

$$E\left[\int_0^u H_t^2\,dt\right] \;<\; \infty.$$

The corresponding space of equivalence classes of $P\otimes\lambda$ versions is denoted by $L_a^2(0,u)$.

Lemma 5.8. *$L_a^2(0,u)$ is a closed linear subspace of $L^2(P\otimes\lambda_{(0,u)})$.*

Proof. It only remains to show that an $L^2(P\otimes\lambda)$ limit of $(\mathcal{F}_t^P)$ adapted processes again has an $(\mathcal{F}_t^P)$ adapted $P\otimes\lambda$-version. Hence consider a sequence $H^n \in \mathcal{L}_a^2(0,u)$ with $H^n \to H$ in $L^2(P\otimes\lambda)$. Then there exists a subsequence (H^{n_k}) such that $H_t^{n_k}(\omega) \to H_t(\omega)$ for $P\otimes\lambda$-almost every $(\omega,t)\in\Omega\times(0,u)$. The process $\widetilde{H}$ defined by $\widetilde{H}_t(\omega) := \lim H_t^{n_k}(\omega)$ if the limit exists, $\widetilde{H}_t(\omega) := 0$ otherwise, is an $(\mathcal{F}_t^P)$ adapted version of H. □

We can now identify the class of integrands H for which the stochastic integral $H_\bullet B$ is well-defined as a limit of integrals of predictable step functions in $M_c^2([0,u])$:

Theorem 5.9 (Admissible integrands for Brownian motion). *For any $u\in(0,\infty]$,*

$$\overline{\mathscr{E}_u} \;=\; L_a^2(0,u).$$

Proof. Since $\mathscr{E}\subseteq\mathcal{L}_a^2(0,u)$ it only remains to show the inclusion "$\supseteq$". Hence fix a process $H\in\mathcal{L}_a^2(0,u)$. We will prove in several steps that H can be approximated by simple predictable processes w.r.t. the $L^2(P\otimes\lambda_{(0,u)})$ norm:

(1). Suppose first that H is bounded and has almost surely continuous trajectories. Then for $u<\infty$, H is in $\overline{\mathscr{E}_u}$ by Theorem 5.7. For $u=\infty$, H is still in $\overline{\mathscr{E}_u}$ provided there exists $t_0\in(0,\infty)$ such that H_t vanishes for $t\ge t_0$.

(2). Now suppose that (H_t) is bounded and, if $u=\infty$, vanishes for $t\ge t_0$. To prove $H\in\overline{\mathscr{E}_u}$ we approximate H by continuous adapted processes. To this end let $\psi_n:\mathbb{R}\to[0,\infty), n\in\mathbb{N}$, be continuous functions such that $\psi(s)=0$ for $s\notin(0,1/n)$ and $\int_{-\infty}^{\infty}\psi_n(s)\,ds=1$. Let $H^n := H*\psi_n$, i.e.,

$$H_t^n(\omega) \;=\; \int_0^{1/n} H_{t-\varepsilon}(\omega)\psi_n(\varepsilon)\,d\varepsilon, \tag{5.2.11}$$

where we set $H_t := 0$ for $t\le 0$. We prove that

(a) $H^n\to H$ in $L^2(P\otimes\lambda_{(0,u)})$, and

(b) $H^n\in\overline{\mathscr{E}_u}$ for any $n\in\mathbb{N}$.

Combining (a) and (b), we see that H is in $\overline{\mathscr{E}_u}$ as well.

(a) Since H is in $\mathcal{L}^2(P \otimes \lambda_{(0,u)})$, we have

$$\int_0^u H_t(\omega)^2 \, dt \;<\; \infty \tag{5.2.12}$$

for P-almost every ω. It is a standard fact from analysis that (5.2.12) implies

$$\int_0^u |H_t^n(\omega) - H_t(\omega)|^2 \, dt \quad \longrightarrow \quad 0 \qquad \text{as } n \to \infty.$$

By dominated convergence, we obtain

$$E\left[\int_0^u |H_t^n - H_t|^2 \, dt\right] \quad \longrightarrow \quad 0 \qquad \text{as } n \to \infty \tag{5.2.13}$$

because H is bounded, the sequence (H_n) is uniformly bounded, and H and H^n vanish for $t \geq t_0 + 1$.

(b) This is essentially a consequence of part (1) of the proof. We sketch how to verify that H^n satisfies the assumptions made there:

- The sample paths $t \mapsto H_t^n(\omega)$ are continuous for all ω.
- $|H_t^n|$ is bounded by $\sup |H|$.
- The map $(\omega, t) \mapsto H_t^n(\omega)$ is product measurable by (5.2.11) and Fubini's Theorem, because the map $(\omega, t, \varepsilon) \mapsto H_{t-\varepsilon}(\omega)\psi_\varepsilon(\omega)$ is product measurable.
- If the process (H_t) is *progressively measurable*, i.e., if the map $(s, \omega) \mapsto H_s(\omega)$ $(s \in (0,t), \omega \in \Omega)$ is measurable w.r.t. the product σ-algebra $\mathcal{B}(0,t) \otimes \mathcal{F}_t^P$ for any $t \geq 0$, then (H_t^n) is $(\mathcal{F}_t^P)$ adapted by (5.2.11) and Fubini's Theorem. This is for example the case if (H_t) is right continuous or left continuous.
- In general, one can prove that (H_t) has a progressively measurable modification, whence (H_t^n) has an $(\mathcal{F}_t^P)$ adapted modification. We omit the details.

(3). We finally prove that general $H \in \mathcal{L}_a^2(0,u)$ are contained in $\overline{\mathscr{E}_u}$. This is a consequence of (2), because we can approximate H by the processes

$$H_t^n \;:=\; ((H_t \wedge n) \vee (-n)) \cdot I_{(0,n)}(t), \qquad n \in \mathbb{N}.$$

These processes are bounded, they vanish for $t \geq n$, and $H^n \to H$ in $L^2(P \otimes \lambda_{(0,u)})$. By (2), H^n is contained in $\overline{\mathscr{E}_u}$ for any n, so H is in $\overline{\mathscr{E}_u}$ as well.

□

Remark (Riemann sum approximations). For discontinuous integrands, the predictable Riemann sum approximations considered above do not converge to the stochastic integral in general. However, one can prove that for $u < \infty$ any process $H \in L^2_a(0, u)$ is the limit of the simple predictable processes

$$H_t^n \;=\; \sum_{i=1}^{2^n-1} 2^n \int_{(i-1)2^{-n}u}^{i2^{-n}u} H_s \, ds \cdot I_{(i2^{-n}u,(i+1)2^{-n}u]}(t)$$

w.r.t. the $L^2(P \otimes \lambda_{[0,u)})$ norm, cf. [XXXSteele: "Stochastic calculus and financial applications", Sect 6.6]. Therefore, the stochastic integral $\int_0^t H \, dB$ can be approximated for $t \leq u$ by the correspondingly modified Riemann sums.

For continuous martingales, a similar statement as in Theorem 5.9 holds provided the angle bracket process is absolutely continuous. Let $\mathcal{L}^2_a(0, u; M)$ denote the linear space of all product-measurable, $(\mathcal{F}_t^P)$ adapted stochastic processes $(\omega, t) \mapsto H_t(\omega)$ such that

$$E\left[\int_0^u H_t^2 \, d\langle M\rangle_t\right] \;<\; \infty.$$

The corresponding space of equivalence classes w.r.t. $P_{\langle M\rangle}$ is denoted by $L^2_a(0, u; M)$.

Exercise (Admissible integrands w.r.t. martingales). Suppose that (M_t) is a continuous square integrable $(\mathcal{F}_t)$ martingale. Show that if almost surely, $t \mapsto \langle M\rangle_t$ is absolutely continuous, then the closure $\overline{\mathcal{E}_u}$ of the elementary processes w.r.t. the $L^2(P_{\langle M\rangle})$ norm is given by

$$\overline{\mathcal{E}_u} \;=\; L^2_a(0, u; M).$$

5.3 Localization

Square-integrability of the integrand is still an assumption that we would like to avoid, since it is not always easy to verify or may even fail to hold. The key to extending the class of admissible integrands further is localization, which enables us to define a stochastic integral w.r.t. a continuous martingale for any continuous adapted process. The price we have to pay is that for integrands that are not square integrable, the Itô integral is in general not a martingale, but only a local martingale.

Throughout this section we assume that M_t is a continuous square integrable martingale with absolutely continuous angle bracket process $\langle M\rangle_t$.

Local dependence on integrand and integrator

The approximations considered in the last section imply that the stochastic integral depends locally both on the integrand and on the integrator in the following sense:

Corollary 5.10. *Suppose that $T : \Omega \to [0,\infty]$ is a random variable, $M, \widetilde{M}$ are square integrable martingales with absolutely continuous angle bracket processes $\langle M\rangle, \langle\widetilde{M}\rangle$, and $H, \widetilde{H}$ are processes in $\mathcal{L}^2_a(0,\infty;M)$, $\mathcal{L}^2_a(0,\infty;\widetilde{M})$ respectively, such that almost surely, $H_t = \widetilde{H}_t$ for any $t \in [0,T)$ and $M_t = \widetilde{M}_t$ for any $t \in [0,T]$. Then almost surely,*

$$\int_0^t H_s\, dM_s = \int_0^t \widetilde{H}_s\, d\widetilde{M}_s \qquad \text{for any } t \in [0,T]. \tag{5.3.1}$$

Proof. We go through the same approximations as in the proof of Theorem 5.9 above:

(1). Suppose first that H_t and $\widetilde{H}_t$ are almost surely continuous and bounded, and there exists $t_0 \in \mathbb{R}_+$ such that $H_t = \widetilde{H}_t = 0$ for $t \geq t_0$. Let (π_n) be a sequence of partitions with $\mathrm{mesh}(\pi_n) \to 0$. Then by Theorem 5.7,

$$\int_0^t H\, dM = \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (M_{s'\wedge t} - M_s), \qquad \text{and}$$

$$\int_0^t \widetilde{H}\, d\widetilde{M} = \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \widetilde{H}_s \cdot (\widetilde{M}_{s'\wedge t} - \widetilde{M}_s)$$

with P-almost sure uniform convergence on finite time-intervals along a common subsequence. For $t \leq T$ the right-hand sides coincide, and thus (5.3.1) holds true.

(2). Now suppose that H and $\widetilde{H}$ are bounded and $H_t = \widetilde{H}_t = 0$ for $t \geq t_0$. Then the approximations

$$H_t^n \;=\; \int_0^{1/n} H_{t-\varepsilon}\psi_n(\varepsilon)\, d\varepsilon, \qquad \widetilde{H}_t^n \;=\; \int_0^{1/n} \widetilde{H}_{t-\varepsilon}\psi_n(\varepsilon)$$

(with ψ_n defined as in the proof of Theorem 5.9 and $H_t := \widetilde{H}_t := 0$ for $t < 0$) coincide for $t \leq T$. Hence by (1), on $\{t \leq T\}$,

$$\int_0^t H\, dM \;=\; \lim \int_0^t H^n\, dM \;=\;=\; \lim \int_0^t \widetilde{H}^n\, d\widetilde{M} \;=\; \int_0^t \widetilde{H}\, d\widetilde{M},$$

where the convergence holds again almost surely uniformly in t along a subsequence.

(3). Finally, in the general case the assertion follows by approximating H and $\widetilde{H}$ by the bounded processes

$$H_t^n \;=\; ((H_t \wedge n) \vee (-n)) \cdot I_{[0,n]}(t), \;\; \widetilde{H}_t^n \;=\; ((\widetilde{H}_t \wedge n) \vee (-n)) \cdot I_{[0,n]}(t).$$

□

Itô integrals for locally square-integrable integrands

Let M be a continuous square integrable martingale with absolutely continuous angle bracket process $\langle M \rangle$, and let $T : \Omega \to [0, \infty]$ be an $(\mathcal{F}_t^P)$ stopping time. We will also be interested in the case where $T = \infty$. Let $\mathcal{L}^2_{a,\mathrm{loc}}(0, T; M)$ denote the linear space consisting of all stochastic processes $(t, \omega) \mapsto H_t(\omega)$ defined for $t \in [0, T(\omega))$ such that the trivially extended process

$$\widetilde{H}_t \;:=\; \begin{cases} H_t & \text{for } t < T, \\ 0 & \text{for } t \geq T, \end{cases}$$

is product measurable in (t, ω), adapted w.r.t. the filtration $(\mathcal{F}_t^P)$, and

$$t \mapsto H_t(\omega) \quad \text{is in } \mathcal{L}^2_{\mathrm{loc}}([0, T(\omega)), d\langle M \rangle(\omega)) \quad \text{for } P\text{-a.e. } \omega. \tag{5.3.2}$$

Here for $u \in (0, \infty]$, the space $\mathcal{L}^2_{\text{loc}}([0, u), d\langle M\rangle(\omega))$ consists of all measurable functions $f : [0, u) \to [-\infty, \infty]$ such that $\int_0^s f(t)^2 \, d\langle M\rangle_t(\omega) < \infty$ for any $s \in (0, u)$. In particular, it contains all continuous functions.

From now on, we use the notation $H_t \cdot I_{\{t<T\}}$ for the trivial extension $(\widetilde{H}_t)_{0\le t<\infty}$ of a process $(H_t)_{0\le t<T}$ beyond the stopping time T. Locally square integrable adapted processes allow for a localization by stopping times:

Lemma 5.11 (Localization by stopping). *If $(H_t)_{0\le t<T}$ is a process in $\mathcal{L}^2_{a,loc}(0, T; M)$ then there exists an increasing sequence $(T_n)_{n\in\mathbb{N}}$ of $(\mathcal{F}^P_t)$ stopping times such that $T = \sup T_n$ almost surely, and*

$$H_t \cdot I_{\{t<T_n\}} \in \mathcal{L}^2_a(0, \infty; M) \qquad \text{for any } n \in \mathbb{N}.$$

Proof. One easily verifies that the random variables T_n defined by

$$T_n := \inf \left\{ 0 \le t < T : \int_0^t H_s^2 \, d\langle M\rangle_s \ge n \right\} \wedge T, \qquad n \in \mathbb{N}, \tag{5.3.3}$$

are $(\mathcal{F}^P_t)$ stopping times. Moreover, for almost every ω, the function $t \mapsto H_t(\omega)$ is in $\mathcal{L}^2_{\text{loc}}([0, T), d\langle M\rangle(\omega))$. Hence the function $t \mapsto \int_0^t H_s(\omega)^2 \, d\langle M\rangle_s$ is increasing and finite on $[0, T(\omega))$, and therefore $T_n(\omega) \nearrow T(\omega)$ as $n \to \infty$. Since T_n is an $(\mathcal{F}^P_t)$ stopping time, the process $H_t \cdot I_{\{t<T_n\}}$ is $(\mathcal{F}^P_t)$-adapted, and by (5.3.3),

$$E\left[\int_0^\infty (H_s \cdot I_{\{s<T_n\}})^2 \, d\langle M\rangle_s\right] = E\left[\int_0^{T_n} H_s^2 \, d\langle M\rangle_s\right] \le n \qquad \text{for any } n.$$

□

A sequence of stopping times as in the lemma will also be called a **localizing sequence**. We can now extend the definition of the Itô integral to locally square-integrable adapted integrands:

Definition (Itô integral with locally square integrable integrand). *For a process $H \in \mathcal{L}^2_{a,loc}(0, T; M)$, the Itô stochastic integral w.r.t. the martingale M is defined for $t \in [0, T)$ by*

$$\int_0^t H_s \, dM_s := \int_0^t H_s \cdot I_{\{s<\hat{T}\}} \, dM_s \qquad \text{for any } t \in [0, \hat{T}] \tag{5.3.4}$$

whenever $\hat{T}$ is an $(\mathcal{F}_t^P)$ stopping time such that $H_t \cdot I_{\{t<\hat{T}\}} \in \mathcal{L}_a^2(0,\infty;M)$.

Theorem 5.12. *For $H \in \mathcal{L}^2_{a,loc}(0,T;M)$ the Itô integral $t \mapsto \int_0^t H_s dM_s$ is almost surely well defined by (5.3.4) as a continuous process on $[0,T)$.*

Proof. We have to verify that the definition does not depend on the choice of the localizing stopping times. This is a direct consequence of Corollary 5.10: Suppose that $\hat{T}$ and $\widetilde{T}$ are stopping times such that $H_t \cdot I_{\{t<\hat{T}\}}$ and $H_t \cdot I_{\{t<\widetilde{T}\}}$ are both in $\mathcal{L}_a^2(0,\infty;M)$. Since the two trivially extended processes agree on $[0,\hat{T}\wedge\widetilde{T})$, Corollary 5.10 implies that almost surely,

$$\int_0^t H_s \cdot I_{\{s<\hat{T}\}}\, dM_s \;=\; \int_0^t H_s \cdot I_{\{s<\widetilde{T}\}}\, dM_s \qquad \text{for any } t \in [0,\hat{T}\wedge\widetilde{T}).$$

Hence, by Lemma 5.11, the stochastic integral is well defined on $[0,T)$. □

Stochastic integrals as local martingales

Itô integrals w.r.t. square integrable martingales are not necessarily martingales if the integrands are not square integrable. However, they are still local martingales in the sense of the definition stated below.

Definition (Predictable stopping time). *An $(\mathcal{F}_t^P)$ stopping time T is called **predictable** iff there exists an increasing sequence $(T_k)_{k\in\mathbb{N}}$ consisting of $(\mathcal{F}_t^P)$ stopping times such that $T_k < T$ on $\{T \neq 0\}$ for any k, and $T = \sup T_k$.*

Example (Hitting time of a closed set). The hitting time T_A of a closed set A by a continuous adapted process is predictable, as it can be approximated from below by the hitting times T_{A_k} of the neighbourhoods $A_k = \{x : \mathrm{dist}(x,A) < 1/k\}$. On the other hand, the hitting time of an open set is not predictable in general.

Definition (Local martingale). *Suppose that $T : \Omega \to [0, \infty]$ is a predictable stopping time. A stochastic process $M_t(\omega)$ defined for $0 \le t < T(\omega)$ is called a **local martingale up to time** T, if and only if there exists an increasing sequence (T_k) of stopping times with $T = \sup T_k$ such that for any $k \in \mathbb{N}$, $T_k < T$ on $\{T > 0\}$, and the stopped process $(M_{t \wedge T_k})$ is a martingale for $t \in [0, \infty)$.*

Recall that by the Optional Stopping Theorem, a continuous martingale stopped at a stopping time is again a martingale. Therefore, any continuous martingale $(M_t)_{t \ge 0}$ is a local martingale up to $T = \infty$. Even if (M_t) is assumed to be uniformly integrable, the converse implication fails to hold:

Exercise (A uniformly integrable local martingale that is not a martingale). Let $x \in \mathbb{R}^3$ with $x \neq 0$, and suppose that (B_t) is a three-dimensional Brownian motion with initial value $B_0 = x$. Prove that the process $M_t = 1/|B_t|$ is a uniformly integrable local martingale up to $T = \infty$, but (M_t) is not a martingale.

On the other hand, note that if (M_t) is a continuous local martingale up to $T = \infty$, and the family $\{M_{t \wedge T_k} : k \in \mathbb{N}\}$ is uniformly integrable for each *fixed* $t \ge 0$, then (M_t) is a martingale, because for $0 \le s \le t$

$$E[M_t \mid \mathcal{F}_s] \;=\; \lim_{k \to \infty} E[M_{t \wedge T_k} \mid \mathcal{F}_s] \;=\; \lim_{k \to \infty} M_{s \wedge T_k} \;=\; M_s$$

with convergence in L^1.

As a consequence of the definition of the Itô integral by localization, we immediately obtain:

Theorem 5.13 (Itô integrals as local martingales). *Suppose that T is a predictable stopping time w.r.t. $(\mathcal{F}_t^P)$. Then for any $H \in \mathcal{L}^2_{a,loc}(0, T; M)$, the Itô integral process $t \mapsto \int_0^t H_s \, dM_s$ is a continuous local martingale up to time T.*

Proof. We can choose an increasing sequence (T_k) of stopping times such that $T_k < T$ on $\{T > 0\}$ and $H_t \cdot I_{\{t<T_k\}} \in \mathcal{L}_a^2(0,\infty;M)$ for any k. Then, by definition of the Itô integral,

$$\int_0^{t\wedge T_k} H_s\,dM_s \;=\; \int_0^{t\wedge T_k} H_s \cdot I_{\{s<T_k\}}\,dM_s \qquad \text{almost surely for any } k\in\mathbb{N},$$

and the right-hand side is a continuous martingale in $M_c^2([0,\infty))$. □

The theorem shows that for a predictable $(\mathcal{F}_t^P)$ stopping time T, the Itô map $H \mapsto \int_0^\bullet H\,dM$ extends to a linear map

$$\mathcal{J} \;:\; L^2_{\mathrm{loc}}(0,T;M) \;\longrightarrow M_{c,\mathrm{loc}}([0,T)),$$

where $L^2_{\mathrm{loc}}(0,T;M)$ is the space of equivalence classes of processes in $\mathcal{L}^2_{\mathrm{loc}}(0,T;M)$ that coincide for $P_{\langle M\rangle}$-a.e. (ω,t), and $M_{c,\mathrm{loc}}([0,T))$ denotes the space of equivalence classes of continuous local $(\mathcal{F}_t^P)$ martingales up to time T w.r.t. P-almost sure coincidence. Note that different notions of equivalence are used for the integrands and the integrals.

We finally observe that continuous local martingales (and hence stochastic integrals w.r.t. continuous martingales) can always be localized by a sequence of *bounded* martingales in $M_c^2([0,\infty)$:

Exercise (Localization by bounded martingales). Suppose that (M_t) is a continuous local martingale up to time T, and (T_k) is a localizing sequence of stopping times.

(1). Show that

$$\widetilde{T}_k \;=\; T_k \wedge \inf\{t\geq 0 \;:\; |M_t|\geq k\}\wedge k$$

is another localizing sequence, and for all k, the stopped processes $\left(M_{t\wedge\widetilde{T}_k}\right)_{t\in[0,\infty)}$ are bounded martingales in $M_c^2([0,\infty))$.

(2). Show that if $T=\infty$ then $\hat{T}_k := \inf\{t\geq 0 : |M_t|\geq k\}$ is also a localizing sequence for M.

Approximation by Riemann-Itô sums

If the integrand (H_t) of a stochastic integral $\int H\, dB$ has continuous sample paths then local square integrability always holds, and the stochastic integral is a limit of Riemann-Itô sums: Let (π_n) be a sequence of partition of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$.

Theorem 5.14. *Suppose that T is a predictable stopping time, and $(H_t)_{0\le t<T}$ is a stochastic process defined for $t < T$. If the sample paths $t \mapsto H_t(\omega)$ are continuous on $[0, T(\omega))$ for any ω, and the trivially extended process $H_t \cdot I_{\{t<T\}}$ is $(\mathcal{F}_t^P)$ adapted, then H is in $\mathcal{L}^2_{a,loc}(0, T; M)$, and for any $t \ge 0$,*

$$\int_0^t H_s\, dM_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (M_{s'\wedge t} - M_s) \qquad \text{on } \{t < T\} \tag{5.3.5}$$

with convergence in probability.

Proof. Let $\lfloor t\rfloor_n = \max\{s \in \pi_n : s \le t\}$ denote the next partition point below t. By continuity,

$$H_t \cdot I_{\{t<T\}} \;=\; \lim_{n\to\infty} H_{\lfloor t\rfloor_n} \cdot I_{\{t<T\}}.$$

Hence $(H_t \cdot I_{\{t<T\}})$ is $(\mathcal{F}_t^P)$ adapted. It is also product-measurable, because

$$H_{\lfloor t\rfloor_n} \cdot I_{\{t<T\}} \;=\; \sum_{s\in\pi_n} H_s \cdot I_{\{s<T\}} \cdot I_{[s,s')}(t) \cdot I_{(0,\infty)}(T - t).$$

By continuity, $t \mapsto H_t(\omega)$ is locally bounded for every ω, and thus H is in $\mathcal{L}^2_{\mathrm{a,loc}}(0, T; M)$. Moreover, suppose that (T_k) is a sequence of stopping times approaching T from below in the sense of the definition of a predictable stopping time given above. Then

$$\widetilde{T}_k \;:=\; T_k \wedge \inf\{t \ge 0 : |H_t| \ge k\}, \qquad k \in \mathbb{N},$$

is a localizing sequence of stopping times with $H_t \cdot I_{\{t<T_k\}}$ in $\mathcal{L}^2_a(0,T;M)$ for any k, and $\widetilde{T}_k \nearrow T$. Therefore, by definition of the Itô integral and by Theorem 5.7,

$$\begin{aligned} \int_0^t H_s\, dM_s &= \int_0^t H_s \cdot I_{\{s<\widetilde{T}_k\}}\, dM_s = \int_0^t H_s \cdot I_{\{s\le\widetilde{T}_k\}}\, dM_s \\ &= \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} H_s \cdot (M_{s'\wedge t} - M_s) \qquad \text{on } \{t \le \widetilde{T}_k\} \end{aligned}$$

w.r.t. convergence in probability. Since

$$P\left[\{t<T\} \setminus \bigcup_k \{t \le \widetilde{T}_k\}\right] = 0,$$

we obtain (5.3.5). □

Chapter 6

Itô's formula and pathwise integrals

Our approach to Itô's formula in this chapter follows that of [Föllmer: Stochastic Analysis, Vorlesungsskript Uni Bonn WS91/92]. We start with a heuristic derivation of the formula that will be the central topic of this chapter.

Suppose that $s \mapsto X_s$ is a function from $[0,t]$ to $\mathbb{R}$, and F is a smooth function on $\mathbb{R}$. If (π_n) is a sequence of partitions of the interval $[0,t]$ with $\mathrm{mesh}(\pi_n) \to 0$ then by Taylor's theorem

$$F(X_{s'}) - F(X_s) \;=\; F'(X_s)\cdot(X_{s'}-X_s) + \frac{1}{2}F''(X_s)\cdot(X_{s'}-X_s)^2 + \text{higher order terms.}$$

Summing over $s \in \pi_n$ we obtain

$$F(X_t) - F(X_0) \;=\; \sum_{s\in\pi_n} F'(X_s)\cdot(X_{s'}-X_s) + \frac{1}{2}F''(X_s)\cdot(X_{s'}-X_s)^2 + \ldots \tag{6.0.1}$$

We are interested in the limit of this formula as $n \to \infty$.

(a) Classical case, e.g. X continuously differentiable For $X \in C^1$ we have

$$\begin{aligned} X_{s'} - X_s &= \frac{dX_s}{ds}(s'-s) + O(|s-s'|^2), \qquad \text{and} \\ (X_{s'}-X_s)^2 &= O(|s-s'|^2). \end{aligned}$$

Therefore, the second order terms can be neglected in the limit of (6.0.1) as $\mathrm{mesh}(\pi_n) \to 0$. Similarly, the higher order terms can be neglected, and we obtain the limit equation

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\,dX_s, \tag{6.0.2}$$

or, in differential notation,

$$dF(X_t) \;=\; F'(X_t)\,dX_t, \tag{6.0.3}$$

Of course, (6.0.3) is just the chain rule of classical analysis, and (6.0.2) is the equivalent chain rule for Stieltjes integrals, cf. Section 6.1 below.

(b) X_t Brownian motion If (X_t) is a Brownian motion then

$$E[(X_{s'} - X_s)^2] \;=\; s' - s.$$

Summing these expectations over $s \in \pi_n$, we obtain the value t independently of n. This shows that the sum of the second order terms in (6.0.1) can not be neglected anymore. Indeed, as $n \to \infty$, a law of large numbers type result implies that we can almost surely replace the squared increments $(X_{s'} - X_s)^2$ in (6.0.1) asymptotically by their expectation values. The higher order terms are on average $O(|s' - s|^{3/2})$ whence their sum can be neglected. Therefore, in the limit of (6.0.1) as $n \to \infty$ we obtain the modified chain rule

$$F(X_t) - F(X_0) \;=\; \int_0^t F'(X_s)\,dX_s + \frac{1}{2}\int_0^t F''(X_s)\,ds \tag{6.0.4}$$

with probability one. The equation (6.0.4) is the basic version of Itô's celebrated formula.

In Section 6.1, we will first introduce Stieltjes integrals and the chain rule from Stieltjes calculus systematically. In Section 6.2 we prove a general version of Itô's formula for continuous functions with finite quadratic variation in dimension one. Here the setup and the proof are still purely deterministic. As an aside we obtain a pathwise definition for stochastic integrals involving only a single one-dimensional process due to Föllmer. After computing the quadratic variation of Brownian motion in Section 6.3, we consider first consequences of Itô's formula for Brownian motions and continuous martingales. Section 6.4 contains extensions to the multivariate and time-dependent case, as well as further applications.

6.1 Stieltjes integrals and chain rule

In this section, we define Lebesgue-Stieltjes integrals w.r.t. deterministic functions of finite variation, and prove a corresponding chain rule. The resulting calculus can then be applied path by path to stochastic processes with sample paths of finite variation.

Lebesgue-Stieltjes integrals

Fix $u \in (0, \infty]$, and suppose that $t \mapsto A_t$ is a right-continuous and non-decreasing function on $[0, u)$. Then $A_t - A_0$ is the distribution function of the positive measure μ on $(0, u)$ determined uniquely by

$$\mu_A[(s,t]] \;=\; A_t - A_s \qquad \text{for any } 0 \leq s \leq t < u.$$

Therefore, we can define integrals of type $\int_s^t H_s \, dA_s$ as Lebesgue integrals w.r.t. the measure μ_A. We extend μ trivially to the interval $[0, u)$, so $\mathcal{L}^1_{\text{loc}}([0, u), \mu_A)$ is the space of all functions $H : [0, u) \to \mathbb{R}$ that are integrable w.r.t. μ_A on any interval $(0, t)$ with $t < u$. Then for any $u \in [0, \infty]$ and any function $H \in \mathcal{L}^1_{\text{loc}}([0, u), \mu_A)$, the **Lebesgue-Stieltjes integral of H w.r.t. A** is defined by

$$\int_s^t H_r \, dA_r \;:=\; \int H_r \cdot I_{(s,t]}(r) \mu_A(dr) \qquad \text{for } 0 \leq s \leq t < u.$$

It is easy to verify that the definition is consistent, i.e., varying u does not change the definition of the integrals, and that $t \mapsto \int_0^t H_r \, dA_r$ is again a right-continuous function.

For an arbitrary right-continuous function $A : [0, u) \to \mathbb{R}$, the (first order) variation of A on an interval $[0, t)$ is defined by

$$V_t^{(1)}(A) \;:=\; \sup_\pi \sum_{s \in \pi} |A_{s' \wedge t} - A_{s \wedge t}| \qquad \text{for } 0 \leq t < u,$$

where the supremum is over all partitions π of $\mathbb{R}_+$. The function $t \mapsto A_t$ is said to be **(locally) of finite variation** on the interval $[0, u)$ iff $V_t^{(1)}(A) < \infty$ for any $t \in [0, u)$.

Any right-continuous function of finite variation can be written as the difference of two non-decreasing right-continuous functions. In fact, we have

$$A_t \;=\; A_t^{\nearrow} - A_t^{\searrow} \tag{6.1.1}$$

with

$$A_t^{\nearrow} \;=\; \sup_{\pi} \sum_{s\in\pi} (A_{s'\wedge t} - A_{s\wedge t})^+ \;=\; \frac{1}{2}(V_t^{(1)}(A) + A_t), \tag{6.1.2}$$

$$A_t^{\searrow} \;=\; \sup_{\pi} \sum_{s\in\pi} (A_{s'\wedge t} - A_{s\wedge t})^- \;=\; \frac{1}{2}(V_t^{(1)}(A) - A_t). \tag{6.1.3}$$

Exercise. Prove that if A_t is right-continuous and is locally of finite variation on $[0, u)$ then the functions $V_t^{(1)}(A)$, $A_t^{\nearrow}$ and $A_t^{\searrow}$ are all right-continuous and non-decreasing for $t < u$.

Remark (Hahn-Jordan decomposition). The functions $A_t^{\nearrow} - A_0^{\nearrow}$ and $A_t^{\searrow} - A_0^{\searrow}$ are again distribution functions of positive measures μ_A^+ and μ_A^- on $(0, u)$. Correspondingly, $A_t - A_0$ is the distribution function of the signed measure

$$\mu_A[B] \;:=\; \mu_A^+[B] - \mu_A^-[B], \qquad B \in \mathcal{B}(0, u), \tag{6.1.4}$$

and $V_t^{(1)}$ is the distribution of the measure $|\mu_A| = \mu_A^+ - \mu_A^-$. It is a consequence of (6.1.5) and (6.1.6) that the measures μ_A^+ and μ_A^- are singular, i.e., the mass is concentrated on disjoint sets S^+ and S^-. The decomposition (6.1.7) is hence a particular case of the Hahn-Jordan decomposition of a signed measure μ of finite variation into a positive and a negative part, and the measure $|\mu|$ is the total variation measure of μ, cf. e.g. [Alt: Lineare Funktionalanalysis].

We can now apply (6.1.1) to define Lebesgue-Stieltjes integrals w.r.t. functions of finite variation. A function is integrable w.r.t. a signed measure μ if and only if it is integrable w.r.t. both the positive part μ^+ and the negative part μ^-. The Lebesgue integral w.r.t. μ is then defined as the difference of the Lebesgue integrals w.r.t. μ^+ and μ^-. Correspondingly, we define the Lebesgue-Stieltjes integral w.r.t. a function A_t of finite variation as the Lebesgue integral w.r.t. the associated signed measure μ_A:

Definition. *Suppose that $t \mapsto A_t$ is right-continuous and locally of finite variation on $[0, u)$. Then the* ***Lebesgue-Stieltjes integral w.r.t.*** *A is defined by*

$$\int_s^t H_r \, dA_r \; := \; \int H_r \cdot I_{(s,t]}(r) \, dA_r^{\nearrow} - \int H_r \cdot I_{(s,t]}(r) \, dA_r^{\searrow}, \qquad 0 \le s \le t < u,$$

for any function $H \in \mathcal{L}^1_{loc}((0,u), |dA|)$ where

$$\mathcal{L}^1_{loc}((0,u), |dA|) := \mathcal{L}^1_{loc}((0,u), dA^{\nearrow}) \cap \mathcal{L}^1_{loc}((0,u), dA^{\searrow})$$

is the intersection of the local $\mathcal{L}^1$ spaces w.r.t. the positive measures $dA^{\nearrow} = \mu_A^+$ and $dA^{\searrow} = \mu_A^-$ on $[0,u)$, or, equivalently, the local $\mathcal{L}^1$ space w.r.t. the total variation measure $|dA| = |\mu_A|$.

Remark. (1). *Simple integrands:* If $H_t = \sum_{i=0}^{n-1} c_i \cdot I_{(t_i, t_{i+1}]}$ is a step function with $0 \le t_0 < t_1 < \ldots < t_n < u$ and $c_0, c_1, \ldots, c_{n-1} \in \mathbb{R}$ then

$$\int_0^t H_s \, dA_s \; = \; \sum_{i=0}^{n-1} c_i \cdot (A_{t_{i+1} \wedge t} - A_{t_i \wedge t}).$$

(2). *Continuous integrands; Riemann-Stieltjes integral:* If $H : [0,u) \to \mathbb{R}$ is a continuous function then the Stieltjes integral can be approximated by Riemann sums:

$$\int_0^t H_s \, dA_s \; = \; \lim_{n \to \infty} \sum_{\substack{s \in \pi_n \\ s < t}} H_s \cdot (A_{s' \wedge t} - A_s), \qquad t \in [0,u),$$

for any sequence (π_n) of partitions of $\mathbb{R}_+$ such that $\mathrm{mesh}(\pi_n) \to 0$. For the proof note that the step functions

$$H_r^n \; = \; \sum_{\substack{s \in \pi_n \\ s < t}} H_s \cdot I_{(s,s']}(r), \qquad r \in [0,u),$$

converge to H_r pointwise on $(0, u)$ by continuity. Moreover, again by continuity, H_r is locally bounded on $[0, u)$,and hence the sequence H_r^n is locally uniformly bounded. Therefore,

$$\int H_r^n I_{(0,t]}(r)\, dA_r \;\longrightarrow\; \int H_r I_{(0,t]}(r)\, dA_r$$

for any $t < n$ by the dominated convergence theorem.

(3). *Absolutely continuous integrators:* If A_t is an absolutely continuous function on $[0, u)$ then A_t has locally finite variation

$$V_t^{(1)}(A) \;=\; \int_0^t |A_s'|\, ds \;<\; \infty \qquad \text{for } t \in [0, u).$$

The signed measure μ_A with distribution function $A_t - A_0$ is then absolutely continuous w.r.t. Lebesgue measure with Radon-Nikodym density

$$\frac{d\mu_A}{dt}(t) \;=\; A_t' \qquad \text{for almost every } t \in [0, u).$$

Therefore,

$$\mathcal{L}^1_{\text{loc}}([0, u), |dA|) \;=\; \mathcal{L}^1_{\text{loc}}([0, u), |A'|dt),$$

and the Lebesgue-Stieltjes integral of a locally integrable function H is given by

$$\int_0^t H_s\, dA_s \;=\; \int_0^t H_s A_s'\, ds \qquad \text{for } t \in [0, u).$$

In the applications that we are interested in, the integrand will mostly be continuous, and the integrator absolutely continuous. Hence Remarks (2) and (3) above apply.

The chain rule in Stieltjes calculus

We are now able to prove Itô's formula in the special situation where the integrator has finite variation. In this case, the second order correction disappears, and Itô's formula reduces to the classical chain rule from Stieltjes calculus:

Theorem 6.1 (Fundamental Theorem of Stieltjes Calculus). *Suppose that $A : [0,u) \to \mathbb{R}$ is a continuous function of locally finite variation. Then for any $F \in C^2(\mathbb{R})$,*

$$F(A_t) - F(A_0) \;=\; \int_0^t F'(A_s)\,dA_s \qquad \forall t \in [0,u). \tag{6.1.5}$$

Proof. Let $t \in [0,u)$ be given. Choose a sequence of partitions (π_n) of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$, and let

$$\delta A_s \;:=\; A_{s'\wedge t} - A_{s\wedge t} \qquad \text{for } s \in \pi_n,$$

where, as usual, s' denotes the next partition point. By Taylor's formula, for $s \in \pi_n$ with $s < t$ we have

$$F(A_{s'\wedge t}) - F(A_s) \;=\; F'(A_s)\delta A_s + \frac{1}{2}F''(Z_s)\cdot(\delta A_s)^2,$$

where Z_s is an intermediate value between A_s and $A_{s'\wedge t}$. Summing over $s \in \pi_n$, we obtain

$$F(A_t) - F(A_0) \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} F'(A_s)\delta A_s + \frac{1}{2}\sum_{\substack{s\in\pi_n\\ s<t}} F''(Z_s)(\delta A_s)^2. \tag{6.1.6}$$

As $n \to \infty$, the first (Riemann) sum converges to the Stieltjes integral $\int_0^t F'(A_s)\,dA_s$ by continuity of $F'(A_s)$, cf. Remark (2) above.

To see that the second sum converges to zero, note that the range of the continuous function A restricted to $[0,t]$ is a bounded interval. Since F'' is continuous by assumption, F'' is bounded on this range by a finite constant c. As Z_s is an intermediate value between A_s and $A_{s'\wedge t}$, we obtain

$$\left|\sum_{\substack{s\in\pi_n\\ s<t}} F''(Z_s)(\delta A_s)^2\right| \;\le\; c\cdot\sum_{\substack{s\in\pi_n\\ s<t}} (\delta A_s)^2 \;\le\; c\cdot V_t^{(1)}(A)\cdot \sup_{\substack{s\in\pi_n\\ s<t}} |\delta A_s|.$$

Since $V_t^{(1)}(A) < \infty$, and A is a uniformly continuous function on $[0,t]$, the right hand side converges to 0 as $n \to \infty$. Hence we obtain (6.1.5) in the limit of (6.1.6) as $n \to \infty$. □

To see that (6.1.5) can be interpreted as a chain rule, we write the equation in differential form:

$$dF(A) \;=\; F'(A)dA. \tag{6.1.7}$$

In general, the equation (6.1.7) is to be understood mathematically only as an abbreviation for the integral equation (6.1.5). For intuitive arguments, the differential notation is obviously much more attractive than the integral form of the equation. However, for the differential form to be useful at all, we should be able to multiply the equation (6.1.7) by another function, and still obtain a valid equation. This is indeed possible due to the next result, which states briefly that if $dI = H\,dA$ then also $G\,dI = GH\,dA$:

Theorem 6.2 (Stieltjes integrals w.r.t. Stieltjes integrals). *Suppose that $I_s = \int_0^s H_r dA_r$ where $A : [0,u) \to \mathbb{R}$ is a function of locally finite variation, and $H \in \mathcal{L}^1_{loc}([0,u), |dA|)$. Then the function $s \mapsto I_s$ is again right continuous with locally finite variation $V_t^{(1)}(I) \leq \int_0^t |H|\,|dA| < \infty$, and, for any function $G \in \mathcal{L}^1_{loc}([0,u), |dI|)$,*

$$\int_0^t G_s\,dI_s \;=\; \int_0^t G_s H_s\,dA_s \qquad \textit{for } t \in [0,u). \tag{6.1.8}$$

Proof. Right continuity of I_t and the upper bound for the variation are left as an exercise.

We now use Riemann sum approximations to prove (6.1.8) if G is continuous. For a partition $0 = t_0 < t_1 < \ldots < t_k = t$, we have

$$\sum_{i=0}^{n-1} G_{t_i}(I_{t_{i+1}} - I_{t_i}) \;=\; \sum_{i=0}^{n-1} G_{t_i} \cdot \int_{t_i}^{t_{i+1}} H_s\,dA_s \;=\; \int_0^t G_{\lfloor s \rfloor} H_s\,dA_s$$

where $\lfloor s \rfloor$ denotes the largest partition point $\leq s$. Choosing a sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$, the integral on the right hand side converges to the Lebesgue-Stieltjes integral $\int_0^t G_s H_s \, dA_s$ by continuity of G and the dominated convergence theorem, whereas the Riemann sum on the left hand side converges to $\int_0^t G_s \, dI_s$. Hence (6.1.8) holds for continuous G. The equation for general $G \in \mathcal{L}^1_{\mathrm{loc}}([0,u), |dI|)$ follows then by standard arguments. $\square$

6.2 Quadratic variation and Itô's formula

Our next goal is to derive a generalization of the chain rule from Stieltjes calculus to continuous functions that are not of finite variation. Examples of such functions are typical sample paths of Brownian motion. As pointed out above, an additional term will appear in the chain rule in this case.

Quadratic variation

Consider once more the approximation (6.1.6) that we have used to prove the fundamental theorem of Stieltjes calculus. We would like to identify the limit of the last sum $\sum_{s \in \pi_n} F''(Z_s)(\delta A_s)^2$ when A has unfinite variation on finite intervals. For $F'' = 1$ this limit is called the quadratic variation of A if it exists:

Definition. *Let $u \in (0, \infty]$ and let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. The* ***quadratic variation*** *$[X]_t$ of a continuous function $X : [0,u) \to \mathbb{R}$ w.r.t. the sequence (π_n) is defined by*

$$[X]_t \;=\; \lim_{n \to \infty} \sum_{s \in \pi_n} (X_{s' \wedge t} - X_{s \wedge t})^2 \qquad \text{for } t \in [0,u)$$

whenever the limit exists.

WARNINGS (Dependence on partition, classical 2-variation)

(1). The quadratic variation should not be confused with the classical 2-variation defined by

$$V_t^{(2)}(X) := \sup_{\pi} \sum_{s \in \pi} |X_{s' \wedge t} - X_{s \wedge t}|^2$$

where the supremum is over all partitions π. The classical 2-variation $V_t^{(2)}(X)$ is strictly positive for any function X that is not constant on $[0,t]$ whereas $[X]_t$ vanishes in many cases, cf. Example (1) below.

(2). In general, the quadratic variation may depend on the sequence of partitions considered. See however Examples (1) and (3) below.

Example. (1). *Functions of finite variation:* For any continuous function $A : [0,u) \to \mathbb{R}$ of locally finite variation, the quadratic variation along (π_n) vanishes:

$$[A]_t = 0 \qquad \text{for any } t \in [0,u).$$

In fact, for $\delta A_s = A_{s' \wedge t} - A_{s \wedge t}$ we have

$$\sum_{s \in \pi_n} |\delta A_s|^2 \leq V_t^{(1)}(A) \cdot \sup_{\substack{s \in \pi_n \\ s < t}} |\delta A_s| \to 0 \qquad \text{as } n \to \infty$$

by uniform continuity and since $V_t^{(1)}(A) < \infty$.

(2). *Perturbations by functions of finite variation:* If the quadratic variation $[X]_t$ of X w.r.t. (π_n) exists, then $[X+A]_t$ also exists, and

$$[X+A]_t = [X]_t.$$

This holds since

$$\sum |\delta(X+A)|^2 = \sum (\delta X)^2 + 2\sum \delta X \delta A + \sum (\delta A)^2,$$

and the last two sums converge to 0 as $\mathrm{mesh}(\pi_n) \to 0$ by Example (1) and the Cauchy-Schwarz inequality.

(3). *Brownian motion:* If $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion then P-almost surely,

$$[B]_t = t \qquad \text{for all } t \geq 0$$

w.r.t. any *fixed* sequence (π_n) of partitions such that $\mathrm{mesh}(\pi_n) \to 0$, cf. Theorem 6.6 below.

(4). *Itô processes:* If $I_t = \int_0^t H_s\, dB_s$ is the stochastic integral of a process $H \in \mathcal{L}^2_{a,\mathrm{loc}}(0,\infty)$ w.r.t. Brownian motion then almost surely, the quadratic variation w.r.t. a fixed sequence of partitions is

$$[I]_t = \int_0^t H_s^2\, ds \qquad \text{for all } t \geq 0.$$

(5). *Continuous martingales:* $[M]$ exists and is almost surely finite, see below.

Note that in Examples (3), (4) and (5), the exceptional sets may depend on the sequence (π_n). If it exists, the quadratic variation $[X]_t$ is a non-decreasing function in t. In particular, Stieltjes integrals w.r.t. $[X]$ are well-defined provided $[X]$ is right continuous.

Lemma 6.3. *Suppose that $X : [0,u) \to \mathbb{R}$ is a continuous function. If the quadratic variation $[X]_t$ along (π_n) exists for $t \in [0,u)$, and $t \mapsto [X]_t$ is continuous then*

$$\sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (X_{s'\wedge t} - X_s)^2 \quad \longrightarrow \quad \int_0^t H_s\, d[X]_s \qquad \text{as } n\to\infty \tag{6.2.1}$$

for any continuous function $H : [0,u) \to \mathbb{R}$ and any $t \geq 0$.

Remark. Heuristically, the assertion of the lemma says that

$$\text{“}\int H\, d[X] = \int H\, (dX)^2\text{”},$$

i.e., the infinitesimal increments of the quadratic variation are something like squared infinitesimal increments of X. This observation is crucial for controlling the second order terms in the Taylor expansion used for proving the Itô-Doeblin formula.

Proof. The sum on the left-hand side of (6.2.1) is the integral of H w.r.t. the finite positive measure

$$\mu_n := \sum_{\substack{s\in\pi_n\\ s<t}} (X_{s'\wedge t} - X_s)^2 \cdot \delta_s$$

on the interval $[0,t)$. The distribution function of μ_n is

$$F_n(u) =: \sum_{\substack{s\in\pi_n\\ s\le t}} (X_{s'\wedge t}-X_s)^2, \qquad u\in[0,t].$$

As $n\to\infty$, $F_n(u)\to[X]_u$ for any $u\in[0,t]$ by continuity of X. Since $[X]_u$ is a continuous function of u, convergence of the distribution functions implies weak convergence of the measures μ_n to the measure $d[X]$ on $[0,t)$ with distribution function $[X]$. Hence,

$$\int H_s\mu_n(ds) \longrightarrow \int H_s\, d[X]_s \qquad \text{as } n\to\infty$$

for any continuous function $H:[0,t]\to\mathbb{R}$. □

Itô's formula and pathwise integrals in $\mathbb{R}^1$

We are now able to complete the proof of the following purely deterministic (pathwise) version of the one-dimensional Itô formula going back to [Föllmer: Calcul d'Itô sans probabilités, Sém. Prob XV, LNM850XXX]:

Theorem 6.4 (Itô's formula without probability). *Suppose that $X:[0,u)\to\mathbb{R}$ is a continuous function with continuous quadratic variation $[X]$ w.r.t. (π_n). Then for any function F that is C^2 in a neighbourhood of $X([0,u))$, and for any $t\in[0,u)$, the Itô integral*

$$\int_0^t F'(X_s)\,dX_s = \lim_{n\to\infty}\sum_{\substack{s\in\pi_n\\ s<t}} F'(X_s)\cdot(X_{s'\wedge t}-X_s) \tag{6.2.2}$$

exists, and Itô's formula

$$F(X_t)-F(X_0) = \int_0^t F'(X_s)\,dX_s + \frac{1}{2}\int_0^t F''(X_s)\,d[X]_s \tag{6.2.3}$$

holds. In particular, if the quadratic variation $[X]$ does not depend on (π_n) then the Itô integral (6.2.2) does not depend on (π_n) either.

Note that the theorem **implies the existence** of $\int_0^t f(X_s)\,dX_s$ for any function $f \in C^1(\mathbb{R})$! Hence this type of Itô integrals can be defined in a purely deterministic way without relying on the Itô isometry. Unfortunately, the situation is more complicated in higher dimensions, cf. **??** below.

Proof. Fix $t \in [0,u)$ and $n \in \mathbb{N}$. As before, for $s \in \pi_n$ we set $\delta X_s = X_{s'\wedge t} - X_{s\wedge t}$ where s' denotes the next partition point. Then as above we have

$$F(X_t) - F(X_0) \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_s)\delta X_s + \frac{1}{2}\sum_{\substack{s\in\pi_n\\ s<t}} F''(Z_s^{(n)})(\delta X_s)^2 \tag{6.2.4}$$

$$\;=\; \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_s)\delta X_s + \frac{1}{2}\sum_{\substack{s\in\pi_n\\ s<t}} F''(X_s)(\delta X_s)^2 + \sum_{\substack{s\in\pi_n\\ s<t}} R_s^{(n)}, \tag{6.2.5}$$

where $Z_s^{(n)}$ is an intermediate point between X_s and $X_{s'\wedge t}$, and $R_s^{(n)} := \frac{1}{2}(F''(Z_s^{(n)}) - F''(X_s)) \cdot (\delta X_s)^2$. As $n \to \infty$, the second sum on the right hand side of (6.2.4) converges to $\int_0^t F''(X_s)\,d[X]_s$ by Lemma 6.3. We claim that the sum of the remainders $R_s^{(n)}$ converges to 0. To see this note that $Z_s^{(n)} = X_r$ for some $r \in [s, s'\wedge t]$, whence

$$|R_s^{(n)}| \;=\; |F''(Z_s^{(n)}) - F''(X_s)| \cdot (\delta X_s)^2 \;\le\; \frac{1}{2}\varepsilon_n(\delta X_s)^2,$$

where

$$\varepsilon_n \;:=\; \sup_{\substack{a,b\in[0,t]\\ |a-b|\le \mathrm{mesh}(\pi_n)}} |F''(X_a) - F''(X_b)|.$$

As $n \to \infty$, ε_n converges to 0 by uniform continuity of $F'' \circ X$ on the interval $[0,t]$. Thus

$$\sum |R_s^{(n)}| \;\le\; \frac{1}{2}\varepsilon_n \sum_{\substack{s\in\pi_n\\ s<t}} (\delta X_s)^2 \;\to\; 0 \qquad \text{as well,}$$

because the sum of the squared increments converges to the finite quadratic variation $[X]_t$.

We have shown that all the terms on the right hand side of (6.2.4) except the first

Riemann-Itô sum converge as $n \to \infty$. Hence, by (6.2.4), the limit $\int_0^t F'(X_s)\, dX_s$ of the Riemann Itô sums also exists, and the limit equation (6.2.2) holds. □

Remark. (1). In differential notation, we obtain the Itô chain rule

$$dF(X) \;=\; F'(X)\, dX + \frac{1}{2}F''(X)\, d[X]$$

which includes a second order correction term due to the quadratic variation. A justification for the differential notation is given in Section **??**.

(2). For functions X with $[X] = 0$ we recover the classical chain rule $dF(X) = F'(X)\, dX$ from Stieltjes calculus as a particular case of Itô's formula.

Example. (1). *Exponentials:* Choosing $F(x) = e^x$ in Itô's formula, we obtain

$$e^{X_t} - e^{X_0} \;=\; \int_0^t e^{X_s}\, dX_s + \frac{1}{2}\int_0^t e^{X_s}\, d[X]_s,$$

or, in differential notation,

$$de^X \;=\; e^X\, dX + \frac{1}{2}e^X\, d[X].$$

Thus e^X does *not* solve the Itô differential equation

$$dZ \;=\; Z\, dX \tag{6.2.6}$$

if $[X] \neq 0$. An appropriate renormalization is required instead. We will see below that the correct solution of (6.2.6) is given by

$$Z_t \;=\; \exp\left(X_t - [X]/2\right),$$

cf. the first example below Theorem 6.18.

(2). *Polynomials:* Similarly, choosing $F(x) = x^n$ for some $n \in \mathbb{N}$, we obtain

$$dX^n \;=\; nX^{n-1}\, dX + \frac{n(n-1)}{2}X^{n-2}\, [X].$$

Again, X^n does not solve the equation $dX^n = nX^{n-1}\, dX$. Here, the appropriate renormalization leads to the Hermite polynomials $: X :^n$, cf. the second example below Theorem 6.18.

The chain rule for anticipative integrals

The form of the second order correction term appearing in Itô's formula depends crucially on choosing non-anticipative Riemann sum approximations. For limits of anticipative Riemann sums, we obtain different correction terms, and hence also different notions of integrals.

Theorem 6.5. *Suppose that $X : [0,u) \to \mathbb{R}$ is continuous with continuous quadratic variation $[X]$ along (π_n). Then for any function F that is C^2 in a neighbourhood of $X([0,u))$ and for any $t \geq 0$, the* ***backward Itô integral***

$$\int_0^t F'(X_s)\,\hat{d}X_s \;:=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} F'(X_{s'\wedge t})\cdot(X_{s'\wedge t}-X_s),$$

and the ***Stratonovich integral***

$$\int_0^t F'(X_s)\,\circ dX_s \;:=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \frac{1}{2}(F'(X_s)+F'(X_{s'\wedge t}))\cdot(X_{s'\wedge t}-X_s)$$

exist, and

$$F(X_t)-F(X_0) \;=\; \int_0^t F'(X_s)\,\hat{d}X_s - \frac{1}{2}\int_0^t F''(X_s)\,d[X]_s \qquad (6.2.7)$$

$$= \int_0^t F'(X_s)\,\circ dX_s. \qquad (6.2.8)$$

Proof. The proof of the backward Itô formula (6.2.7) is completely analogous to that of Itô's formula. The Stratonovich formula (6.2.8) follows by averaging the Riemann sum approximations to the forward and backward Itô rule. □

Note that Stratonovich integrals satisfy the classical chain rule

$$\circ dF(X) \;=\; F'(X)\,\circ dX.$$

This makes them very attractive for various applications. For example, in stochastic differential geometry, the chain rule is of fundamental importance to construct stochastic processes that stay on a given manifold. Therefore, it is common to use Stratonovich instead of Itô calculus in this context, cf. the corresponding example in the next section. On the other hand, Stratonovich calculus has a significant disadvantage against Itô calculus: The Stratonovich integrals

$$\int_0^t H_s \circ dB_s = \lim_{n\to\infty} \sum \frac{1}{2}(H_s + H_{s'\wedge t})(B_{s'\wedge t} - B_s)$$

w.r.t. Brownian motion typically are not martingales, because the coefficients $\frac{1}{2}(H_s + H_{s'\wedge t})$ are not predictable.

6.3 Itô's formula for Brownian motion and martingales

Our next aim is to compute the quadratic variation and to state Itô's formula for typical sample paths of Brownian motion. More generally, we will show that the quadratic variation exists almost surely for continuous local martingales.

Let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. We note first that for any function $t \mapsto X_t$ the identity

$$X_t^2 - X_0^2 = \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t}^2 - X_s^2) = V_t^n + 2I_t^n \tag{6.3.1}$$

with

$$V_t^n = \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t} - X_s)^2 \quad \text{and} \quad I_t^n = \sum_{\substack{s\in\pi_n \\ s<t}} X_s \cdot (X_{s'\wedge t} - X_s)$$

holds. The equation (6.3.1) is a discrete approximation of Itô's formula for the function $F(x) = x^2$. The remainder terms in the approximation vanish in this particular case.

Note that by (6.3.1), the quadratic variation $[X]_t = \lim_{n\to\infty} V_t^n$ exists if and only if the Riemann sum approximations I_t^n to the Itô integral $\int_0^t X_s\, dX_s$ converge:

$$\exists\ [X]_t = \lim_{n\to\infty} V_t^n \quad \iff \quad \exists \int_0^t X_s\, dX_s = \lim_{n\to\infty} I_t^n.$$

Now suppose that (X_t) is a continuous martingale with $E[X_t^2] < \infty$ for any $t \geq 0$. Then the Riemann sum approximations (I_t^n) are continuous martingales for any $n \in \mathbb{N}$. Therefore, by the maximal inequality, for a given $u > 0$, the processes (I_t^n) and (V_t^n) converge uniformly for $t \in [0, u]$ in $L^2(P)$ if and only if the random variables I_u^n or V_u^n respectively converge in $L^2(P)$.

Quadratic variation of Brownian motion

For the sample paths of a Brownian motion B, the quadratic variation $[B]$ exists almost surely along any *fixed* sequence of partitions (π_n) with $\mathrm{mesh}(\pi_n) \to 0$, and $[B]_t = t$ a.s. In particular, $[B]$ is a *deterministic* function that does not depend on (π_n). The reason is a law of large numbers type effect when taking the limit of the sum of squared increments as $n \to \infty$.

Theorem 6.6 (P. Lévy). *If (B_t) is a one-dimensional Brownian motion on $(\Omega, \mathcal{A}, P)$ then as $n \to \infty$*

$$\sup_{t\in[0,u]} \left| \sum_{\substack{s\in\pi_n \\ s<t}} (B_{s'\wedge t} - B_s)^2 - t \right| \quad \longrightarrow \quad 0 \qquad P\text{-a.s. and in } L^2(\Omega, \mathcal{A}, P) \tag{6.3.2}$$

for any $u \in (0, \infty)$, and for each sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$.

Warning. (1). Although the almost sure limit in (6.3.2) does not depend on the sequence (π_n), the exceptional set may depend on the chosen sequence!

(2). The classical quadratic variation $V_t^{(2)}(B) = \sup_\pi \sum_{s\in\pi} (\delta B_s)^2$ is almost surely infinite for any $t \geq 0$. The classical p-variation is almost surely finite if and only if $p > 2$.

Proof. (1). *L^2-convergence for fixed t:* As usual, the proof of L^2 convergence is comparatively simple. For $V_t^n = \sum_{s\in\pi_n} (\delta B_s)^2$ with $\delta B_s = B_{s'\wedge t} - B_{s\wedge t}$, we have

$$\begin{aligned} E[V_t^n] &= \sum_{s\in\pi_n} E[(\delta B_s)^2] = \sum_{s\in\pi_n} \delta s = t, \qquad \text{and} \\ \mathrm{Var}[V_t^n] &= \sum_{s\in\pi_n} \mathrm{Var}[(\delta B_s)^2] = \sum_{s\in\pi_n} E[((\delta B_s)^2 - \delta s)^2] \\ &= E[(Z^2-1)^2] \cdot \sum_{s\in\pi_n} (\delta s)^2 \le \text{const.} \cdot t \cdot \mathrm{mesh}(\pi_n) \end{aligned}$$

where Z is a standard normal random variable. Hence, as $n \to \infty$,

$$V_t^n - t = V_t^n - E[V_t^n] \to 0 \qquad \text{in } L^2(\Omega, \mathcal{A}, P).$$

Moreover, by (6.3.1), $V_t^n - V_t^m = I_t^n - I_t^m$ is a continuous martingale for any $n, m \in \mathbb{N}$. Therefore, the maximal inequality yields uniform convergence of V_t^n to t for t in a finite interval in the $L^2(P)$ sense.

(2). *Almost sure convergence if $\sum \mathrm{mesh}(\pi_n) < \infty$:* Similarly, by applying the maximal inequality to the process $V_t^n - V_t^m$ and taking the limit as $m \to \infty$, we obtain

$$P\left[\sup_{t\in[0,u]} |V_t^n - t| > \varepsilon\right] \le \frac{2}{\varepsilon^2} E[(V_t^n - t)^2] \le \text{const.} \cdot t \cdot \mathrm{mesh}(\pi_n)$$

for any given $\varepsilon > 0$ and $u \in (0, \infty)$. If $\sum \mathrm{mesh}(\pi_n) < \infty$ then the sum of the probabilities is finite, and hence $\sup_{t\in[0,u]} |V_t^n - t| \to 0$ almost surely by the Borel-Cantelli Lemma.

(3). *Almost sure convergence if $\sum \mathrm{mesh}(\pi_n) = \infty$:* In this case, almost sure convergence can be shown by the backward martingale convergence theorem. We refer to Proposition 2.12 in [Revuz, YorXXX], because for our purposes almost sure convergence w.r.t arbitrary sequences of partitions is not essential.

□

Itô's formula for Brownian motion

By Theorem 6.6, we can apply Theorem 6.7 to almost every sample path of a one-dimensional Brownian motion (B_t):

Theorem 6.7 (Itô's formula for Brownian motion). *Suppose that $F \in C^2(I)$ where $I \subseteq \mathbb{R}$ be an open interval. Then almost surely,*

$$F(B_t) - F(B_0) = \int_0^t F'(B_s)\, dB_s + \frac{1}{2}\int_0^t F''(B_s)\, ds \qquad \text{for all } t < T, \tag{6.3.3}$$

where $T = \inf\{t \geq 0 \,:\, B_t \notin I\}$ is the first exit time from I.

Proof. For almost every ω, the quadratic variation of $t \mapsto B_t(\omega)$ along a fixed sequence of partitions is t. Moreover, for any $r < T(\omega)$, the function F is C^2 on a neighbourhood of $\{B_t(\omega) : t \in [0,r]\}$. The assertion now follows from Theorem 6.7 by noting that the pathwise integral and the Itô integral as defined in Section 5 coincide almost surely since both are limits of Riemann-Itô sums w.r.t. uniform convergence for t in a finite interval, almost surely along a common (sub)sequence of partitions. □

Consequences

(1). *Doob decomposition in continuous time:* The Itô integral $M_t^F = \int_0^t F'(B_s)\, dB_s$ is a local martingale up to T, and M_t^F is a square integrable martingale if $I = \mathbb{R}$ and F' is bounded. Therefore, (6.3.3) can be interpreted as a *continuous time Doob decomposition* of the process $(F(B_t))$ into the (local) martingale part M^F and an adapted process of finite variation. This process takes over the role of the predictable part in discrete time.

In particular, we obtain:

Corollary 6.8 (Martingale problem for Brownian motion). *Brownian motion is a solution of the martingale problem for the operator* $\mathscr{L} = \frac{1}{2}\frac{d^2}{dx^2}$ *with domain* $\mathrm{Dom}(\mathscr{L}) = \{F \in C^2(\mathbb{R}) : \frac{dF}{dx}$ *is bounded*$\}$*, i.e., the process*

$$M_t^F = F(B_t) - F(B_0) - \int_0^t (\mathscr{L}f)(B_s)\,ds$$

is a martingale for any $F \in \mathrm{Dom}(\mathscr{L})$.

The corollary demonstrates how Itô's formula can be applied to obtain solutions of martingale problems, cf./ below for generalizations.

(2). *Kolmogorov's forward equation:* Taking expectation values in (6.3.3), we recover Kolmogorov's equation

$$E[F(B_t)] = E[F(B_0)] + \int_0^t E[(\mathscr{L}F)(B_s)]\,ds \qquad \forall\, t \geq 0$$

for any $F \in C_b^2(\mathbb{R})$. In differential form,

$$\frac{d}{dt}E[F(B_t)] = \frac{1}{2}E[(\mathscr{F}'')(B_t)].$$

(3). *Computation of expected values:* The Itô formula can be applied in many ways to compute expectation values:

Example. (a) For any $n \in \mathbb{N}$, the process

$$B_t^n - \frac{n(n-1)}{2}\int_0^t B_s^{n-2}\,ds = n \cdot \int_0^t B_s^{n-1}\,dB_s$$

is a martingale. By taking expectation values for $t = 1$ we obtain the recursion

$$\begin{aligned} E[B_1^n] &= \frac{n(n-1)}{2}\int_0^1 E[B_s^{n-2}]\,ds = \frac{n(n-1)}{2}\int_0^1 s^{n-2/2}\,ds \cdot E[B_1^{n-2}] \\ &= (n-1)\cdot E[B_1^{n-2}] \end{aligned}$$

for the moments of the standard normally distributed random variable B_1. Of course this identity can be obtained directly by integration by parts in the Gaussian integral $\int x^n \cdot e^{-x^2/2}\, dx$.

(b) For $\alpha \in \mathbb{R}$, the process

$$\exp(\alpha B_t) - \frac{\alpha^2}{2} \int_0^t \exp(\alpha B_s)\, ds \;=\; \alpha \int_0^t \exp(\alpha B_s)\, dB_s$$

is a martingale because $E[\int_0^t \exp(2\alpha B_s)\, ds] < \infty$. Denoting by $T_b = \min\{t \geq 0 \,:\, B_t = b\}$ the first passage time to a level $b > 0$, we obtain the identity

$$E\left[\int_0^{T_b} \exp(\alpha B_s)\, ds\right] \;=\; \frac{2}{\alpha^2}(e^{\alpha b} - 1) \qquad \text{for any } \alpha > 0$$

by optional stopping and dominated convergence.

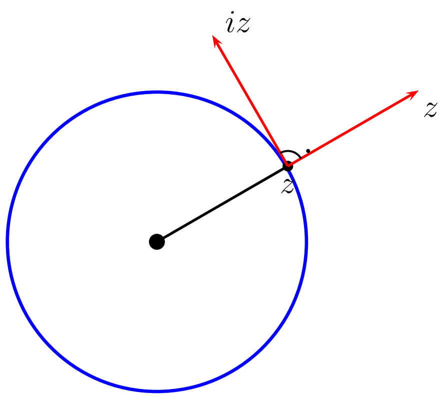
Itô's formula is also the key tool to derive or solve stochastic differential equations for various stochastic processes of interest:

Example (Brownian motion on $\mathbf{S}^1$). Brownian motion on the unit circle $S^1 = \{z \in \mathbb{C} \,:\, |z| = 1\}$ is the process given by

$$Z_t \;=\; \exp(iB_t) \;=\; \cos B_t + i \cdot \sin B_t$$

where (B_t) is a standard Brownian motion on $\mathbb{R}^1$. Itô's formula yields the stochastic differential equation

$$dZ_t \;=\; \mathfrak{t}(Z_t)\, dB_t - \frac{1}{2}\mathfrak{n}(Z_t)\, dt, \tag{6.3.4}$$



where $\mathfrak{t}(z) = iz$ is the unit tangent vector to S^1 at the point z, and $\mathfrak{n}(z) = z$ is the outer normal vector. If we would omit the correction term $-\frac{1}{2}\mathfrak{n}(Z_t)\,dt$ in (6.3.4), the solution to the s.d.e. would not stay on the circle. This is contrary to classical o.d.e. where the correction term is not required. For Stratonovich integrals, we obtain the modified equation

$$\circ dZ_t \;=\; \mathfrak{t}(Z_t)\,\circ dB_t,$$

which does not involve a correction term!

Quadratic variation of continuous martingales

Next, we will show that the sample paths of continuous local martingales almost surely have finite quadratic variation. Let (M_t) be a continuous local martingale, and fix a sequence (π_n) of partitions of $\mathbb{R}_+$ with $\text{mesh}(\pi_n) \to 0$. Let

$$V_t^n \;=\; \sum_{s\in\pi_n} (M_{s'\wedge t} - M_{s\wedge t})^2$$

denote the quadratic variation of M along π_n. Recall the crucial identity

$$M_t^2 - M_0^2 \;=\; \sum_{s\in\pi_n} \left(M_{s'\wedge t}^2 - M_{s\wedge t}^2\right) \;=\; V_t^n \;+\; 2I_t^n \tag{6.3.5}$$

where $I_t^n = \sum_{s\in\pi_n} M_s(M_{s'\wedge t} - M_{s\wedge t})$ are the Riemann sum approximations to the Itô integral $\int_0^t M\,dM$. The identity shows that V_t^n converges (uniformly) as $n\to\infty$ if and only if the same holds for I_t^n. Moreover, in this case, we obtain the limit equation

$$M_t^2 - M_0^2 \;=\; [M]_t \;+\; 2\int_0^t M_s\,dM_s \tag{6.3.6}$$

which is exactly Itô's equation for $F(x) = x^2$.

Theorem 6.9 (Existence of quadratic variation). *Suppose that (M_t) is a continuous local martingale on $(\Omega, \mathcal{A}, P)$. Then there exist a continuous non-decreasing process $t\mapsto [M]_t$ and a continuous local martingale $t\mapsto \int_0^t M\,dM$ such that as $n\to\infty$,*

$$\sup_{s\in[0,t]} |V_s^n - [M]_s| \;\to\; 0 \qquad \textit{and} \qquad \sup_{s\in[0,t]} \left| I_s^n - \int_0^s M\,dM \right| \;\to\; 0$$

in probability for any $t \geq 0$, and in $L^2(P)$ respectively if M is bounded. Moreover, the identity (6.3.6) holds.

Notice that in the theorem, we do not assume the existence of an angle bracket process $\langle M \rangle$. Indeed, the theorem proves that for continuous local martingales, the angle bracket process always exists and it coincides almost surely with the quadratic variation process $[M]$! We point out that for discontinuous martingales, $\langle M \rangle$ and $[M]$ do not coincide.

Proof. We first assume that M is a bounded martingale: $|M_t| \leq C$ for some finite constant C. We then show that (I_n) is a Cauchy sequence in the Hilbert space $M_c^2([0,t])$ for any given $t \in \mathbb{R}_+$. To this end let $n, m \in \mathbb{N}$. We assume without loss of generality that $\pi_m \subseteq \pi_n$, otherwise we compare to a common refinement of both partitions. For $s \in \pi_n$, we denote the next partition point in π_n by s', and the previous partition point in π_m by $\lfloor s \rfloor_m$. Fix $t \geq 0$. Then

$$\begin{aligned} I_t^n - I_t^m &= \sum_{\substack{s\in\pi_n \\ s<t}} (M_s - M_{\lfloor s\rfloor_m})\,(M_{s'\wedge t} - M_s), \qquad \text{and hence} \\ \|I^n - I^m\|^2_{M^2([0,t])} &= E\left[(I_t^n - I_t^m)^2\right] \\ &= \sum_{\substack{s\in\pi_n \\ s<t}} E\left[(M_s - M_{\lfloor s\rfloor_m})^2\,(M_{s'\wedge t} - M_s)^2\right] \\ &\leq E\left[\delta_m^2\right]^{1/2}\, E\left[\left(\sum (\delta M_s)^2\right)^2\right]^{1/2}, \end{aligned} \tag{6.3.7}$$

where $\delta_m := \sup\{|M_s - M_r|^2 : |s - r| \leq \mathrm{mesh}(\pi_m)\}$. Here we have used that the non-diagonal summands cancel because M is a martingale.

Since M is bounded and continuous, dominated convergence shows that $E[\delta_m^2] \to 0$ as $m \to \infty$. Furthermore,

$$\begin{aligned} E\left[\left(\sum_s (\delta M_s)^2\right)^2\right] &= E\left[\sum_s (\delta M_s)^4\right] + E\left[\sum_{r,s:r<s} (\delta M_r)^2 (\delta M_s)^2\right] \\ &\leq 4C^2\, E\left[\sum_s (\delta M_s)^2\right] + 2\, E\left[\sum_r (\delta M_r)^2 E\left[\sum_{s>r} (\delta M_s)^2 | \mathcal{F}_r\right]\right] \\ &\leq 6\, C^2\, E[M_t^2 - M_0^2] \leq 6\, C^4 < \infty. \end{aligned}$$

Here we have used that by the martingale property,

$$
\begin{aligned}
E\left[\sum_s (\delta M_s)^2\right] &= E[M_t^2 - M_0^2] \;\leq\; C^2, \qquad \text{and} \\
E\left[\sum_{s>r} (\delta M_s)^2 | \mathcal{F}_r\right] &= E\left[M_t^2 - M_r^2 | \mathcal{F}_r\right] \;\leq\; C^2.
\end{aligned}
$$

By (6.3.7), $\|I^n - I^m\|^2_{M^2([0,t])} \to 0$ as $n, m \to \infty$. Hence $(I^n_s)_{s\in[0,t]}$ converges uniformly as $n \to \infty$ in the $L^2(P)$ sense. By (6.3.5), $(V^n_s)_{s\in[0,t]}$ converges uniformly as $n \to \infty$ in the $L^2(P)$ sense as well. Hence the limits $\int_0^\bullet M\,dM$ and $[M]$ exist, the stochastic integral is in $M^2_c([0,t])$, and the identity (6.3.6) holds.

It remains to extend the result from bounded martingales to local martingales. If M is a continuous local martingale then there exists a sequence of stopping times $T_k \uparrow \infty$ such that the stopped processes $(M_{T_k \wedge t})_{t\geq 0}$ are continuous bounded martingales. Hence the corresponding quadratic variations $[M_{T_k\wedge \bullet}]$ converge uniformly in the $L^2(P)$ sense for any finite t and k. Therefore, the approximations V^n_t for the quadratic variation of M converge uniformly in the $L^2(P)$ sense on each of the random intervals $[0, T_k \wedge t]$, and thus for any $\varepsilon, \delta > 0$,

$$
P\left[\sup_{s\leq t} |V^n_s - [M]_s| > \varepsilon\right] \;\leq\; P\left[t > T_k\right] + P\left[\sup_{s\leq T_k} |V^n_s - [M]_s| > \varepsilon\right] \;\leq\; \delta
$$

for k, n sufficiently large. □

Having shown the existence of the quadratic variation $[M]$ for continuous local martingales, we observe next that $[M]$ is always non-trivial if M is not constant:

Theorem 6.10 (Non-constant continuous martingales have non-trivial quadratic variation). *Suppose that (M_t) is a continuous local martingale. If $[M]_t = 0$ almost surely for some $t \geq 0$, then M is almost surely constant on the interval $[0,t]$.*

Proof. Again, we assume at first that M is a bounded martingale. Then the Itô integral $\int_0^\bullet M\,dM$ is a martingale as well. Therefore, by (6.3.6),

$$
\|M - M_0\|^2_{M^2([0,t])} \;=\; E[(M_t - M_0)^2] \;=\; E[M_t^2 - M_0^2] \;=\; E[\,[M]_t\,] \;=\; 0,
$$

i.e., $M_s = M_0$ for any $s \in [0, t]$. In the general case, the assertion follows once more by localization. $\square$

The theorem shows in particular that every local martingale with continuous finite variation paths is almost surely constant, i.e., *the Doob type decomposition of a continuous stochastic process into a local martingale and a continuous finite variation process starting at* 0 *is unique up to equivalence.* As a consequence we observe that the quadratic variation is the *unique* angle bracket process of M. In particular, up to modification on measure zero sets, $[M]$ does not depend on the chosen partition sequence (π_n):

Corollary 6.11 (Quadratic variation as unique angle bracket process). *Suppose that* (M_t) *is a continuous local martingale. Then* $[M]$ *is the up to equivalence unique continuous process of finite variation such that* $[M]_0 = 0$ *and* $M_t^2 - [M]_t$ *is a local martingale.*

Proof. By (6.3.6), $M_t^2 - [M]_t$ is a continuous local martingale. To prove uniqueness, suppose that (A_t) and $(\widetilde{A}_t)$ are continuous finite variation processes with $A_0 = \widetilde{A}_0 = 0$ such that both $M_t^2 - A_t$ and $M_t^2 - \widetilde{A}_t$ are local martingales. Then $A_t - \widetilde{A}_t$ is a continuous local martingale as well. Since the paths have finite variation, the quadratic variation of $A - \widetilde{A}$ vanishes. Hence almost surely, $A_t - \widetilde{A}_t = A_0 - \widetilde{A}_0 = 0$ for all t. $\square$

From continuous martingales to Brownian motion

A remarkable consequence of Itô's formula for martingales is that any continuous local martingale (M_t) (up to $T = \infty$) with quadratic variation given by $[M]_t = t$ for any $t \geq 0$ is a Brownian motion ! In fact, for $0 \leq s \leq t$ and $p \in \mathbb{R}$, Itô's formula yields

$$e^{ipM_t} - e^{ipM_s} = ip \int_s^t e^{ipM_r}\, dM_r - \frac{p^2}{2} \int_s^t e^{ipM_r}\, dr$$

where the stochastic integral can be identified as a local martingale. From this identity it is not difficult to conclude that the increment $M_t - M_s$ is conditionally independent of $\mathcal{F}_s^M$ with characteristic function

$$E[e^{ip(M_t-M_s)}] \;=\; e^{-p^2(t-s)/2} \qquad \text{for any } p \in \mathbb{R},$$

i.e., (M_t) has independent increments with distribution $M_t - M_s \sim N(0, t-s)$.

Theorem 6.12 (P. Lévy 1948). *A continuous local martingale $(M_t)_{t\in[0,\infty)}$ is a Brownian motion if and only if almost surely,*

$$[M]_t = t \qquad \textit{for any } t \geq 0.$$

Exercise (Lévy's characterization of Brownian motion). Extend the sketch above to a proof of Theorem 6.12.

Lévy's Theorem is the basis for many important developments in stochastic analysis including transformations and weak solutions for stochastic differential equations. An extension to the multi-dimensional case with a detailled proof, as well as several applications, are contained in Section 11.1 below.

One remarkable consequence of Lévy's characterization of Brownian motion is that every continuous local martingale can be represented as a time-changed Brownian motion (in general possibly on an extended probability space):

Exercise (Continuous local martingales as time-changed Brownian motions). Let $(M_t)_{t\in[0,\infty)}$ be a continuous local martingale, and assume for simplicity that $t \mapsto [M]_t$ is almost surely strictly increasing with $\lim_{t\to\infty}[M]_t = \infty$. Prove that there exists a Brownian motion $(B_t)_{t\in[0,\infty)}$ such that

$$M_t \;=\; B_{[M]_t} \qquad \text{for } t \in [0,\infty). \tag{6.3.8}$$

Hint: Set $B_a = M_{T_a}$ where $T_a = [M]^{-1}(a) = \inf\{t \geq 0 : [M]_t = a\}$, and verify by Lévy's characterization that B is a Brownian motion.

In a more general form, the representation of continuous local martingales as time-changed Brownian motions is due to Dambis and Dubins-Schwarz (1965), cf. [37] or Section 11.2 below for details. Remarkably, even before Itô, Wolfgang Doeblin, the son of Alfred Doeblin, had developed an alternative approach to stochastic calculus where stochastic integrals are defined as time changes of Brownian motion. Doeblin died when fighting as a French soldier at the German front in World War II, and his results that were hidden in a closed envelope at the Académie de Sciences have become known and been published only recently, more than fifty years after their discovery, cf. [Doeblin, Sur l'équation de Kolmogoroff, 1940/2000], [Yor: Présentation du pli cacheté, C.R.Acad.Sci. Paris 2000].

6.4 Multivariate and time-dependent Itô formula

We now extend Itô's formula to $\mathbb{R}^d$-valued functions and stochastic processes. Let $u \in (0,\infty]$ and suppose that $X : [0,u) \to D, X_t = (X_t^{(1)},\dots,X_t^{(d)})$, is a continuous function taking values in an open set $D \subseteq \mathbb{R}^d$. As before, we fix a sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. For a function $F \in C^2(D)$, we have similarly as in the one-dimensional case:

$$\begin{aligned} F(X_{s'\wedge t}) - F(X_s) &= \nabla F(X_s)\cdot(X_{s'\wedge t}-X_s) + \\ &\quad \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)(X^{(i)}_{s'\wedge t}-X^{(i)}_s)(X^{(j)}_{s'\wedge t}-X^{(j)}_s) + R_s^{(n)} \end{aligned} \tag{6.4.1}$$

for any $s \in \pi_n$ with $s < t$ where the dot denotes the Euclidean inner product $R_s^{(n)}$ is the remainder term in Taylor's formula. We would like to obtain a multivariate Itô formula by summing over $s \in \pi_n$ with $s < t$ and taking the limit as $n \to \infty$. A first problem that arises in this context is the identification of the limit of the sums

$$\sum_{\substack{s\in\pi_n \\ s<t}} g(X_s)\delta X_s^{(i)}\delta X_s^{(j)}$$

for a continuous function $g : D \to \mathbb{R}$ as $n \to \infty$.

Covariation

Suppose that $X, Y : [0, u) \to \mathbb{R}$ are continuous functions with continuous quadratic variations $[X]_t$ and $[Y]_t$ w.r.t. (π_n).

Definition. *The function*

$$[X, Y]_t \;=\; \lim_{n\to\infty} \sum_{s\in\pi_n} (X_{s'\wedge t} - X_{s\wedge t})(Y_{s'\wedge t} - Y_{s\wedge t}), \qquad t \in [0, u),$$

is called the ***covariation of*** X ***and*** Y ***w.r.t.*** (π_n) *if the limit exists.*

The covariation $[X, Y]_t$ is the bilinear form corresponding to the quadratic form $[X]_t$. In particular, $[X, X] = [X]$. Furthermore:

Lemma 6.13 (Polarization identity). *The covariation $[X, Y]_t$ exists and is a continuous function in t if and only if the quadratic variation $[X + Y]_t$ exists and is continuous respectively. In this case,*

$$[X, Y]_t \;=\; \frac{1}{2}([X + Y]_t - [X]_t - [Y]_t).$$

Proof. For $n \in \mathbb{N}$ we have

$$2 \sum_{s\in\pi_n} \delta X_s \delta Y_s \;=\; \sum_{s\in\pi_n} (\delta X_s + \delta Y_s)^2 - \sum_{s\in\pi_n} (\delta X_s)^2 - \sum_{s\in\pi_n} (\delta Y_s)^2.$$

The assertion follows as $n \to \infty$ because the limits $[X]_t$ and $[Y]_t$ of the last two terms are continuous functions by assumption. □

Remark. Note that by the polarization identity, the covariation $[X, Y]_t$ is the difference of two increasing functions, i.e., $t \mapsto [X, Y]_t$ has finite variation.

Example. (1). *Functions and processes of finite variation:* If Y has finite variation then $[X, Y]_t = 0$ for any $t \geq 0$. Indeed,

$$\left| \sum_{s\in\pi_n} \delta X_s \delta Y_s \right| \;\leq\; \sup_{s\in\pi_n} |\delta X_s| \cdot \sum_{s\in\pi_n} |\delta Y_s|$$

and the right hand side converges to 0 by uniform continuity of X on $[0,t]$. In particular, we obtain again

$$[X+Y] \;=\; [X]+[Y]+2[X,Y] \;=\; [X].$$

(2). *Independent Brownian motions:* If (B_t) and $(\widetilde{B}_t)$ are independent Brownian motions on a probability space $(\Omega,\mathcal{A},P)$ then for any given sequence (π_n),

$$[B,\widetilde{B}]_t \;=\; \lim_{n\to\infty}\sum_{s\in\pi_n}\delta B_s\delta\widetilde{B}_s \;=\; 0 \qquad \text{for any } t\geq 0$$

P-almost surely. For the proof note that $(B_t+\widetilde{B}_t)/\sqrt{2}$ is again a Brownian motion, whence

$$[B,\widetilde{B}]_t \;=\; [(B+\widetilde{B})/\sqrt{2}]_t-\frac{1}{2}[B]_t-\frac{1}{2}[\widetilde{B}]_t \;=\; t-\frac{t}{2}-\frac{t}{2} \;=\; 0 \qquad \text{almost surely.}$$

(3). *Itô processes:* If $I_t=\int_0^t G_s\,dB_s$ and $\mathcal{F}_t \;=\; \int\limits_0^t H_s\,dB_s$ with continuous adapted processes (G_t) and (H_t) and Brownian motions (B_t) and $(\widetilde{B}_t)$ then

$$[I,J]_t \;=\; 0 \qquad \text{if } B \text{ and } \widetilde{B} \text{ are independent, and} \tag{6.4.2}$$

$$[I,J]_t \;=\; \int\limits_0^t G_sH_s\,ds \qquad \text{if } B=\widetilde{B}, \tag{6.4.3}$$

cf. Theorem **??** below.

More generally, under appropriate assumptions on G,H,X and Y, the identity

$$[I,J]_t \;=\; \int\limits_0^t G_sH_s\,d[X,Y]_s$$

holds for Itô integrals $I_t=\int\limits_0^t G_s\,dX_s$ and $J_t=\int\limits_0^t H_s\,dY_s$, cf. e.g. Corollary **??**.

Itô to Stratonovich conversion

The covariation also occurs as the correction term in Itô compared to Stratonovich integrals:

Theorem 6.14. *If the Itô integral*

$$\int_0^t X_s\, Y_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} X_s \delta Y_s$$

and the covariation $[X,Y]_t$ *exists along a sequence* (π_n) *of partitions with* $\mathrm{mesh}(\pi_n) \to 0$ *then the corresponding backward Itô integral* $\int_0^t X_s\, \hat{d}Y_s$ *and the Stratonovich integral* $\int_0^t X_s \circ dY_s$ *also exist, and*

$$\begin{aligned} \int_0^t X_s\, \hat{d}Y_s &= \int_0^t X_s\, Y_s + [X,Y]_t, \qquad \textit{and} \\ \int_0^t X_s \circ dY_s &= \int_0^t X_s\, Y_s + \frac{1}{2}[X,Y]_t. \end{aligned}$$

Proof. This follows from the identities

$$\begin{aligned} \sum X_{S'\wedge t}\delta Y_s &= \sum X_s \delta Y_s + \sum \delta X_s \delta Y_s, \qquad \text{and} \\ \sum \frac{1}{2}(X_s + X_{s'\wedge t})\delta Y_s &= \sum X_s \delta Y_s + \frac{1}{2}\sum \delta X_s \delta Y_s. \end{aligned}$$

□

Itô’s formula in $\mathbb{R}^d$

By the polarization identity, if $[X]_t$, $[Y]_t$ and $[X+Y]_t$ exist and are continuous then $[X,Y]_t$ is a continuous function of finite variation.

Lemma 6.15. *Suppose that* X, Y *and* $X+Y$ *are continuous function on* $[0,u)$ *with continuous quadratic variations w.r.t.* (π_n). *Then*

$$\sum_{\substack{s\in\pi_n\\ s<t}} H_s (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s) \quad\longrightarrow\quad \int_0^t H_s\, d[X,Y]_s \qquad \textit{as } n\to\infty$$

for any continuous function $H : [0, u) \to \mathbb{R}$ *and any* $t \geq 0$.

Proof. The assertion follows from Lemma 6.3 by polarization. □

By Lemma 6.15, we can take the limit as $\mathrm{mesh}(\pi_n) \to 0$ in the equation derived by summing (6.4.2) over all $s \in \pi_n$ with $s < t$. In analogy to the one-dimensional case, this yields the following multivariate version of the pathwise Itô formula:

Theorem 6.16 (Multivariate Itô formula without probability). *Suppose that* $X : [0, u) \to D \subseteq \mathbb{R}^d$ *is a continuous function with continuous covariations* $[X^{(i)}, X^{(j)}]_t, 1 \leq i, j \leq d$, *w.r.t.* (π_n). *Then for any* $F \in C^2(D)$ *and* $t \in [0, u)$,

$$F(X_t) \;=\; F(X_0) + \int_0^t \nabla F(X_s) \cdot dX_s + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(X_s)\, d[X^{(i)}, X^{(j)}]_s,$$

where the Itô integral is the limit of Riemann sums along (π_n):

$$\int_0^t \nabla F(X_s) \cdot dX_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \nabla F(X_s) \cdot (X_{s'\wedge t} - X_s). \tag{6.4.4}$$

The details of the proof are similar to the one-dimensional case and left as an exercise to the reader. Note that the theorem shows in particular that the Itô integral in (6.4.4) is independent of the sequence (π_n) if the same holds for the covariations $[X^{(i)}, X^{(j)}]$.

Remark (Existence of pathwise Itô integrals). The theorem implies the existence of the Itô integral $\int_0^t b(X_s) \cdot dX_s$ if $b = \nabla F$ is the gradient of a C^2 function $F : D \subseteq \mathbb{R}^d \to \mathbb{R}$. In contrast to the one-dimensional case, not every C^1 vector field $b : D \to \mathbb{R}^d$ is a gradient. Therefore, for $d \geq 2$ we do **not** obtain existence of $\int_0^t b(X_s) \cdot dX_s$ for any $b \in C^1(D, \mathbb{R}^d)$ from Itô's formula. In particular, *we do not know in general* if the integrals $\int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^{(i)}, 1 \leq i \leq d$, exists and if

$$\int_0^t \nabla F(X_s) \cdot dX_s \;=\; \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x_i}(X_s)\, dX_s^{(i)}.$$

If (X_t) is a Brownian motion this is almost surely the case by the existence proof for Itô integrals w.r.t. Brownian motion from Section 5.

Example (Itô's formula for Brownian motion in $\mathbb{R}^d$). Suppose that $B_t = (B_t^{(1)}, \dots, B_t^{(d)})$ is a d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$. Then the component processes $B_t^{(1)}, \dots, B_t^{(d)}$ are independent one-dimensional Brownian motions. Hence for a given sequence of partitions (π_n) with $\mathrm{mesh}(\pi_n) \to 0$, the covariations $[B^{(i)}, B^{(j)}]$, $1 \le i, j \le d$, exists almost surely by Theorem 6.6 and the example above, and

$$[B^{(i)}, B^{(j)}] \;=\; t \cdot \delta_{ij} \qquad \forall t \ge 0$$

P-almost surely. Therefore, we can apply Itô's formula to almost every trajectory. For an open subset $D \subseteq \mathbb{R}^d$ and a function $F \in C^2(D)$ we obtain:

$$F(B_t) \;=\; F(B_0) + \int_0^t \nabla F(B_s) \cdot dB_s + \frac{1}{2} \int_0^t \Delta F(B_s)\, ds \qquad \forall t < T_{D^C} \quad P\text{-a.s.} \tag{6.4.5}$$

where $T_{D^C} := \inf\{t \ge 0 \;:\; B_t \notin D\}$ denotes the first exit time from D. As in the one-dimensional case, (6.4.5) yields a decomposition of the process $F(B_t)$ into a continuous local martingale and a continuous process of finite variation, cf. Section **??** for applications.

Product rule, integration by parts

As a special case of the multivariate Itô formula, we obtain the following integration by parts identity for Itô integrals:

Corollary 6.17. *Suppose that $X, Y : [0, u) \to \mathbb{R}$ are continuous functions with continuous quadratic variations $[X]$ and $[Y]$, and continuous covariation $[X, Y]$. Then*

$$X_t Y_t - X_0 Y_0 \;=\; \int_0^t \begin{pmatrix} Y_s \\ X_s \end{pmatrix} \cdot d\begin{pmatrix} X_s Y_s \end{pmatrix} + [X, Y]_t \qquad \textit{for any } t \in [0, u). \tag{6.4.6}$$

If one, or, equivalently, both of the Itô integrals $\int_0^t Y_s\,dX_s$ *and* $\int_0^t X_s\,dY_s$ *exist then (6.4.6) yields*

$$X_tY_t - X_0Y_0 \;=\; \int_0^t Y_s\,dX_s + \int_0^t X_s\,dY_s + [X,Y]_t. \tag{6.4.7}$$

Proof. The identity (6.4.6) follows by applying Itô's formula in $\mathbb{R}^2$ to the process (X_t, Y_t) and the function $F(x,y) = xy$. If one of the integrals $\int_0^t Y\,dX$ or $\int_0^t X\,dY$ exists, then the other exists as well, and

$$\int_0^t \begin{pmatrix} Y_s \\ X_s \end{pmatrix} \cdot d\begin{pmatrix} X_s \\ Y_s \end{pmatrix} \quad = \quad \int_0^t Y_s\,dX_s + \int_0^t X_s\,dY_s.$$

□

As it stands, (6.4.7) is an integration by parts formula for Itô integrals which involves the correction term $[X,Y]_t$. In differential notation, it is a product rule for Itô differentials:

$$d(XY) \;=\; X\,dY + Y\,dX + [X,Y].$$

Again, in Stratonovich calculus a corresponding product rule holds without the correction term $[X,Y]$:

$$\circ d(XY) \;=\; X\,\circ dY + Y\,\circ dX.$$

Remark / Warning (Existence of $\int$ X dY, Lévy area). Under the conditions of the theorem, the Itô integrals $\int_0^t X\,dY$ and $\int_0^t Y\,dX$ do not necessarily exist! The following statements are equivalent:

(1). The Itô integral $\int_0^t Y_s\,dX_s$ exists (along (π_n)).

(2). The Itô integral $\int_0^t X_s\,dY_s$ exists.

(3). The **Lévy area** $A_t(X,Y)$ defined by

$$A_t(X,Y) \;=\; \int_0^t (Y\,dX - X\,dY) \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} (Y_s\Delta X_s - X_s\Delta Y_s)$$

exists.

Hence, if the Lévy area $A_t(X,Y)$ is given, the stochastic integrals $\int X\,dY$ and $\int Y\,dX$ can be constructed pathwise. Pushing these ideas further leads to the rough paths theory developed by T. Lyons and others, cf. [Lyons, St. Flour], [Friz: Rough paths theory].

Example (Integrating finite variation processes w.r.t. Brownian motion). If (H_t) is an adapted process with continuous sample paths of finite variation and (B_t) is a one-dimensional Brownian motion then $[H,B] = 0$, and hence

$$H_tB_t - H_0B_0 \;=\; \int_0^t H_s\,dB_s + \int_0^t B_s\,dH_s.$$

This integration by parts identity can be used as an alternative definition of the stochastic integral $\int_0^t H\,dB$ for integrands of finite variation, which can then again be extended to general integrands in $\mathcal{L}^2_a(0,t)$ by the Itô isometry.

Time-dependent Itô formula

The multi-dimensional Itô formula can be applied to functions that depend explicitly on the time variable t or on the quadratic variation $[X]_t$. For this purpose we simply add t or $[X]_t$ respectively as an additional component to the function, i.e., we apply the multi-dimensional Itô formula to $Y_t = (t, X_t)$ or $Y_t = (t, [X]_t)$ respectively.

Theorem 6.18. *Suppose that $X : [0,u) \to \mathbb{R}^d$ is a continuous function with continuous covariations $[X^{(i)}, X^{(j)}]_t$, along (π_n), and let $F \in C^2(A([0,u))\times\mathbb{R}^d)$. If $A : [0,u) \to \mathbb{R}$ is a continuous function of finite variation then the integral*

$$\int_0^t \nabla_x F(A_s, X_s)\cdot dX_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \nabla_x F(A_s,X_s)\cdot (X_{s'\wedge t} - X_s)$$

exists, and the Itô formula

$$F(A_t, X_t) \;=\; F(0, X_0) + \int_0^t \nabla_x F(A_s, X_s) \cdot dX_s + \int_0^t \frac{\partial F}{\partial a}(A_s, X_s)\, dA_s + \tag{6.4.8}$$

$$\frac{1}{2}\sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(A_s, X_s)\, d[X^{(i)}, X^{(j)}]_s \tag{6.4.9}$$

holds for any $t \geq 0$. Here $\partial F/\partial a$ denotes the derivative of $F(a, x)$ w.r.t. the first component, and $\nabla_x F$ and $\partial^2 F/\partial x_i \partial x_j$ are the gradient and the second partial derivatives w.r.t. the other components. The most important application of the theorem is for $A_t = t$. Here we obtain the time-dependent Itô formula

$$dF(t, X_t) \;=\; \nabla_x F(t, X_t) \cdot dX_t + \frac{\partial F}{\partial t}(t, X_t)\, dt + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 F}{\partial x_i \partial x_j}(t, X_t)\, d[X^{(i)}, X^{(j)}]_t. \tag{6.4.10}$$

Similarly, if $d = 1$ and $A_t = [X]_t$ then we obtain

$$dF([X]_t, X_t) \;=\; \frac{\partial F}{\partial t}([X]_t, X_t)\, dt + \left(\frac{\partial F}{\partial a} + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right)([X]_t, X_t)\, d[X]_t. \tag{6.4.11}$$

If $(X)_t$ is a Brownian motion and $d = 1$ then both formulas coincide.

Proof. Let $Y_t = (Y_t^{(0)}, Y_t^{(1)}, \ldots, Y_t^{(d)}) := (A_t, X_t)$. Then $[Y^{(0)}, Y^{(i)}]_t = 0$ for any $t \geq 0$ and $0 \leq i \leq d$ because $Y_t^{(0)} = A_t$ has finite variation. Therefore, by Itô's formula in $\mathbb{R}^{d+1}$,

$$F(A_t, X_t) \;=\; F(A_0, X_0) + I_t + \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 F}{\partial x_i \partial x_j}(A_s, X_s)\, d[X^{(i)}, X^{(j)}]_s$$

where

$$\begin{aligned} I_t \;&=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \nabla^{\mathbb{R}^{d+1}} F(A_s, X_s) \cdot \begin{pmatrix} A_{s'\wedge t} - A_s \\ X_{s'\wedge t} - X_s \end{pmatrix} \\ &=\; \lim_{n\to\infty} \left(\sum \frac{\partial F}{\partial a}(A_s, X_s)(A_{s'\wedge t} - A_s) + \sum \nabla_x F(A_s, X_s) \cdot (X_{s'\wedge t} - X_s) \right). \end{aligned}$$

The first sum on the right hand side converges to the Stieltjes integral $\int_0^t \frac{\partial F}{\partial a}(A_s, X_s)\, dA_s$ as $n \to \infty$. Hence, the second sum also converges, and we obtain (6.4.7) in the limit as $n \to \infty$. □

Note that if $h(t,x)$ is a solution of the dual heat equation

$$\frac{\partial h}{\partial t} + \frac{1}{2}\frac{\partial^2 h}{\partial x^2} \;=\; 0 \qquad \text{for } t \geq 0, x \in \mathbb{R}, \tag{6.4.12}$$

then by (6.4.11),

$$h([X]_t, X_t) \;=\; h(0, X_0) + \int_0^t \frac{\partial h}{\partial x}([X]_s, X_s)\, dX_s.$$

In particular, if (X_t) is a Brownian motion, or more generally a local martingale, then $h([X]_t, X_t)$ is also a local martingale. The next example considers two situations where this is particular interesting:

Example. (1). *Itô exponentials:* For any $\alpha \in \mathbb{R}$, the function

$$h(t,x) \;=\; \exp(\alpha x - \alpha^2 t/2)$$

satisfies (6.4.12) and $\partial h/\partial x = \alpha h$. Hence the function

$$Z_t^{(\alpha)} \;:=\; \exp\left(\alpha X_t - \frac{1}{2}\alpha^2 [X]_t\right)$$

is a solution of the Itô differential equation

$$dZ_t^{(\alpha)} \;=\; \alpha Z_t^{(\alpha)}\, dX_t$$

with initial condition $Z_0^{(\alpha)} = 1$. This shows that in Itô calculus, the functions $Z_t^{(\alpha)}$ are the correct replacements for the exponential functions. The additional factor $\exp(-\alpha^2 [X]_t/2)$ should be thought of as an appropriate renormalization in the continuous time limit.

For a Brownian motion (X_t), we obtain the exponential martingales as generalized exponentials.

(2). *Hermite polynomials:* For $n = 0, 1, 2, \ldots$, the Hermite polynomials

$$h_n(t,x) \;=\; \frac{\partial^n}{\partial \alpha^n} \exp(\alpha x - \frac{1}{2}\alpha^2 t)|_{\alpha=0}$$

also satisfy (6.4.12). The first Hermite polynomials are $1, x, x^2 - t, x^3 - 3tx, \ldots$. Note also that

$$\exp(\alpha x - \alpha^2 t/2) \;=\; \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} h_n(t,x)$$

by Taylor's theorem. Moreover, the following properties can be easily verified:

$$h_n(1,x) \;=\; e^{x^2/2}(-1)^n \frac{d^n}{dx^n} e^{-x^2/2} \qquad \text{for any } x \in \mathbb{R}, \tag{6.4.13}$$

$$h_n(t,x) \;=\; t^{n/2} h_n(1, x/\sqrt{t}) \qquad \text{for any } t \geq 0, x \in \mathbb{R}, \tag{6.4.14}$$

$$\frac{\partial h_n}{\partial x} \;=\; n h_{n-1}, \qquad \frac{\partial h_n}{\partial t} + \frac{1}{2}\frac{\partial^2 h_n}{\partial x^2} \;=\; 0. \tag{6.4.15}$$

For example, (6.4.13) holds since

$$\exp(\alpha x - \alpha^2/2) \;=\; \exp(-(x-a)^2/2)\exp(x^2/2)$$

yields

$$h_N(1,x) \;=\; \exp(x^2/2)(-1)^n \left.\frac{d^n}{d\beta^n}\exp(-\beta^2/2)\right|_{\beta=x},$$

and (6.4.14) follows from

$$\begin{aligned} \exp(\alpha x - \alpha^2 t/2) \;&=\; \exp(\alpha\sqrt{t}\cdot(x/\sqrt{t}) - (\alpha\sqrt{t})^2/2) \\ &=\; \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} t^{n/2} h_n(1, x/\sqrt{t}). \end{aligned}$$

By (6.4.13) and (6.4.14), h_n is a polynomial of degree n. For any $n \geq 0$, the function

$$H_t^{(n)} \;:=\; h_n([X]_t, X_t)$$

is a solution of the Itô equation

$$dH_t^{(n)} \;=\; n H_t^{(n-1)}\, dX_t. \tag{6.4.16}$$

Therefore, the Hermite polynomials are appropriate replacements for the ordinary monomials x^n in Itô calculus. If $X_0 = 0$ then $H_0^{(n)} = 0$ for $n \geq 1$, and we obtain inductively

$$H_t^{(0)} = 1, \quad H_t^{(1)} = \int_0^t dX_s, \quad H_t^{(2)} = \int H_s^{(1)}\, dX_s = \int_0^t \int_0^s dX_r\, dX_s,$$

and so on.

Corollary 6.19 (Itô 1951). *If $X : [0,u) \to \mathbb{R}$ is continuous with $X_0 = 0$ and continuous quadratic variation then for $t \in [0,u)$,*

$$\int_0^t \int_0^{s_n} \cdots \int_0^{s_2} dX_{s_1} \cdots dX_{s_{n-1}}\, dX_{s_n} = \frac{1}{n!} h_n([X]_t, X_t).$$

Proof. The equation follows from (6.4.16) by induction on n. □

Iterated Itô integrals occur naturally in Taylor expansions of Itô calculus. Therefore, the explicit expression from the corollary is valuable for numerical methods for stochastic differential equations, cf. Section **??** below.

Chapter 7

Brownian Motion and Partial Differential Equations

The stationary and time-dependent Itô formula enable us to work out the connection of Brownian motion to several partial differential equations involving the Laplace operator in detail. One of the many consequences is the evaluation of probabilities and expectation values for Brownian motion by p.d.e. methods. More generally, Itô's formula establishes a link between stochastic processes and analysis that is extremely fruitful in both directions.

Suppose that (B_t) is a d-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$ such that *every* sample path $t \mapsto B_t(\omega)$ is continuous. We first note that Itô's formula shows that Brownian motion solves the martingale problem for the operator $\mathscr{L} = \frac{1}{2}\Delta$ in the following sense:

Corollary 7.1 (Time-dependent martingale problem). *The process*

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

is a continuous $(\mathcal{F}_t^B)$ martingale for any C^2 function $F : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ with bounded first derivatives. Moreover, M^F is a continuous local martingale up to $T_{D^C} = \inf\{t \geq 0 \,:\, B_t \notin D\}$ for any $F \in C^2([0,\infty) \times D)$, $D \subseteq \mathbb{R}^d$ open.

Proof. By the continuity assumptions one easily verifies that M^F is $(\mathcal{F}_t^B)$ adapted. Moreover, by the time-dependent Itô formula (6.4.10),

$$M_t^F = \int_0^t \nabla_x F(s, B_s) \cdot dB_s \qquad \text{for } t < T_{D^C},$$

which implies the claim. □

Choosing a function F that does not explicitly depend on t, we obtain in particular that

$$M_t^F = F(B_t) - F(B_0) - \int_0^t \frac{1}{2}\Delta F(B_s)\, ds$$

is a martingale for any $f \in C_b^2(\mathbb{R}^d)$, and a local martingale up to T_{D^C} for any $F \in C^2(D)$.

7.1 Dirichlet problem, recurrence and transience

As a first consequence of Corollary 7.1 we can now complete the proof of the stochastic representation for solutions of the Dirichlet problem that has been already mentioned in Section 3.2 above. By solving the Dirichlet problem for balls explicitly, we will then study recurrence, transience and polar sets for multi-dimensional Brownian motion.

The Dirichlet problem revisited

Suppose that $h \in C^2(D) \cap C(\overline{D})$ is a solution of the Dirichlet problem

$$\Delta h = 0 \quad \text{on } D, \quad h = f \quad \text{on } \partial D, \tag{7.1.1}$$

for a bounded open set $D \subset \mathbb{R}^d$ and a continuous function $f : \partial D \to \mathbb{R}$. If (B_t) is under P_x a continuous Brownian motion with $B_0 = x$ P_x-almost surely, then by Corollary 7.1, the process $h(B_t)$ is a local $(\mathcal{F}_t^B)$ martingale up to T_{D^C}. By applying the optional stopping theorem with a localizing sequence of bounded stopping times $S_n \nearrow T_{D^C}$, we obtain

$$h(x) \;=\; E_x[h(B_0)] \;=\; E_x[h(B_{S_n})] \qquad \text{for any } n \in \mathbb{N}.$$

Since $P_x[T_{D^C} < \infty] = 1$ and h is bounded on $\overline{D}$, dominated convergence then yields the stochastic representation

$$h(x) \;=\; E_x[h(B_{T_{D^C}})] \;=\; E_x[f(B_{T_{D^C}})] \qquad \text{for any } x \in \mathbb{R}^d.$$

We thus have shown:

Theorem 7.2 (Stochastic representation for solutions of the Dirichlet problem). *Suppose that D is a bounded open subset of $\mathbb{R}^d$, f is a continuous function on the boundary ∂D, and $h \in C^2(D) \cap C(\overline{D})$ is a solution of the Dirichlet problem (7.1.1). Then*

$$h(x) \;=\; E_x[f(B_T)] \quad \textit{for any } x \in \overline{D}.$$

We will generalize this result substantially in Theorem 7.5 below. Before, we apply the Dirichlet problem to study recurrence and transience of Brownian motions:

Recurrence and transience of Brownian motion in $\mathbb{R}^d$

Let (B_t) be a d-dimensional Brownian motion on $(\Omega, \mathcal{A}, P)$ with initial value $B_0 = x_0, x_0 \neq 0$. For $r \geq 0$ let

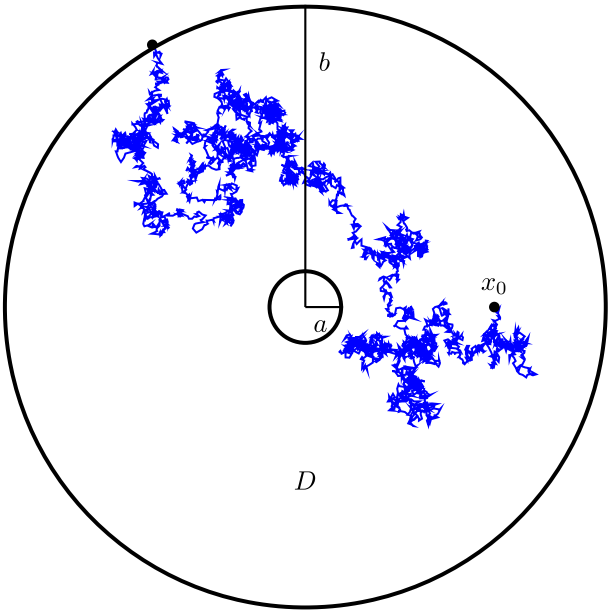
$$T_r \;=\; \inf\{t > 0 \;:\; |B_t| = r\}.$$

We now compute the probabilities $P[T_a < T_b]$ for $a < |x_0| < b$. Note that this is a multi-dimensional analogue of the *classical ruin problem*. To compute the probability for given a, b we consider the domain

$$D \;=\; \{x \in \mathbb{R}^d \;:\; a < |x| < b\}.$$

For $b < \infty$, the first exit time T_{D^C} is almost surely finite,

$$T_{D^C} \;=\; \min(T_a, T_b), \qquad \text{and} \qquad P[T_a < T_b] \;=\; P[|B_{T_{D^C}}| = a].$$



Suppose that $h \in C(\overline{U}) \cap C^2(U)$ is a solution of the Dirichlet problem

$$\Delta h(x) = 0 \qquad \text{for all } x \in D, \qquad h(x) = \begin{cases} 1 & \text{if } |x| = a, \\ 0 & \text{if } |x| = b. \end{cases} \tag{7.1.2}$$

Then $h(B_t)$ is a bounded local martingale up to T_{D^C} and optional stopping yields

$$P[T_a < T_b] = E[h(B_{T_{D^C}})] = h(x_0). \tag{7.1.3}$$

By rotational symmetry, the solution of the Dirichlet problem (7.1.2) can be computed explicitly. The Ansatz $h(x) = f(|x|)$ leads us to the boundary value problem

$$\frac{d^2 f}{dr^2}(|x|) + \frac{d-1}{|x|}\frac{df}{dr}(|x|) = 0, \qquad f(a) = 1, f(b) = 0,$$

for a second order ordinary differential equation. Solutions of the o.d.e. are linear combinations of the constant function 1 and the function

$$\phi(s) := \begin{cases} s & \text{for } d = 1, \\ \log s & \text{for } d = 2, \\ s^{2-d} & \text{for } d \geq 3. \end{cases}$$

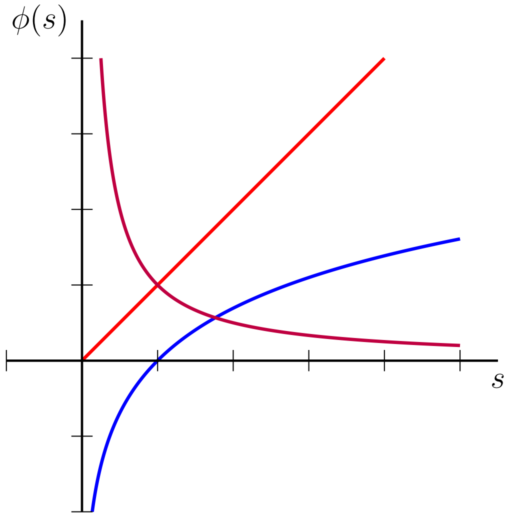


Figure 7.1: The function $\phi(s)$ for different values of d: red ($d = 1$), blue ($d = 2$) and purple ($d = 3$)

Hence, the unique solution f with boundary conditions $f(a) = 1$ and $f(b) = 0$ is

$$f(r) \;=\; \frac{\phi(b) - \phi(r)}{\phi(b) - \phi(a)}.$$

Summarizing, we have shown:

Theorem 7.3 (Ruin problem in $\mathbb{R}^d$). *For $a, b > 0$ with $a < |x_0| < b$,*

$$P[T_a < T_b] \;=\; \frac{\phi(b) - \phi(|x_0|)}{\phi(b) - \phi(a)}, \qquad \textit{and}$$

$$P[T_b < \infty] \;=\; \begin{cases} 1 & \textit{for } d \leq 2 \\ (a/|x_0|)^{d-2} & \textit{for } d > 2. \end{cases}$$

Proof. The first equation follows by 6.4.12. Moreover,

$$P[T_a < \infty] \;=\; \lim_{b \to \infty} P[T_a < T_b] \;=\; \begin{cases} 1 & \text{for } d \leq 2 \\ \phi(|x_0|)/\phi(a) & \text{for } d \geq 3. \end{cases}$$

□

Corollary 7.4. *For a Brownian motion in $\mathbb{R}^d$ the following statements hold for any initial value $x_0 \in \mathbb{R}^d$:*

*(1). If $d \leq 2$ then every non-empty ball $D \subseteq \mathbb{R}^d$ is **recurrent**, i.e., the last visit time of D is almost surely infinite:*

$$L_d \;=\; \sup\{t \geq 0 \,:\, B_t \in D\} \;=\; \infty \qquad P\text{-a.s.}$$

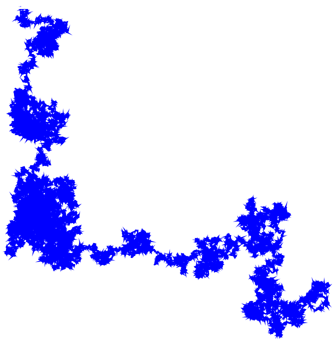
*(2). If $d \geq 3$ then every ball D is **transient**, i.e.,*

$$L_d \;<\; \infty \qquad P\text{-a.s.}$$

*(3). If $d \geq 2$ then every point $x \in \mathbb{R}^d$ is **polar**, i.e.,*

$$P[\,\exists\, t > 0 \,:\, B_t = x] \;=\; 0.$$

We point out that the last statement holds even if the starting point x_0 coincides with x. The first statement implies that a typical Brownian sample path is dense in $\mathbb{R}^2$, whereas by the second statement, $\lim_{t\to\infty} |B_t| = \infty$ almost surely for $d \geq 3$.



Proof.

(1),(2) The first two statements follow from Theorem 7.3 and the Markov property.

(3). For the third statement we assume w.l.o.g. $x = 0$. If $x_0 \neq 0$ then

$$P[T_0 < \infty] \;=\; \lim_{b\to\infty} P[T_0 < T_b]$$

for any $a > 0$. By Theorem 7.3,

$$P[T_0 < T_b] \;\leq\; \inf_{a>0} P[T_a < T_b] \;=\; 0 \qquad \text{for } d \geq 2,$$

whence $T_0 = \infty$ almost surely. If $x_0 = 0$ then by the Markov property,

$$P[\,\exists\, t > \varepsilon \;:\; B_t = 0] \;=\; E[P_{B_\varepsilon}[T_0 < \infty]] \;=\; 0$$

for any $\varepsilon > 0$. thus we again obtain

$$P[T_0 < \infty] \;=\; \lim_{\varepsilon\searrow 0} P[\,\exists\, t > \varepsilon \;:\; B_t = 0] \;=\; 0.$$

□

Remark (Polarity of linear subspaces). For $d \geq 2$, any $(d-2)$ dimensional subspace $V \subseteq \mathbb{R}^d$ is polar for Brownian motion. For the proof note that the orthogonal projection of a one-dimensional Brownian motion onto the orthogonal complement $V^\perp$ is a 2-dimensional Brownian motion.

7.2 Boundary value problems, exit and occupation times

The connection of Brownian motion to boundary value problems for partial differential equations involving the Laplace operator can be extended substantially:

The stationary Feynman-Kac-Poisson formula

Suppose that $f : \partial D \to \mathbb{R}, V : D \to \mathbb{R}$ and $g : D \to [0,\infty)$ are continuous functions defined on an open bounded domain $D \subset \mathbb{R}^d$, or on its boundary respectively. We assume that under P_x, (B_t) is Brownian motion with $P_x[B_0 = x] = 1$, and that

$$E_x\left[\exp \int_0^T V^-(B_s)\, ds\right] \;<\; \infty \qquad \text{for any } x \in D, \tag{7.2.1}$$

where $T = T_{D^C}$ is the first exit time from D.
Note that (7.2.1) always holds if V is non-negative.

Theorem 7.5. *If $u \in C^2(D) \cap C(\overline{D})$ is a solution of the boundary problem*

$$\frac{1}{2}\Delta u(x) = V(x)u(x) - g(x) \qquad \text{for } x \in D \tag{7.2.2}$$

$$u(x) = f(x) \qquad \text{for } x \in \partial D, \tag{7.2.3}$$

and (7.2.1) holds then

$$\begin{aligned} u(x) &= E_x\left[\exp\left(-\int_0^T V(B_s)\,ds\right)\cdot f(B_T)\right] + \\ &\quad E_x\left[\int_0^T \exp\left(-\int_0^t V(B_s)\,ds\right)\cdot g(B_t)\,dt\right] \end{aligned} \tag{7.2.4}$$

for any $x \in D$.

Remark. Note that we *assume* the existence of a smooth solution of the boundary value problem (7.2.2). Proving that the function u defined by (7.2.4) is a solution of the b.v.p. without assuming existence is much more demanding.

Proof. By continuity of V and (B_s), the sample paths of the process

$$A_t = \int_0^t V(B_s)\,ds$$

are C^1 and hence of finite variation for $t < T$. Let

$$X_t = e^{-A_t}u(B_t), \qquad t < T.$$

Applying Itô's formula with $F(a,b) = e^{-a}u(b)$ yields the decomposition

$$\begin{aligned} dX_t &= e^{-A_t}\nabla u(B_t)\cdot dB_t - e^{-A_t}u(B_t)\,dA_t + \frac{1}{2}e^{-A_t}\Delta u(B_t)\,dt \\ &= e^{-A_t}\nabla u(B_t)\cdot dB_t + e^{-A_t}\left(\frac{1}{2}\Delta u - V\cdot u\right)(B_t)\,dt \end{aligned}$$

of X_t into a local martingale up to time T and an absolutely continuous part. Since u is a solution of (7.2.2), we have $\frac{1}{2}\Delta u - Vu = -g$ on D. By applying the optional stopping theorem with a localizing sequence $T_n \nearrow T$ of stopping times, we obtain the representation

$$\begin{aligned} u(x) = E_x[X_0] &= E_x[X_{T_n}] + E_x\left[\int_0^{T_n} e^{-A_t}g(B_t)\,dt\right] \\ &= E_x[e^{-A_{T_n}}u(B_{T_n})] + E_x\left[\int_0^{T_n} e^{-A_t}g(B_t)\,dt\right] \end{aligned}$$

for $x \in D$. The assertion (7.2.4) now follows provided we can interchange the limit as $n \to \infty$ and the expectation values. For the second expectation on the right hand side this is possible by the monotone convergence theorem, because $g \geq 0$. For the first expectation value, we can apply the dominated convergence theorem, because

$$\left|e^{-A_{T_n}}u(B_{T_n})\right| \leq \exp\left(\int_0^T V^-(B_s)\,ds\right)\cdot\sup_{y\in\overline{D}}|u(y)| \qquad \forall n \in \mathbb{N},$$

and the majorant is integrable w.r.t. each P_x by Assumption 7.2.1. $\square$

Remark (Extension to diffusion processes). A corresponding result holds under appropriate assumptions if the Brownian motion (B_t) is replaced by a diffusion process (X_t) solving a stochastic differential equation of the type $dX_t = \sigma(X_t)\,dB_t + b(X_t)\,dt$, and the operator $\frac{1}{2}\Delta$ in (7.2.2) is replaced by the generator

$$\mathscr{L} = \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x)\cdot\nabla, \quad a(x) = \sigma(x)\sigma(x)^\top,$$

of the diffusion process, cf. **??**. The theorem hence establishes a general connection between Itô diffusions and boundary value problems for linear second order elliptic partial differential equations.

By Theorem 7.5 we can compute many interesting expectation values for Brownian motion by solving appropriate p.d.e. We now consider various corresponding applications.

Let us first recall the Dirichlet problem where $V \equiv 0$ and $g \equiv 0$. In this case, $u(x) = E_x[f(B_t)]$. We have already pointed out in the last section that this can be used to compute exit distributions and to study recurrence, transience and polarity of linear subspaces for Brownian motion in $\mathbb{R}^d$. A second interesting case of Theorem 7.5 is the stochastic representation for solutions of the Poisson equation:

Poisson problem and mean exit time

If V and f vanish in Theorem 7.3, the boundary value problem (7.2.2) reduces to the boundary value problem

$$\frac{1}{2}\Delta u = -g \quad \text{on } D, \quad u = 0 \quad \text{on } D,$$

for the Poisson equation. The solution has the stochastic representation

$$u(x) = E_x\left[\int_0^T g(B_t)\,dt\right], \qquad x \in D, \tag{7.2.5}$$

which can be interpreted as an average cost accumulated by the Brownian path before exit from the domain D. In particular, choosing $g \equiv 1$, we can compute the mean exit time

$$u(x) = E_x[T]$$

from D for Brownian motion starting at x by solving the corresponding Poisson problem.

Example. If $D = \{x \in \mathbb{R}^d : |x| < r\}$ is a ball around 0 of radius $r > 0$, then the solution $u(x)$ of the Poisson problem

$$\frac{1}{2}\Delta u(x) = \begin{cases} -1 & \text{for } |x| < r \\ 0 & \text{for } |x| = r \end{cases}$$

can be computed explicitly. We obtain

$$E_x[T] = u(x) = \frac{r^2 - |x|^2}{d} \qquad \text{for any } x \in D.$$

Occupation time density and Green function

If (B_t) is a Brownian motion in $\mathbb{R}^d$ then the corresponding Brownian motion with absorption at the first exit time from the domain D is the Markov process (X_t) with state space $D \cup \{\Delta\}$ defined by

$$X_t = \begin{cases} B_t & \text{for } t < T \\ \Delta & \text{for } t \geq T \end{cases},$$

where Δ is an extra state added to the state space. By setting $g(\Delta) = 0$, the stochastic representation (7.2.5) for a solution of the Poisson problem can be written in the form

$$u(x) = E_x\left[\int_0^\infty g(X_t)\,dt\right] = \int_0^\infty (p_t^D g)(x)\,dt, \tag{7.2.6}$$

where

$$p_t^D(x, A) = P_x[X_t \in A], \qquad A \subseteq \mathbb{R}^d measurable,$$

is the transition function for the absorbed process (X_t). Note that for $A \subset \mathbb{R}^d$,

$$p_t^D(x, A) = P_x[B_t \in A \text{ and } t < T] \leq p_t(x, A) \tag{7.2.7}$$

where p_t is the transition function of Brownian motion on $\mathbb{R}^d$. For $t > 0$ and $x \in \mathbb{R}^d$, the transition function $p_t(x, \bullet)$ of Brownian motion is absolutely continuous. Therefore, by (7.2.7), the sub-probability measure $p_t^D(x, \bullet)$ restricted to $\mathbb{R}^d$ is also absolutely continuous with non-negative density

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

The function p_t^D is called the **heat kernel on the domain** D w.r.t. absorption on the boundary. Note that

$$G^D(x, y) = \int_0^\infty p_t^D(x, y)\,dt$$

is an **occupation time density**, i.e., it measures the average time time a Brownian motion started in x spends in a small neighbourhood of y before it exits from the Domain

D. By (7.2.6), a solution u of the Poisson problem $\frac{1}{2}\Delta u = -g$ on D, $u = 0$ on ∂D, can be represented as

$$u(x) = \int_D G^D(x,y)g(y)\,dy \qquad \text{for } x \in D.$$

This shows that the occupation time density $G^D(x,y)$ is the **Green function** (i.e., the fundamental solution of the Poisson equation) **for the operator $\frac{1}{2}$ with Dirichlet boundary conditions on the domain D.**

Note that although for domains with irregular boundary, the Green's function might not exist in the classical sense, the function $G^D(x,y)$ is always well-defined!

Stationary Feynman-Kac formula and exit time distributions

Next, we consider the case where g vanishes and $f \equiv 1$ in Theorem 7.5. Then the boundary value problem (7.2.4) takes the form

$$\frac{1}{2}\Delta u = Vu \quad \text{on } D, \quad u = 1 \quad \text{on } \partial D. \tag{7.2.8}$$

The p.d.e. $\frac{1}{2}\Delta u = Vu$ is a stationary Schrödinger equation. We will comment on the relation between the Feynman-Kac formula and Feynman's path integral formulation of quantum mechanics below. For the moment, we only note that for the solution of (**??**), the stochastic representation

$$u(x) = E_x\left[\exp\left(-\int_0^T V(B_t)\,dt\right)\right]$$

holds for $x \in D$.

As an application, we can, at least in principle, compute the full distribution of the exit time T. In fact, choosing $V \equiv \alpha$ for some constant $\alpha > 0$, the corresponding solution u_α of (7.2.8) yields the Laplace transform

$$u_\alpha(x) = E_x[e^{-\alpha T}] = \int_0^\infty e^{-\alpha t}\mu_x(dt) \tag{7.2.9}$$

of $\mu_x = P_x \circ T^{-1}$.

Example (Exit times in $\mathbb{R}^1$). Suppose $d = 1$ and $D = (-1,1)$. Then (7.2.8) with $V = \alpha$ reads

$$\frac{1}{2}u_\alpha''(x) \;=\; \alpha u_\alpha(x) \quad \text{ for } x \in (-1,1), \quad u_\alpha(1) \;=\; u_\alpha(-1) \;=\; 1.$$

This boundary value problem has the unique solution

$$u_\alpha(x) \;=\; \frac{\cosh(x \cdot \sqrt{2\alpha})}{\cosh(\sqrt{2\alpha})} \qquad \text{for } x \in [-1,1].$$

By inverting the Laplace transform (7.2.9), one can now compute the distribution μ_x of the first exit time T from $(-1,1)$. It turns out that μ_x is absolutely continuous with density

$$f_T(t) = \frac{1}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} \left((4n+1+x)e^{-\frac{(4n+1+x)^2}{2t}} + (4n+1-x)e^{-\frac{(4n+1-x)^2}{2t}} \right), \quad t \geq 0.$$

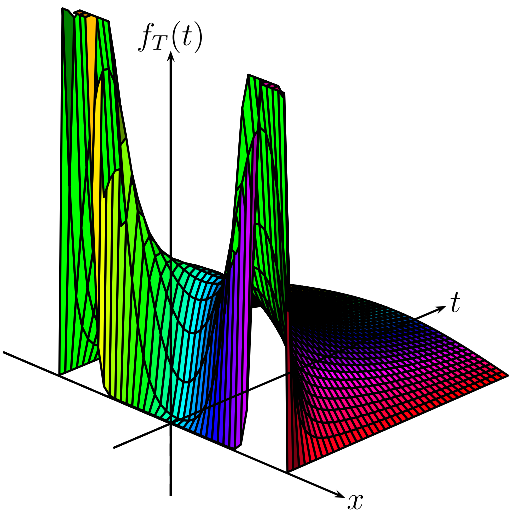


Figure 7.2: The density of the first exit time T depending on the starting point $x \in [-1,1]$ and the time $t \in (0,2]$.

Boundary value problems in $\mathbb{R}^{\mathrm{d}}$ and total occupation time

Suppose we would like to compute the distribution of the total occupation time

$$\int_0^\infty I_A(B_s)\, ds$$

of a bounded domain $A \subset \mathbb{R}^d$ for Brownian motion. This only makes sense for $d \geq 3$, since for $d \leq 2$, the total occupation time of any non-empty open set is almost surely infinite by recurrence of Brownian motion in $\mathbb{R}^1$ and $\mathbb{R}^2$. The total occupation time is of the form $\int_0^\infty V(B_s)\, ds$ with $V = I_A$. Therefore, we should in principle be able to apply Theorem 7.3, but we have to replace the exit time T by $+\infty$ and hence the underlying bounded domain D by $\mathbb{R}^d$.

Corollary 7.6. *Suppose $d \geq 3$ and let $V : \mathbb{R}^d \to [0,\infty)$ be continuous. If $u \in C^2(\mathbb{R}^d)$ is a solution of the boundary value problem*

$$\frac{1}{2}\Delta u \;=\; Vu \quad \text{on } \mathbb{R}^d, \qquad \lim_{|x|\to\infty} u(x) \;=\; 1 \tag{7.2.10}$$

then

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^\infty V(B_t)\, dt\right)\right] \qquad \text{for any } x \in \mathbb{R}^d.$$

Proof. Applying the stationary Feynman-Kac formula on an open bounded subset $D \subset \mathbb{R}^d$, we obtain the representation

$$u(x) \;=\; E_x\left[u(B_{T_{D^C}}) \exp\left(-\int_0^{T_{D^C}} V(B_t)\, dt\right)\right] \tag{7.2.11}$$

by Theorem 7.3. Now let $D_n = \{x \in \mathbb{R}^d : |x| < n\}$. Then $T_{D_n^C} \nearrow \infty$ as $n \to \infty$. Since $d \geq 3$, Brownian motion is transient, i.e., $\lim_{t\to\infty} |B_t| = \infty$, and therefore by (7.2.10)

$$\lim_{n\to\infty} u(B_{T_{D_n^C}}) \;=\; 1 \qquad P_x\text{-almost surely for any } x.$$

Since u is bounded and V is non-negative, we can apply dominated convergence in (7.2.11) to conclude

$$u(x) \;=\; E_x\left[\exp\left(-\int_0^\infty V(B_t)\,dt\right)\right].$$

□

Let us now return to the computation of occupation time distributions. consider a bounded subset $A \subset \mathbb{R}^d$, $d \geq 3$, and let

$$v_\alpha(x) \;=\; E_x\left[\exp\left(-\alpha\int_0^\infty I_A(B_s)\,ds\right)\right], \qquad \alpha > 0,$$

denote the Laplace transform of the total occupation time of A. Although $V = \alpha I_A$ is not a continuous function, a representation of v_α as a solution of a boundary problem holds:

Exercise. Prove that if $A \subset \mathbb{R}^d$ is a bounded domain with smooth boundary ∂A and $u_\alpha \in C^1(\mathbb{R}^d) \cap C^2(\mathbb{R}^d \setminus \partial A)$ satisfies

$$\frac{1}{2}\Delta u_\alpha \;=\; \alpha I_A u_\alpha \quad \text{on } \mathbb{R}^d \setminus \partial A, \qquad \lim_{|x|\to\infty} u_\alpha(x) \;=\; 1, \tag{7.2.12}$$

then $v_\alpha = u_\alpha$.

Remark. The condition $u_\alpha \in C^1(\mathbb{R}^d)$ guarantees that u_α is a weak solution of the p.d.e. (7.2.10) on all of $\mathbb{R}^d$ including the boundary ∂U.

Example (Total occupation time of the unit ball in $\mathbb{R}^3$). Suppose $A = \{x \in \mathbb{R}^3 : |x| < 1\}$. In this case the boundary value problem (7.2.10) is rotationally symmetric. The ansatz $u_\alpha(x) = f_\alpha(|x|)$, yields a Bessel equation for f_α on each of the intervals $(0,1)$ and $(1,\infty)$:

$$\frac{1}{2}f_\alpha''(r) + r^{-1}f_\alpha'(r) \;=\; \alpha f_\alpha(r) \quad \text{for } r < 1, \quad \frac{1}{2}f_\alpha''(r) + r^{-1}f_\alpha(r) \;=\; 0 \quad \text{for } r > 1.$$

Taking into account the boundary condition and the condition $u_\alpha \in C^1(\mathbb{R}^d)$, one obtains the rotationally symmetric solution

$$u_\alpha(x) = \begin{cases} 1 + \left(\dfrac{\tanh(\sqrt{2\alpha})}{\sqrt{2\alpha}} - 1\right) \cdot r^{-1} & \text{for } r \in [1,\infty), \\ \dfrac{\sinh(\sqrt{2\alpha}r)}{\sqrt{2\alpha}\cosh\sqrt{2\alpha}} \cdot r^{-1} & \text{for } r \in (0,1) \\ \dfrac{1}{\cosh(\sqrt{2\alpha})} & \text{for } r = 0 \end{cases} .$$

of (7.2.10), and hence an explicit formula for v_α. In particular, for $x = 0$ we obtain the simple formula

$$E_0\left[\exp\left(-\alpha \int_0^\infty I_A(B_t)\,dt\right)\right] = u_\alpha(0) = \frac{1}{\cosh(\sqrt{2\alpha})}.$$

The right hand side has already appeared in the example above as the Laplace transform of the exit time distribution of a one-dimensional Brownian motion starting at 0 from the interval $(-1,1)$. Since the distribution is uniquely determined by its Laplace transform, we have proven the remarkable fact that the total occupation time of the unit ball for a standard Brownian motion in $\mathbb{R}^3$ has the same distribution as the first exit time from the unit ball for a standard one-dimensional Brownian motion:

$$\int_0^\infty I_{\{|B_t^{\mathbb{R}^3}|<1\}}\,dt \sim \inf\{t > 0 : |B_t^{\mathbb{R}^3}| > 1\}.$$

This is a particular case of a theorem of Ciesielski and Taylor who proved a corresponding relation between Brownian motion in $\mathbb{R}^{d+2}$ and $\mathbb{R}^d$ for arbitrary d.

7.3 Heat Equation and Time-Dependent Feynman-Kac Formula

Itô's formula also yields a connection between Brownian motion (or, more generally, solutions of stochastic differential equations) and parabolic partial differential equations. The parabolic p.d.e. are Kolmogorov forward or backward equations for the corresponding Markov processes. In particular, the time-dependent Feynman-Kac formula shows that the backward equation for Brownian motion with absorption is a heat equation with dissipation.

Brownian Motion with Absorption

Suppose we would like to describe the evolution of a Brownian motion that is absorbed during the evolution of a Brownian motion that is absorbed during an infinitesimal time interval $[t, t+dt]$ with probability $V(t,x)dt$ where x is the current position of the process. We assume that the *absorption rate* $V(t,x)$ is given by a measurable locally-bounded function

$$V : [0,\infty) \times \mathbb{R}^d \;\rightarrow\; [0,\infty).$$

Then the accumulated absorption rate up to time t is given by the increasing process

$$A_t \;=\; \int_0^t V(s, B_s)\, ds, \qquad t \geq 0.$$

We can think of the process A_t as an internal clock for the Brownian motion determining the absorption time. More precisely, we define:

Definition. *Suppose that $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion and T is a with parameter 1 exponentially distributed random variable independent of (B_t). Let Δ be a separate state added to the state space $\mathbb{R}^d$. Then the process (X_t) defined by*

$$X_t \;:=\; \begin{cases} B_t & \text{for } A_t < T, \\ \Delta & \text{for } A_t \geq T, \end{cases}$$

*is called a **Brownian motion with absorption rate** $V(t,x)$, and the random variable*

$$\zeta \;:=\; \inf\{t \geq 0 \;:\; A_t \geq T\}$$

*is called the **absorption time**.*

A justification for the construction is given by the following informal computation: For an infinitesimal time interval $[t, t + dt]$ and almost every ω,

$$\begin{aligned} P[\zeta \le t + dt \mid (B_s)_{s\ge 0}, \zeta > t](\omega) &= P[A_{t+dt}(\omega) \ge T \mid A_t(\omega) < T] \\ &= P[A_{t+dt}(\omega) - A_t(\omega) \ge T] \\ &= P[V(t, B_t(\omega))dt \ge T] \\ &= V(t, B_t(\omega))dt \end{aligned}$$

by the memoryless property of the exponential distribution, i.e., $V(t, x)$ is indeed the infinitesimal absorption rate.

Rigorously, it is not difficult to verify that (X_t) is a Markov process with state space $\mathbb{R}^d \cup \{\Delta\}$ where Δ is an absorbing state. The Markov process is time-homogeneous if $V(t, x)$ is independent of t.

For a measurable subset $D \subseteq \mathbb{R}^d$ and $t \ge 0$ the distribution μ_t of X_t is given by

$$\begin{aligned} \mu_t[D] &= P[X_t \in D] \;=\; P[B_t \in D \text{ and } A_t < T] \\ &= E[P[A_t < T \mid (B_t)]\,;\; B_t \in D] \\ &= E\left[\exp\left(-\int_0^t V(s, B_s)\, ds\right) \;;\; B_t \in D\right]. \end{aligned} \tag{7.3.1}$$

Itô's formula can be used to prove a Kolmogorov type forward equation:

Theorem 7.7 (Forward equation for Brownian motion with absorption). *The sub-probability measures μ_t on $\mathbb{R}^d$ solve the heat equation*

$$\frac{\partial \mu_t}{\partial t} = \frac{1}{2}\Delta\mu_t - V(t, \bullet)\mu_t \tag{7.3.2}$$

in the following distributional sense:

$$\int f(x)\mu_t(dx) - \int f(x)\mu_0(dx) = \int_0^t \int \left(\frac{1}{2}\Delta f(x) - V(s, x)f(x)\right)\mu_s(dx)\, ds$$

for any function $f \in C_0^2(\mathbb{R}^d)$.

Here $C_0^2(\mathbb{R}^d)$ denotes the space of C^2-functions with compact support. Under additional regularity assumptions it can be shown that μ_t has a smooth density that solves (7.3.1) in the classical sense. The equation (7.3.1) describes heat flow with cooling when the heat at x at time t dissipates with rate $V(t,x)$.

Proof. By (7.3.1),

$$\int f\,d\mu_t \;=\; E[\exp(-A_t)\,;\,f(B_t)] \tag{7.3.3}$$

for any bounded measurable function $f:\mathbb{R}^d\to\mathbb{R}$. For $f\in C_0^2(\mathbb{R}^d)$, an application of Itô's formula yields

$$e^{-A_t}f(B_t) \;=\; f(B_0)+M_t+\int_0^t e^{-A_s}f(B_s)V(s,B_s)\,ds+\frac{1}{2}\int_0^t e^{-A_s}\Delta f(B_s)\,ds,$$

for $t\geq 0$, where (M_t) is a local martingale. Taking expectation values for a localizing sequence of stopping times and applying the dominated convergence theorem subsequently, we obtain

$$E[e^{-A_t}f(B_t)] \;=\; E[f(B_0)]+\int_0^t E[e^{-A_s}(\frac{1}{2}\Delta f-V(s,\bullet)f)(B_s)]\,ds.$$

Here we have used that $\frac{1}{2}\Delta f(x)-V(s,x)f(x)$ is uniformly bounded for $(s,x)\in[0,t]\times\mathbb{R}^d$, because f has compact support and V is locally bounded. The assertion now follows by (7.3.3). $\square$

Exercise (Heat kernel and Green's function). The transition kernel for Brownian motion with time-homogeneous absorption rate $V(x)$ restricted to $\mathbb{R}^d$ is given by

$$p_t^V(x,D) \;=\; E_x\left[\exp\left(-\int_0^t V(B_s)\,ds\right)\;;\;B_t\in D\right].$$

(1). Prove that for any $t>0$ and $x\in\mathbb{R}^d$, the sub-probability measure $p_t^V(x,\bullet)$ is absolutely continuous on $\mathbb{R}^d$ with density satisfying

$$0\;\leq\;p_t^V(x,y)\;\leq\;(2\pi t)^{-d/2}\exp(-|x-y|^2/(2t)).$$

(2). Identify the occupation time density

$$G^V(x,y) \;=\; \int_0^\infty p_t^V(x,y)\,dt$$

as a fundamental solution of an appropriate boundary value problem. Adequate regularity may be assumed.

Time-dependent Feynman-Kac formula

In Theorem 7.7 we have applied Itô's formula to prove a Kolmogorov type forward equation for Brownian motion with absorption. To obtain a corresponding backward equation, we have to reverse time:

Theorem 7.8 (Feynman-Kac). *Fix $t > 0$, and let $f : \mathbb{R}^d \to \mathbb{R}$ and $V, g : [0,t] \times \mathbb{R}^d \to \mathbb{R}$ be continuous functions. Suppose that f is bounded, g is non-negative, and V satisfies*

$$E_x\left[\exp \int_0^t V(t-s, B_s)^-\,ds\right] \;<\; \infty \qquad \text{for all } x \in \mathbb{R}^d. \tag{7.3.4}$$

If $u \in C^{1,2}((0,t] \times \mathbb{R}^d) \cap C([0,t] \times \mathbb{R}^d)$ is a bounded solution of the heat equation

$$\begin{aligned} \frac{\partial u}{\partial s}(s,x) \;&=\; \frac{1}{2}\Delta u(s,x) - V(s,x)u(s,x) + g(s,x) \qquad \text{for } s \in (0,t], x \in \mathbb{R}^d, \\ u(0,x) \;&=\; f(x), \end{aligned} \tag{7.3.5}$$

then u has the stochastic representation

$$\begin{aligned} u(t,x) \;=\; & E_x\left[f(B_t)\exp\left(-\int_0^t V(t-s,B_s)\,ds\right)\right] + \\ & E_x\left[\int_0^t g(t-r,B_r)\exp\left(-\int_0^r V(t-s,B_s)\,ds\right)\,dr\right]. \end{aligned}$$

Remark. The equation (7.3.5) describes heat flow with sinks and dissipation.

Proof. We first reverse time on the interval $[0,t]$. The function

$$\hat{u}(s,x) \;=\; u(t-s,x)$$

solves the p.d.e.

$$\begin{aligned}\frac{\partial \hat{u}}{\partial s}(s,x) &= -\frac{\partial u}{\partial t}(t-s,x) \;=\; -\left(\frac{1}{2}\Delta u - Vu + g\right)(t-s,x)\\ &= -\left(\frac{1}{2}\Delta \hat{u} - \hat{V}\hat{u} + \hat{g}\right)(s,x)\end{aligned}$$

on $[0,t]$ with terminal condition $\hat{u}(t,x) \;=\; f(x)$. Now let $X_r = \exp(-A_r)\hat{u}(r,B_r)$ for $r \in [0,t]$, where

$$A_r \;:=\; \int_0^r \hat{V}(s,B_s)\,ds \;=\; \int_0^r V(t-s,B_s)\,ds.$$

By Itô's formula, we obtain for $\tau \in [0,t]$,

$$\begin{aligned}X_\tau - X_0 &= M_\tau - \int_0^\tau e^{-A_r}\hat{u}(r,B_r)\,dA_r + \int_0^\tau e^{-A_r}\left(\frac{\partial \hat{u}}{\partial s} + \frac{1}{2}\Delta\hat{u}\right)(r,B_r)\,dr\\ &= M_\tau + \int_0^\tau e^{-A_r}\left(\frac{\partial \hat{u}}{\partial s} + \frac{1}{2}\Delta\hat{u} - \hat{V}\hat{u}\right)(r,B_r)\,dr\\ &= M_\tau - \int_0^\tau e^{-A_r}\hat{g}(r,B_r)\,dr\end{aligned}$$

with a local martingale $(M_\tau)_{\tau\in[0,t]}$ vanishing at 0. Choosing a corresponding localizing sequence of stopping times T_n with $T_n \nearrow t$, we obtain by the optional stopping theorem and by dominated convergence,

$$\begin{aligned}u(t,x) &= \hat{u}(0,x) \;=\; E_x[X_0]\\ &= E_x[X_t] + E_x\left[\int_0^t e^{-A_r}\hat{g}(r,B_r)\,dr\right]\\ &= E_x[e^{-A_t}u(0,B_t)] + E_x\left[\int_0^t e^{-A_r}g(t-r,B_r)\,dr\right].\end{aligned}$$

□

Remark (Extension to diffusion processes). Again a similar result holds under a appropriate regularity assumptions for Brownian motion replaced by a solution of a s.d.e. $dX_t = \sigma(X_t)dB_t + b(X_t)dt$ and $\frac{1}{2}\Delta$ replaced by the corresponding generator, cf. **??**.

Occupation times and arc-sine law

The Feynman-Kac formula can be used to study the distribution of occupation times of Brownian motion. We consider an example where the distribution can be computed explicitly: The proportion of time during the interval $[0, t]$ spent by a one-dimensional standard Brownian motion (B_t) in the interval $(0, \infty)$. Let

$$A_t \;=\; \lambda(\{s \in [0,t] \;:\; B_s > 0\}) \;=\; \int_0^t I_{(0,\infty)}(B_s)\, ds.$$

Theorem 7.9 (Arc-sine law of P.Lévy). *For any $t > 0$ and $\theta \in [0, 1]$,*

$$P_0[A_t/t \le \theta] \;=\; \frac{2}{\pi}\arcsin\sqrt{\theta} \;=\; \frac{1}{\pi}\int_0^\theta \frac{ds}{\sqrt{s(1-s)}}.$$

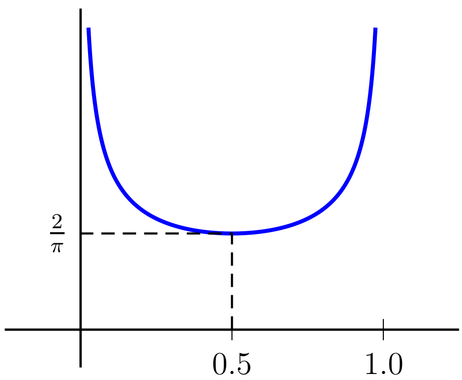


Figure 7.3: Density of A_t/t.

Note that the theorem shows in particular that a law of large numbers does **not** hold! Indeed, for each $\varepsilon > 0$,

$$P_0\left[\left|\frac{1}{t}\int_0^t I_{(0,\infty)}(B_s)\,ds - \frac{1}{2}\right| > \varepsilon\right] \quad \not\to \quad 0 \qquad \text{as } t \to \infty.$$

Even for large times, values of A_t/t close to 0 or 1 are the most probable. By the functional central limit theorem, the proportion of time that one player is ahead in a long coin tossing game or a counting of election results is also close to the arcsine law. In particular, it is more then 20 times more likely that one player is ahead for more than 98% of the time than it is that each player is ahead between 49% and 51% of the time [Steele].

Before proving the arc-sine law, we give an informal derivation based on the time-dependent Feynman-Kac formula.
The idea for determining the distribution of A_t is again to consider the Laplace transforms

$$u(t,x) \;=\; E_x[\exp(-\beta A_t)], \qquad \beta > 0.$$

By the Feynman-Kac formula, we could expect that u solves the equation

$$\frac{\partial u}{\partial t} \;=\; \frac{1}{2}\frac{\partial^2 u}{\partial x^2} \tag{7.3.6}$$

with initial condition $u(0,x) = 1$. To solve the parabolic p.d.e. (7.3.6), we consider another Laplace transform: The Laplace transform

$$v_\alpha(x) \;=\; \int_0^\infty e^{-\alpha t}u(t,x)\,dt \;=\; E_x\left[\int_0^\infty e^{-\alpha t - \beta A_t}\,dt\right], \qquad \alpha > 0,$$

of a solution $u(t,x)$ of (7.3.6) w.r.t. t. An informal computation shows that v_α should satisfy the o.d.e.

$$\begin{aligned}
\frac{1}{2}v_\alpha'' - \beta I_{(0,\infty)}v_\alpha \;&=\; \int_0^\infty e^{-\alpha t}\left(\frac{1}{2}\frac{\partial^2 u}{\partial x^2} - \beta I_{(0,\infty)}u\right)(t,\bullet)\,dt \\
&=\; \int_0^\infty e^{-\alpha t}\frac{\partial u}{\partial t}(t,\bullet)\,dt \;=\; e^{-\alpha t}u(t,\bullet)|_0^\infty - \alpha\int_0^\infty e^{-\alpha t}u(t,\bullet)\,dt \\
&=\; 1 - \alpha v_\alpha,
\end{aligned}$$

i.e., v_α should be a bounded solution of

$$\alpha v_\alpha - \frac{1}{2} v_\alpha'' + \beta I_{0,\infty} v_\alpha \;=\; g \tag{7.3.7}$$

where $g(x) = 1$ for all x. The solution of (7.3.7) can then be computed explicitly, and yield the arc-sine law by Laplace inversion.

Remark. The method of transforming a parabolic p.d.e. by the Laplace transform into an elliptic equation is standard and used frequently. In particular, the Laplace transform of a transition semigroup $(p_t)_{t\ge 0}$ is the corresponding resolvent $(g_\alpha)_{\alpha\ge 0}, g_\alpha = \int_0^\infty e^{-\alpha t} p_t\, dt$, which is crucial for potential theory.

Instead of trying to make the informal argument above rigorous, one can directly prove the arc-sine law by applying the stationary Feynman-Kac formula:

Exercise. Prove Lévy's arc-sine law by proceeding in the following way:

(1). Let $g \in C_b(\mathbb{R})$. Show that if v_α is a bounded solution of (7.3.7) on $\mathbb{R} \setminus \{0\}$ with $v_\alpha \in C^1(\mathbb{R}) \cap C^2(\mathbb{R} \setminus \{0\})$ then

$$v_\alpha(x) \;=\; E_x\left[\int_0^\infty g(B_t) e^{-\alpha t - \beta A_t}\, dt\right] \qquad \text{for any } x \in \mathbb{R}.$$

(2). Compute a corresponding solution v_α for $g \equiv 1$, and conclude that

$$\int_0^\infty e^{-\alpha t} E_0[e^{-\beta A_t}]\, dt \;=\; \frac{1}{\sqrt{\alpha(\alpha+\beta)}}.$$

(3). Now use the uniqueness of the Laplace inversion to show that the distribution μ_t of A_t/t under $P_\bullet$ is absolutely continuous with density

$$f_{A_t/t}(s) \;=\; \frac{1}{\pi\sqrt{s\cdot(1-s)}}.$$

Chapter 8

Stochastic Differential Equations: Explicit Computations

Suppose that $(B_t)_{t\geq 0}$ is a given Brownian motion defined on a probability space $(\Omega, \mathcal{A}, P)$. We will now study solutions of stochastic differential equations (SDE) of type

$$dX_t \;=\; b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \tag{8.0.1}$$

where b and σ are continuous functions defined on $\mathbb{R}_+ \times \mathbb{R}^d$ or an appropriate subset.

Recall that $\mathcal{F}_t^{B,P}$ denotes the completion of the filtration $\mathcal{F}_t^B \;=\; \sigma(B_s \mid 0 \leq s \leq t)$ generated by the Brownian motion. Let T be an $(\mathcal{F}_t^{B,P})$ stopping time. We call a process $(t, \omega) \mapsto X_t(\omega)$ defined for $t < T(\omega)$ **adapted w.r.t.** $\left(\mathcal{F}_{\mathbf{t}}^{\mathbf{B,P}}\right)$, if the trivially extended process $\widetilde{X}_t \;=\; X_t \cdot I_{\{t<T\}}$ defined by

$$\widetilde{X}_t \;:=\; \begin{cases} X_t & \text{for } t < T \\ 0 & \text{for } t \geq T \end{cases},$$

is $(\mathcal{F}_t^{B,P})$-adapted.

Definition. *An almost surely continuous stochastic process* $(t,\omega) \mapsto X_t(\omega)$ *defined for* $t \in [0, T(\omega))$ *is called a* ***strong solution*** *of the stochastic differential equation (8.0.1) if it is* $(\mathcal{F}_t^{B,P})$*-adapted, and the equation*

$$X_t \;=\; X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \quad \text{for } t \in [0, T) \tag{8.0.2}$$

holds P*-almost surely.*

The terminology "strong" solution will be explained later when we introduce "weak" solutions. The point is that a strong solution is adapted w.r.t. the filtration $(\mathcal{F}_t^{B,P})$ generated by the Brownian motion. Therefore, a strong solution is essentially (up to modification on measure zero sets) a *measurable function of the given Brownian motion*! The concept of strong and weak solutions of SDE is not related to the analytic definition of strong and weak solutions for partial differential equations.

In this section we study properties of solutions and we compute explicit solutions for one-dimensional SDE. We start with an example:

Example (Asset price model in continuous time). A nearby model for an asset price process $(S_n)_{n=0,1,2,\ldots}$ in discrete time is to define S_n recursively by

$$S_{n+1} - S_n \;=\; \alpha_n(S_0, \ldots, S_n) S_n + \sigma_n(S_0, \ldots, S_n) S_n \eta_{n+1}$$

with i.i.d. random variables $\eta_i, i \in \mathbb{N}$, and measurable functions $\alpha_n, \sigma_n : \mathbb{R}^n \to \mathbb{R}$. Trying to set up a corresponding model in continuous time, we arrive at the stochastic differential equation

$$dS_t \;=\; \alpha_t S_t\, dt + \sigma_t S_t\, dB_t \tag{8.0.3}$$

with an $(\mathcal{F}_t)$-Brownian motion (B_t) and $(\mathcal{F}_t^P)$ adapted continuous stochastic processes $(\alpha_t)_{t\geq 0}$ and $(\sigma_t)_{t\geq 0}$, where $(\mathcal{F}_t)$ is a given filtration on a probability space $(\Omega, \mathcal{A}, P)$. The processes α_t and σ_t describe the *instantaneous mean rate of return* and the *volatility*. Both are allowed to be time dependent and random.

In order to compute a solution of (8.0.3), we assume $S_t > 0$ for any $t \geq 0$, and divide the equation by S_t:

$$\frac{1}{S_t}\,dS_t \;=\; \alpha_t\,dt + \sigma_t\,dB_t. \tag{8.0.4}$$

We will prove in Section 8.1 that if an SDE holds then the SDE multiplied by a continuous adapted process also holds, cf. Theorem 8.1. Hence (8.0.4) is equivalent to (8.0.3) if $S_t > 0$. If (8.0.4) would be a classical ordinary differential equation then we could use the identity $d\,\log S_t = \frac{1}{S_t}\,dS_t$ to solve the equation. In Itô calculus, however, the classical chain rule is violated. Nevertheless, it is still useful to compute $d\,\log S_t$ by Itô's formula. The process (S_t) has quadratic variation

$$[S]_t \;=\; \left[\int_0^\bullet \sigma_r S_r\,dB_r\right]_t \;=\; \int_0^t \sigma_r^2 S_r^2\,dr \qquad \text{for any } t \geq 0,$$

almost surely along an appropriate sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$. The first equation holds by (8.0.3), since $t \mapsto \int_0^t \alpha_r S_r\,dr$ has finite variation, and the second identity is proved in Theorem 8.1 below. Therefore, Itô's formula implies:

$$\begin{aligned} d\,\log S_t \;&=\; \frac{1}{S_t}\,dS_t - \frac{1}{2S_t^2}\,d[S]_t \\ &=\; \alpha_t\,dt + \sigma_t\,dB_t - \frac{1}{2}\sigma_t^2\,dt \\ &=\; \mu_t\,dt + \sigma_t\,dB_t, \end{aligned}$$

where $\mu_t := \alpha_t - \sigma_t^2/2$, i.e.,

$$\log S_t - \log S_0 \;=\; \int_0^t \mu_s\,ds + \int_0^t \sigma_s\,dB_s,$$

or, equivalently,

$$S_t \;=\; S_0 \cdot \exp\left(\int_0^t \mu_s\,ds + \int_0^t \sigma_s\,dB_s\right). \tag{8.0.5}$$

Conversely, one can verify by Itô's formula that (S_t) defined by (8.0.5) is indeed a solution of (8.0.3). Thus we have proven existence, uniqueness and an explicit representation for a strong solution of (8.0.3). In the special case when $\alpha_t \equiv \alpha$ and $\sigma_t \equiv \sigma$ are constants in t and ω, the solution process

$$S_t \;=\; S_0 \exp\left(\sigma B_t + (\alpha - \sigma^2/2)t\right)$$

is called a **geometric Brownian motion with parameters α and σ**.

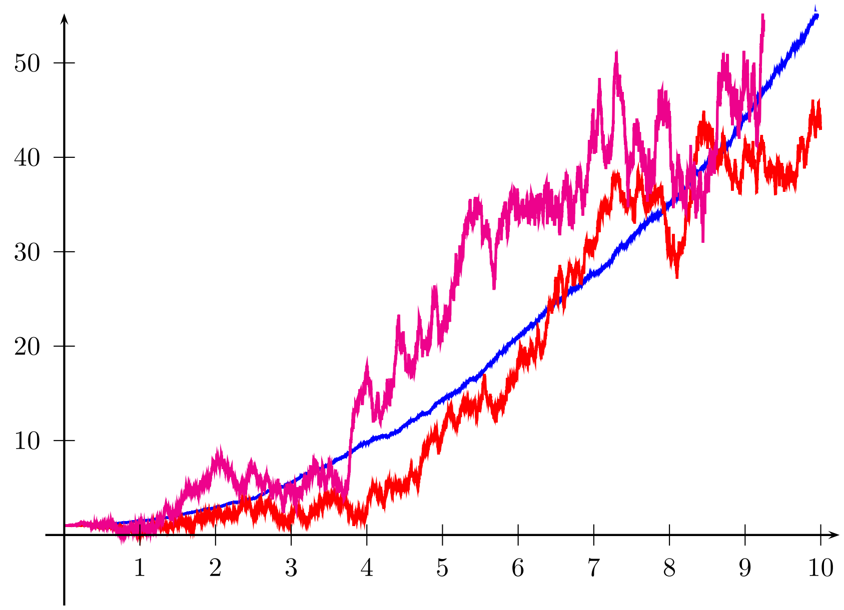


Figure 8.1: Three one dimensional geometric Brownian motions with $\alpha^2 = 1$ and $\sigma = 0.1$ (blue), $\sigma = 1.0$ (red) and $\sigma = 2.0$ (magenta).

8.1 Stochastic Calculus for Itô processes

By definition, any solution of an SDE of the form (8.0.1) is the sum of an absolutely continuous adapted process and an Itô stochastic integral w.r.t. the underlying Brownian motion, i.e.,

$$X_t = A_t + I_t \qquad \text{for } t < T, \tag{8.1.1}$$

where

$$A_t = \int_0^t K_s\, ds \quad \text{and} \quad I_t = \int_0^t H_s\, dB_s \tag{8.1.2}$$

with $(H_t)_{t<T}$ and $(K_t)_{t<T}$ almost surely continuous and $(\mathcal{F}_t^{B,P})$-adapted. A stochastic process of type (8.1.1) is called an **Itô process**. In order to compute and analyze

solutions of SDE we will apply Itô's formula to Itô processes. Since the absolutely continuous process (A_t) has finite variation, classical Stieltjes calculus applies to this part of an Itô process. It remains to consider the stochastic integral part (I_t):

Stochastic integrals w.r.t. Itô processes

Let (π_n) be a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. Recall that for $t \geq 0$,

$$I_t \;=\; \int_0^t H_s\, dB_s \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} H_s \cdot (B_{s'\wedge t} - B_s)$$

w.r.t. convergence in probability on $\{t < T\}$, cf. Theorem 5.14.

Theorem 8.1 (Composition rule and quadratic variation). *Suppose that T is a predictable stopping time and $(H_t)_{t<T}$ is almost surely continuous and adapted.*

(1). For any almost surely continuous, adapted process $(G_t)_{0\leq t<T}$, and for any $t \geq 0$,

$$\lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} G_s(I_{s'\wedge t} - I_s) \;=\; \int_0^t G_s H_s\, dB_s \tag{8.1.3}$$

with convergence in probability on $\{t < T\}$. Moreover, if H is in $\mathcal{L}^2_a([0,a])$ and G is bounded on $[0,a]\times\Omega$ for some $a > 0$, then the convergence holds in $M^2_c([0,a])$ and thus uniformly for $t \in [0,a]$ in the $L^2(P)$ sense.

(2). For any $t \geq 0$, the quadratic variation $[I]_t$ along (π_n) is given by

$$[I]_t \;=\; \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} (I_{s'\wedge t} - I_s)^2 \;=\; \int_0^t H_s^2\, ds \tag{8.1.4}$$

w.r.t. convergence in probability on $\{t < T\}$.

XXX gleich $[I,J]$ berechnen, Beweis analog

Remark (Uniform convergence). Similarly to the proof of Theorem 5.14 one can show that there is a sequence of bounded stopping times $T_k \nearrow T$ such that almost surely along a subsequence, the convergence in (8.1.3) and (8.1.4) holds uniformly on $[0, T_k]$ for any $k \in \mathbb{N}$.

Proof. (1). We first fix $a > 0$ and assume that H is in $\mathcal{L}^2_a([0, a))$ and G is bounded, left-continuous and adapted on $[0, \infty) \times \Omega$. Since $I_{s' \wedge t} - I_s = \int_s^{s' \wedge t} H_r \, dB_r$, we obtain

$$\sum_{\substack{s \in \pi_n \\ s < t}} G_s (I_{s' \wedge t} - I_s) \;=\; \int_0^t G_{\lfloor r \rfloor_n} H_r \, dB_r$$

where $\lfloor r \rfloor_n = \max\{s \in \pi_n : s \leq r\}$ is the next partition point below r. As $n \to \infty$, the right-hand side converges to $\int_0^t G_r H_r \, dB_r$ in $M^2_c([0, a])$ because $G_{\lfloor r \rfloor_n} H_r \to G_r H_r$ in $L^2(P \otimes \lambda_{[0,a)})$ by continuity of G and dominated convergence.

The assertion in the general case now follows by localization: Suppose (S_k) and (T_k) are increasing sequences of stopping times with $T_k \nearrow T$ and $H_t I_{\{t \leq S_k\}} \in \mathcal{L}^2_a([0, \infty))$, and let

$$\widetilde{T}_k \;=\; S_k \wedge T_k \wedge \inf\{t \geq 0 \;:\; |G_t| > k\} \wedge k.$$

Then $\widetilde{T}_k \nearrow T$, the process $H^{(k)}_t := H_t I_{\{t \leq T_k\}}$ is in $\mathcal{L}^2_a([0, \infty))$ the process $G^{(k)}_t := G_t I_{\{t \leq T_k\}}$ is bounded, left-continuous and adapted, and

$$I_s \;=\; \int_0^s H^{(k)}_r \, dB_r, \quad G_s \;=\; G^{(k)}_s \qquad \text{for any } s \in [0, t]$$

holds almost surely on $\{t \leq \widetilde{T}_k\}$. Therefore as $n \to \infty$,

$$\begin{aligned}
\sum_{\substack{s \in \pi_n \\ s < t}} G_s (I_{s' \wedge t} - I_s) \;&=\; \sum_{\substack{s \in \pi_n \\ s < t}} G^{(k)}_s (I_{s' \wedge t} - I_s) \\
&\to\; \int_0^t G^{(k)}_r H^{(k)}_r \, dB_r \;=\; \int_0^t G_r H_r \, dB_r
\end{aligned}$$

uniformly for $t \leq \widetilde{T}_k$ in $L^2(P)$. The claim follows, since

$$P\left[\{t < T\} \setminus \bigcup_k \{t \leq \widetilde{T}_k\}\right] \;=\; 0.$$

(2). We first assume that H is in $\mathcal{L}^2_a([0,\infty))$, continuous and bounded. Then for $s \in \pi_n$,

$$\delta I_s \;=\; I_{s'\wedge t} - I_s \;=\; \int_s^{s'\wedge t} H_r\, dB_r \;=\; H_s \delta B_s + R_s^{(n)}$$

where $R_s^{(n)} \;:=\; \int_s^{s'\wedge t} (H_r - H_{\lfloor r \rfloor_n})\, dB_r$. Therefore,

$$\sum_{\substack{s\in\pi_n\\ s<t}} (\delta I_s)^2 \;=\; \sum_{\substack{s\in\pi_n\\ s<t}} H_s^2 (\delta B_s)^2 + 2 \sum_{\substack{s\in\pi_n\\ s<t}} R_s^{(n)} H_s \delta B_s + \sum_{\substack{s\in\pi_n\\ s<t}} (R_s^{(n)})^2.$$

Since $[B]_t \;=\; t$ almost surely, the first term on the right-hand side converges to $\int_0^t H_s^2\, ds$ with probability one. It remains to show that the remainder terms converge to 0 in probability as $n \to \infty$. This is the case, since

$$\begin{aligned} E\left[\sum (R_s^{(n)})^2\right] \;&=\; \sum E[(R_s^{(n)})^2] \;=\; \sum \int_s^{s'\wedge t} E[(H_r - H_{\lfloor r \rfloor_n})^2]\, dr \\ &=\; \int_0^t E[(H_r - H_{\lfloor r \rfloor_n})^2]\, dr \;\longrightarrow\; 0 \end{aligned}$$

by the Itô isometry and continuity and boundedness of H, whence $\sum (R_s^{(n)})^2 \to 0$ in $\mathcal{L}^1$ and in probability, and $\sum R_s^{(n)} H_s \delta B_s \to 0$ in the same sense by the Schwarz inequality.

For H defined up to a stopping time T, the assertion now follows by a localization procedure similar to the one applied above.

$\square$

The theorem and the corresponding composition rule for Stieltjes integrals suggest that we may define stochastic integrals w.r.t. an Itô process

$$X_t \;=\; X_0 + \int_0^t H_s\, dB_s + \int_0^t K_s\, ds, \qquad t < T,$$

in the following way:

Definition. *Suppose that (B_t) is a Brownian motion on $(\Omega, \mathcal{A}, P)$ w.r.t. a filtration $(\mathcal{F}_t)$, X_0 is an $(\mathcal{F}_0^P)$-measurable random variable, T is a predictable $(\mathcal{F}_t^P)$-stopping time, and (G_t), (H_t) and (K_t) are almost surely continuous, $(\mathcal{F}_t^P)$ adapted processes defined for $t < T$. Then the stochastic integral of (G_t) w.r.t. (X_t) is the Itô process defined by*

$$\int_0^t G_s\,dX_s \;=\; \int_0^t G_s H_s\,dB_s + \int_0^t G_s K_s\,ds, \qquad t < T.$$

By Theorem 8.1, this definition is consistent with a definition by Riemann sum approximations. Moreover, the definition shows that the class of Itô processes w.r.t. a given Brownian motion is *closed under taking stochastic integrals*! In particular, strong solutions of SDE w.r.t. Itô processes are again Itô processes.

Calculus for Itô processes

We summarize calculus rules for Itô processes that are immediate consequences of the definition above and Theorem 8.1: Suppose that (X_t) and (Y_t) are Itô processes, and (G_t), $(\widetilde{G}_t)$ and (H_t) are adapted continuous process that are all defined up to a stopping time T. Then the following calculus rules hold for Itô stochastic differentials:

Linearity:

$$\begin{aligned} d(X + cY) &= dX + c\,dY && \text{for any } c \in \mathbb{R}, \\ (G + cH)\,dX &= G\,dX + cH\,dX && \text{for any } c \in \mathbb{R}. \end{aligned}$$

Composition rule:

$$dY \;=\; G\,dX \quad \Rightarrow \quad \widetilde{G}\,dY \;=\; \widetilde{G}G\,dX,$$

Quadratic variation:

$$dY \;=\; G\,dX \quad \Rightarrow \quad d[Y] \;=\; G^2\,d[X],$$

Itô rule: For any function $F \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R})$,

$$dF(t,X) \;=\; \frac{\partial F}{\partial x}(t,X)\,dX + \frac{\partial F}{\partial t}(t,X)\,dt + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t,X)\,d[X]$$

All equations are to be understood in the sense that the corresponding stochastic integrals over any interval $[0,t], t < T$, coincide almost surely.
The proofs are straightforward. For example, if

$$Y_t \;=\; Y_0 + \int_0^t G_s\,dX_s$$

and

$$X_t \;=\; X_0 + \int_0^t K_s\,ds + \int_0^t H_s\,dB_s$$

then, by the definition above, for $t < T$,

$$Y_t \;=\; Y_0 + \int_0^t G_s K_s\,ds + \int_0^t G_s H_s\,dB_s,$$

and hence

$$\int_0^t \widetilde{G}_s\,dY_s \;=\; \int_0^t \widetilde{G}_s G_s K_s\,ds + \int_0^t \widetilde{G}_s G_s H_s\,dB_s \;=\; \int_0^t \widetilde{G}_s G_s\,dX_s$$

and

$$[Y]_t \;=\; \left[\int_0^\bullet G_s H_s\,dB_s\right]_t \;=\; \int_0^t G_s^2 H_s^2\,ds \;=\; \int_0^t G_s^2\,d[X]_s.$$

Moreover, Theorem 8.1 guarantees that the stochastic integrals in Itô's formula (which are limits of Riemann-Itô sums) coincide with the stochastic integrals w.r.t. Itô processes defined above.

Example (**Option Pricing in continuous time I**). We again consider the continuous time asset price model introduced in the beginning of Chapter 8. Suppose an agent is holding ϕ_t units of a single asset with price process (S_t) at time t, and he invests the remainder $V_t - \phi_t S_t$ of his wealth V_t in the money market with interest rate R_t. We assume that (ϕ_t) and (R_t) are continuous adapted processes. Then the change of wealth in a small time unit should be described by the Itô equation

$$dV_t \;=\; \phi_t\,dS_t + R_t(V_t - \phi_t S_t)\,dt.$$

Similarly to the discrete time case, we consider the discounted wealth process

$$\widetilde{V}_t \;:=\; \exp\left(-\int_0^t R_s\,ds\right) V_t.$$

Since $t \mapsto \int_0^t R_s\,ds$ has finite variation, the Itô rule and the composition rule for stochastic integrals imply:

$$\begin{aligned} d\widetilde{V}_t \;&=\; \exp\left(-\int_0^t R_s\,ds\right) dV_t - \exp\left(-\int_0^t R_s\,ds\right) R_t V_t\,dt \\ &=\; \exp\left(-\int_0^t R_s\,ds\right) \phi_t\,dS_t - \exp\left(-\int_0^t R_s\,ds\right) R_t\phi_t S_t\,dt \\ &=\; \phi_t \cdot \left(\exp\left(-\int_0^t R_s\,ds\right) dS_t - \exp\left(-\int_0^t R_s\,ds\right) R_t S_t\,dt\right) \\ &=\; \phi_t\,d\widetilde{S}_t, \end{aligned}$$

where $\widetilde{S}_t$ is the discounted asset price process. Therefore,

$$\widetilde{V}_t - \widetilde{V}_0 \;=\; \int_0^t \phi_s\,d\widetilde{S}_s \qquad \forall t \geq 0 \quad P\text{-almost surely.}$$

As a consequence, we observe that if $(\widetilde{S}_t)$ is a (local) martingale under a probability measure P_* that is equivalent to P then the discounted wealth process $(\widetilde{V}_t)$ is also a local martingale under P_*. A corresponding probability measure P_* is called an *equivalent martingale* measure or *risk neutral measure*, and can be identified by Girsanov's theorem, cf. Section 9.3 below. Once we have found P_*, option prices can be computed similarly as in discrete time under the additional assumption that the true measure P for the asset price process is equivalent to P_*, see Section 9.4.

The Itô-Doeblin formula in $\mathbb{R}^1$

We will now apply Itô's formula to solutions of stochastic differential equations. Let $b, \sigma \in C(\mathbb{R}_+ \times I)$ where $I \subseteq \mathbb{R}$ is an open interval. Suppose that (B_t) is an $(\mathcal{F}_t)$-Brownian motion on $(\Omega, \mathcal{A}, P)$, and $(X_t)_{0 \le t < T}$ is an $(\mathcal{F}_t^P)$-adapted process with values in I and defined up to an $(\mathcal{F}_t^P)$ stopping time T such that the SDE

$$X_t - X_0 = \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \qquad \text{for any } t < T \tag{8.1.5}$$

holds almost surely.

Corollary 8.2 (Doeblin 1941, Itô 1944). *Let $F \in C^{1,2}(\mathbb{R}_+ \times I)$. Then almost surely,*

$$\begin{aligned} F(t, X_t) - F(0, X_0) &= \int_0^t (\sigma F')(s, X_s)\, dB_s \\ &\quad + \int_0^t \left(\frac{\partial F}{\partial t} + \frac{1}{2}\sigma^2 F'' + bF' \right)(s, X_s)\, ds \qquad \textit{for any } t < T, \end{aligned} \tag{8.1.6}$$

where $F' = \partial F / \partial x$ denotes the partial derivative w.r.t. x.

Proof. Let (π_n) be a sequence of partitions with $\mathrm{mesh}(\pi_n) \to 0$. Since the process $t \mapsto X_0 + \int_0^t b(s, X_s)\, ds$ has sample paths of locally finite variation, the quadratic variation of (X_t) is given by

$$[X]_t = \left[\int_0^\bullet \sigma(s, X_s)\, dB_s \right]_t = \int_0^t \sigma(s, X_s)^2\, ds \qquad \forall t < T$$

w.r.t. almost sure convergence along a subsequence of (π_n). Hence Itô's formula can be applied to almost every sample path of (X_t), and we obtain

$$\begin{aligned} F(t, X_t) - F(0, X_0) &= \int_0^t F'(s, X_s)\, dX_s + \int_0^t \frac{\partial F}{\partial t}(s, X_s)\, ds + \frac{1}{2}\int_0^t F''(s, X_s)\, d[X]_s \\ &= \int_0^t (\sigma F')(s, X_s)\, dB_s + \int_0^t (bF')(s, X_s)\, ds + \int_0^t \frac{\partial F}{\partial t}(s, X_s)\, ds + \frac{1}{2}\int_0^t (\sigma^2 F'')(s, X_s)\, ds \end{aligned}$$

for all $t < T$, P-almost surely. Here we have used (8.1.5) and the fact that the Itô integral w.r.t. X is an almost sure limit of Riemann-Itô sums after passing once more to an appropriate subsequence of (π_n). □

Exercise (Black Scholes partial differential equation). A stock price is modeled by a geometric Brownian Motion (S_t) with parameters $\alpha, \sigma > 0$. We assume that the interest rate is equal to a real constant r for all times. Let $c(t, x)$ be the value of an option at time t if the stock price at that time is $S_t = x$. Suppose that $c(t, S_t)$ is replicated by a hedging portfolio, i.e., there is a trading strategy holding ϕ_t shares of stock at time t and putting the remaining portfolio value $V_t - \phi_t S_t$ in the money market account with fixed interest rate r so that the total portfolio value V_t at each time t agrees with $c(t, S_t)$.

"Derive" the *Black-Scholes partial differential equation*

$$\frac{\partial c}{\partial t}(t, x) + rx\frac{\partial c}{\partial x}(t, x) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 c}{\partial x^2}(t, x) \;=\; rc(t, x) \tag{8.1.7}$$

and the *delta-hedging rule*

$$\phi_t \;=\; \frac{\partial c}{\partial x}(t, S_t) \qquad (=: \text{ Delta }). \tag{8.1.8}$$

Hint: Consider the discounted portfolio value $\widetilde{V}_t = e^{-rt}V_t$ and, correspondingly, the discounted option value $e^{-rt}c(t, S_t)$. Compute the Ito differentials, and conclude that both processes coincide if c is a solution to (8.1.7) and ϕ_t is given by (8.1.8).

Martingale problem for solutions of SDE

The Itô-Doeblin formula shows that if (X_t) is a solution of (8.1.5) then

$$M_t^F = F(t, X_t) - F(0, X_0) - \int_0^t (\mathscr{L}F)(s, X_s)\, ds$$

is a local martingale up to T for any $F \in C^{1,2}(\mathbb{R}_+ \times I)$ and

$$(\mathscr{L}F)(t, x) = \frac{1}{2}\sigma(t, x)^2 F''(t, x) + b(t, x)F'(t, x).$$

In particular, in the time-homogeneous case and for $T = \infty$, any solution of (8.1.5) solves the martingale problem for the operator $\mathscr{L}F = \frac{1}{2}\sigma^2 F'' + bF'$ with domain $C_0^2(I)$. Similarly as for Brownian motion, the martingales identified by the Itô-Doeblin formula can be used to compute various expectation values for the Itô diffusion (X_t). In the next section we will look at first examples.

Remark (Uniqueness and Markov property of strong solutions). If the coefficients are, for example, Lipschitz continuous, then the strong solution of the SDE (8.1.5) is unique, and it has the strong Markov property, i.e., it is a diffusion process in the classical sense (a strong Markov process with continuous sample paths). By the Itô-Doeblin formula, the generator of this Markov process is an extension of the operator $(\mathscr{L}, C_0^2(I))$.

Although in general, uniqueness and the Markov property may not hold for solutions of the SDE (8.1.5), we call any solution of this equation an **Itô diffusion**.

8.2 Stochastic growth

In this section we consider time-homogeneous Itô diffusions taking values in the interval $I = (0, \infty)$. They provide natural models for stochastic growth processes, e.g. in mathematical biology, financial mathematics and many other application fields. Analogue results also hold if I is replaced by an arbitrary non-empty open interval.
Suppose that $(X_t)_{0 \le t < T}$ is a strong solution of the SDE

$$\begin{aligned} dX_t &= b(X_t)\, dt + \sigma(X_t)\, dB_t \qquad \text{for } t \in [0, T), \\ X_0 &= x_0, \end{aligned}$$

with a given Brownian motion $(B_t), x_0 \in (0,\infty)$, and continuous time-homogeneous coefficients $b, \sigma : (0,\infty) \to \mathbb{R}$. We assume that the solution is defined up to the explosion time

$$T \;=\; \sup_{\varepsilon, r>0} T_{\varepsilon,r}, \quad T_{\varepsilon,r} \;=\; \inf\{t \geq 0 \mid X_t \notin (\varepsilon, r)\}.$$

The corresponding generator is

$$\mathscr{L}F \;=\; bF' + \frac{1}{2}\sigma^2 F''.$$

Before studying some concrete models, we show in the general case how harmonic functions can be used to compute exit distributions (e.g. ruin probabilities) and to analyze the asymptotic behaviour of X_t as $t \nearrow T$.

Scale functions and exit distributions

To determine the exit distribution from a finite subinterval $(\varepsilon, r) \subset (0,\infty)$ we compute the harmonic functions of $\mathscr{L}$. For $h \in C^2(0,\infty)$ with $h' > 0$ we obtain:

$$\mathscr{L}h \;=\; 0 \quad \iff \quad h'' \;=\; -\frac{2b}{\sigma^2}h' \quad \iff \quad (\log h')' \;=\; -\frac{2b}{\sigma^2}.$$

Therefore, the two-dimensional vector space of harmonic functions is spanned by the constant function 1 and by the function

$$s(x) \;=\; \int_{x_0}^{x} \exp\left(-\int_{x_0}^{z} \frac{2b(y)}{\sigma(y)^2}\, dy \right) dz.$$

$s(x)$ is called a **scale function** of the process (X_t). It is strictly increasing and harmonic on $(0,\infty)$. Hence we can think of $s : (0,\infty) \to (s(0), s(\infty))$ as a coordinate transformation, and the transformed process $s(X_t)$ is a local martingale up to the explosion time T. Applying the martingale convergence theorem and the optional stopping theorem to $s(X_t)$ one obtains:

Theorem 8.3. *For any $\varepsilon, r \in (0,\infty)$ with $\varepsilon < x_0 < r$ we have:*

(1). The exit time $T_{\varepsilon,r} = \inf\{t \in [0,T) : X_t \notin (\varepsilon, r)\}$ *is almost surely less than* T.

(2). $P[T_\varepsilon < T_r] = P[X_{T_{\varepsilon,r}} = \varepsilon] = \dfrac{s(r) - s(x)}{s(r) - s(\varepsilon)}$.

The proof of Theorem 8.3 is left as an exercise.

Remark. (1). Note that any affine transformation $\widetilde{s}(x) = cs(x) + d$ with constants $c > 0$ and $d \in \mathbb{R}$ is also harmonic and strictly increasing, and hence a scale function. The ratio $(s(r) - s(x))/(s(r) - s(\varepsilon))$ is invariant under non-degenerate affine transformations of s.

(2). The scale function and the ruin probabilities depend only on the ratio $b(x)/\sigma(x)^2$.

Recurrence and asymptotics

We now apply the formula for the exit distributions in order to study the asymptotics of one-dimensional non-degenerate Itô diffusions as $t \nearrow T$. For $\varepsilon \in (0, x_0)$ we obtain

$$\begin{aligned} P[T_\varepsilon < T] &= P[\,T_\varepsilon < T_r \text{ for some } r \in (x_0, \infty)] \\ &= \lim_{r\to\infty} P[T_\varepsilon < T_r] = \lim_{r\to\infty} \frac{s(r) - s(x_0)}{s(r) - s(\varepsilon)}. \end{aligned}$$

In particular, we have

$$P[X_t = \varepsilon \text{ for some } t \in [0,T)] = P[T_\varepsilon < T] = 1$$

if and only if $s(\infty) = \lim_{r\nearrow\infty} s(r) = \infty$.
Similarly, one obtains for $r \in (x_0, \infty)$:

$$P[X_t = r \text{ for some } t \in [0,T)] = P[T_r < T] = 1$$

if and only if $s(0) = \lim_{\varepsilon\searrow 0} s(\varepsilon) = -\infty$.
Moreover,

$$P[X_t \to \infty \text{ as } t \nearrow T] = P\left[\bigcup_{\varepsilon>0} \bigcap_{r<\infty} \{T_r < T_\varepsilon\}\right] = \lim_{\varepsilon\searrow 0} \lim_{r\nearrow\infty} \frac{s(x_0) - s(\varepsilon)}{s(r) - s(\varepsilon)},$$

and

$$P[X_t \to 0 \text{ as } t \nearrow T] \;=\; P\left[\bigcup_{r<\infty}\bigcap_{\varepsilon>0}\{T_\varepsilon < T_r\}\right] \;=\; \lim_{r\nearrow\infty}\lim_{\varepsilon\searrow 0}\frac{s(x_0)-s(\varepsilon)}{s(r)-s(\varepsilon)}.$$

Summarizing, we have shown:

Corollary 8.4 (Asymptotics of one-dimensional Itô diffusions).

(1). If $s(0) = -\infty$ and $s(\infty) = \infty$, then the process (X_t) is recurrent, i.e.,

$$P[X_t = y \quad \text{for some } t \in [0,T)] \;=\; 1 \qquad \text{for any } x_0, y \in (0,\infty).$$

(2). If $s(0) > -\infty$ and $s(\infty) = \infty$ then $\lim_{t\nearrow T} X_t = 0$ almost surely.

(3). If $s(0) = -\infty$ and $s(\infty) < \infty$ then $\lim_{t\nearrow T} X_t = \infty$ almost surely.

(4). If $s(0) > -\infty$ and $s(\infty) < \infty$ then

$$P\left[\lim_{t\nearrow T} X_t = 0\right] \;=\; \frac{s(\infty)-s(x_0)}{s(\infty)-s(0)}$$

and

$$P\left[\lim_{t\nearrow T} X_t = \infty\right] \;=\; \frac{s(x_0)-s(0)}{s(\infty)-s(0)}$$

Intuitively, if $s(0) = -\infty$, in the natural scale the boundary is transformed to $-\infty$, which is not a possible limit for the local martingale $s(X_t)$, whereas otherwise $s(0)$ is finite and approached by $s(X_t)$ with strictly positive probability.

Example. Suppose that $b(x)/\sigma(x)^2 \approx \gamma x^{-1}$ as $x \nearrow \infty$ and $b(x)/\sigma(x)^2 \approx \delta x^{-1}$ as $x \searrow 0$ holds for $\gamma, \delta \in \mathbb{R}$ in the sense that $b(x)/\sigma(x)^2 - \gamma x^{-1}$ is integrable at ∞ and $b(x)/\sigma(x)^2 - \delta x^{-1}$ is integrable at 0. Then $s'(x)$ is of order $x^{-2\gamma}$ as $x \nearrow \infty$ and of order $x^{-2\delta}$ as $x \searrow 0$. Hence

$$s(\infty) \;=\; \infty \quad\iff\quad \gamma \le \frac{1}{2}, \qquad s(0) \;=\; -\infty \quad\iff\quad \delta \ge \frac{1}{2}.$$

In particular, recurrence holds if and only if $\gamma \le \frac{1}{2}$ and $\delta \ge \frac{1}{2}$.

More concrete examples will be studied below.

Remark (Explosion in finite time, Feller's test). Corollary 8.4 does not tell us whether the explosion time T is infinite with probability one. It can be shown that this is always the case if (X_t) is recurrent. In general, *Feller's test for explosions* provides a necessary and sufficient condition for the absence of explosion in finite time. The idea is to compute a function $g \in C(0,\infty)$ such that $e^{-t}g(X_t)$ is a local martingale and to apply the optional stopping theorem. The details are more involved than in the proof of corollary above, cf. e.g. Section 6.2 in [Durrett: Stochastic calculus].

Geometric Brownian motion

A geometric Brownian motion with parameters $\alpha \in \mathbb{R}$ and $\sigma > 0$ is a solution of the s.d.e.

$$dS_t \;=\; \alpha S_t\,dt + \sigma S_t\,dB_t. \tag{8.2.1}$$

We have already shown in the beginning of Section **??** that for $B_0 = 0$, the unique strong solution of (8.2.1) with initial condition $S_0 = x_0$ is

$$S_t \;=\; x_0 \cdot \exp\left(\sigma B_t + (\alpha - \sigma^2/2)t\right).$$

The distribution of S_t at time t is a **lognormal distribution**, i.e., the distribution of $c \cdot e^Y$ where c is a constant and Y is normally distributed. Moreover, one easily verifies that (S_t) is a time-homogeneous Markov process with log-normal transition densities

$$p_t(x,y) \;=\; \frac{1}{\sqrt{2\pi t\sigma^2}}\exp\left(-\frac{(\log(y/x)-\mu t)^2}{2t\sigma^2}\right), \qquad t,x,y>0,$$

where $\mu = \alpha - \sigma^2/2$. By the Law of Large Numbers for Brownian motion,

$$\lim_{t\to\infty} S_t \;=\; \begin{cases} +\infty & \text{if } \mu > 0 \\ 0 & \text{if } \mu < 0 \end{cases}.$$

If $\mu = 0$ then (S_t) is recurrent since the same holds for (B_t).
We now convince ourselves that we obtain the same results via the scale function:
The ratio of the drift and diffusion coefficient is

$$\frac{b(x)}{\sigma(x)^2} \;=\; \frac{\alpha x}{(\sigma x)^2} \;=\; \frac{\alpha}{\sigma^2 x},$$

and hence

$$s'(x) \;=\; \text{const.} \cdot \exp\left(-\int_{x_0}^{x} \frac{2\alpha}{\sigma^2 y}\,dy\right) \;=\; \text{const.} \cdot x^{-2\alpha/\sigma^2}.$$

Therefore,

$$s(\infty) \;=\; \infty \quad\iff\quad 2\alpha/\sigma^2 \le 1, \qquad s(0) \;=\; \infty \quad\iff\quad 2\alpha/\sigma^2 \ge 1,$$

which again shows that $S_t \to \infty$ for $\alpha > \sigma^2/2$, $S_t \to 0$ for $\alpha < \sigma^2/2$, and S_t is recurrent for $\alpha = \sigma^2/2$.

Feller's branching diffusion

Our second growth model is described by the stochastic differential equation

$$dX_t \;=\; \beta X_t\,dt + \sigma\sqrt{X_t}\,dB_t, \qquad X_0 = x_0, \tag{8.2.2}$$

with given constants $\beta \in \mathbb{R}, \sigma > 0$, and values in $\mathbb{R}_+$. Note that in contrast to the equation of geometric Brownian motion, the multiplicative factor $\sqrt{X_t}$ in the noise term is not a linear function of X_t. As a consequence, there is no explicit formula for a solution of (8.2.2). Nevertheless, a general existence result guarantees the existence of a strong solution defined up to the explosion time

$$T \;=\; \sup_{\varepsilon, r > 0} T_{\mathbb{R}\setminus(\varepsilon,r)},$$

cf. **??**. SDEs similar to (8.2.2) appear in various applications.

Example (Diffusion limits of branching processes). We consider a Galton-Watson branching process Z_t^h with time steps $t = 0, h, 2h, 3h, \ldots$ of size $h > 0$, i.e., Z_0^h is a given initial population size, and

$$Z_{t+h}^h \;=\; \sum_{i=1}^{Z_t^h} N_i, t/h \quad \text{ for } t = k\cdot h, k = 0, 1, 2, \ldots,$$

with independent identically distributed random variables $N_{i,k}, i \ge 1, k \ge 0$. The random variable Z_{kh}^h describes the size of a population in the k-th generation when $N_{i,l}$

is the number of offspring of the i-th individual in the l-th generation. We assume that the mean and the variance of the offspring distribution are given by

$$E[N_{i,l}] \;=\; 1+\beta h \quad \text{and} \quad \mathrm{Var}[N_{i,l}] \;=\; \sigma^2$$

for finite constants $\beta, \sigma \in \mathbb{R}$.

We are interested in a scaling limit of the model as the size h of time steps goes to 0. To establish convergence to a limit process as $h \searrow 0$ we rescale the population size by h, i.e., we consider the process

$$X_t^h \;:=\; h \cdot Z_{\lfloor t \rfloor}^h, \qquad t \in [0, \infty).$$

The mean growth ("drift") of this process in one time step is

$$E[X_{t+h}^h - X_t^h \mid \mathcal{F}_t^h] \;=\; h \cdot E[Z_{t+h}^h - Z_t^h \mid \mathcal{F}_t^h] \;=\; h\eta h Z_t^h \;=\; h\beta X_t^h,$$

and the corresponding condition variance is

$$\mathrm{Var}[X_{t+h}^h - X_t^h \mid \mathcal{F}_t^h] \;=\; h^2 \cdot \mathrm{Var}[Z_{t+h}^h - Z_t^h \mid \mathcal{F}_t^h] \;=\; h^2\sigma^2 Z_t^h \;=\; h\sigma^2 X_t^h,$$

where $\mathcal{F}_t^h \;=\; \sigma(N_{i,l} \mid i \geq 1, 0 \leq l \leq k)$ for $t = k \cdot h$. Since both quantities are of order $O(h)$, we can expect a limit process (X_t) as $h \searrow 0$ with drift coefficient $\beta \cdot X_t$ and diffusion coefficient $\sqrt{\sigma^2 X_t}$, i.e., the scaling limit should be a diffusion process solving a s.d.e. of type (8.2.2). A rigorous derivation of this diffusion limit can be found e.g. in Section 8 of [Durrett: Stochastic Calculus].

We now analyze the asymptotics of solutions of (8.2.2). The ratio of drift and diffusion coefficient is $\beta x/(\sigma\sqrt{x})^2 = \beta/\sigma$, and hence the derivative of a scale function is

$$s'(x) \;=\; \text{const.} \cdot \exp(-2\beta x/\sigma).$$

Thus $s(0)$ is always finite, and $s(\infty) = \infty$ if and only if $\beta \leq 1$. Therefore, by Corollary 8.4, in the subcritical and critical case $\beta \leq 1$, we obtain

$$\lim_{t \nearrow T} X_t \;=\; 0 \qquad \text{almost surely,}$$

whereas in the supercritical case $\beta > 1$,

$$P\left[\lim_{t\nearrow T} X_t = 0\right] > 0 \quad \text{and} \quad P\left[\lim_{t\nearrow T} X_t = \infty\right] > 0.$$

This corresponds to the behaviour of Galton-Watson processes in discrete time. It can be shown by Feller's boundary classification for one-dimensional diffusion processes that if $X_t \to 0$ then the process actually dies out almost surely in finite time, cf. e.g. Section 6.5 in [Durrett: Stochastic Calculus]. On the other hand, for trajectories with $X_t \to \infty$, the explosion time T is almost surely infinite and X_t grows exponentially as $t \to \infty$.

Cox-Ingersoll-Ross model

The CIR model is a model for the stochastic evolution of interest rates or volatilities. The equation is

$$dR_t \;=\; (\alpha - \beta R_t)\, dt + \sigma\sqrt{R_t}\, dB_t \qquad R_0 = x_0, \tag{8.2.3}$$

with a one-dimensional Brownian motion (B_t) and positive constants $\alpha, \beta, \sigma > 0$. Although the s.d.e. looks similar to the equation for Feller's branching diffusion, the behaviour of the drift coefficient near 0 is completely different. In fact, the idea is that the positive drift α pushes the process away from 0 so that a recurrent process on $(0, \infty)$ is obtained. We will see that this intuition is true for $\alpha \geq \sigma^2/2$ but not for $\alpha < \sigma^2/2$. Again, there is no explicit solution for the s.d.e. (8.13), but existence of a strong solution holds. The ratio of the drift and diffusion coefficient is $(\alpha - \beta x)/\sigma^2 x$, which yields

$$s'(x) \;=\; \text{const.} \cdot x^{-2\alpha/\sigma^2} e^{2\beta x/\sigma^2}.$$

Hence $s(\infty) = \infty$ for any $\beta > 0$, and $s(0) = \infty$ if and only if $2\alpha \geq \sigma^2$. Therefore, the CIR process is recurrent if and only if $\alpha \geq \sigma^2/2$, whereas $X_t \to 0$ as $t \nearrow T$ almost surely otherwise.

By applying Itô's formula one can now prove that X_t has finite moments, and compute the expectation and variance explicitly. Indeed, taking expectation values in the s.d.e.

$$R_t \;=\; x_0 + \int_0^t (\alpha - \beta R_s)\, ds + \int_0^t \sigma\sqrt{R_s}\, dB_s,$$

we obtain informally

$$\frac{d}{dt}E[R_t] \;=\; \alpha - \beta E[R_t],$$

and hence by variation of constants,

$$E[R_t] \;=\; x_0 \cdot e^{-\beta t} + \frac{\alpha}{\beta}(1 - e^{-\beta t}).$$

To make this argument rigorous requires proving that the local martingale $t \mapsto \int_0^t \sigma\sqrt{R_s}dB_s$ is indeed a martingale:

Exercise. Consider a strong solution $(R_t)_{t\geq 0}$ of (8.13) for $\alpha \geq \sigma^2/2$.

(1). Show by applying Itô's formula to $x \mapsto |x|^p$ that $E[|R_t|^p] < \infty$ for any $t \geq 0$ and $p \geq 1$.

(2). Compute the expectation of R_t, e.g. by applying Itô's formula to $e^{\beta t}x$.

(3). Proceed in a similar way to compute the variance of R_t. Find its asymptotic value $\lim_{t\to\infty} \mathrm{Var}[R_t]$.

8.3 Linear SDE with additive noise

We now consider stochastic differential equations of the form

$$dX_t \;=\; \beta_t C_t\, dt + \sigma_t\, dB_t, \qquad X_0 = x, \tag{8.3.1}$$

where (B_t) is a Brownian motion, and the coefficients are *deterministic* continuous functions $\beta, \sigma : [0, \infty) \to \mathbb{R}$. Hence the drift term $\beta_t X_t$ is linear in X_t, and the diffusion coefficient does not depend on X_t, i.e., the noise increment $\sigma_t\, dB_t$ is proportional to white noise dB_t with a proportionality factor that does not depend on X_t.

Variation of constants

An explicit strong solution of the SDE (8.3.1) can be computed by a "variation of constants" Ansatz. We first note that the general solution in the deterministic case $\sigma_t \equiv 0$ is given by

$$X_t \;=\; \text{const.} \cdot \exp\left(\int_0^t \beta_s\,ds\right).$$

To solve the SDE in general we try the ansatz

$$X_t \;=\; C_t \cdot \exp\left(\int_0^t \beta_s\,ds\right)$$

with a continuous Itô process (C_t) driven by the Brownian motion (B_t). By the Itô product rule,

$$dX_t \;=\; \beta_t X_t\,dt + \exp\left(\int_0^t \beta_s\,ds\right)\,dC_t.$$

Hence (X_t) solves (8.3.1) if and only if

$$dC_t \;=\; \exp\left(-\int_0^t \beta_s\,ds\right)\sigma_t\,dB_t,$$

i.e.,

$$C_t \;=\; C_0 + \int_0^t \exp\left(-\int_0^r \beta_s\,ds\right)\sigma_r\,dB_r.$$

We thus obtain:

Theorem 8.5. *The almost surely unique strong solution of the SDE (8.3.1)with initial value x is given by*

$$X_t^x \;=\; x \cdot \exp\left(-\int_0^t \beta_s\,ds\right) + \int_0^t \exp\left(\int_r^t \beta_s\,ds\right)\sigma_r\,dB_r.$$

Note that the theorem not only yields an explicit solution but it also shows that the solution depends smoothly on the initial value x. The effect of the noise on the solution is additive and given by a Wiener-Itô integral, i.e., an Itô integral with deterministic integrand. The average value

$$E[X_t^x] \;=\; x \cdot \exp\left(\int_0^t B_s\,ds\right), \tag{8.3.2}$$

coincides with the solution in the absence of noise, and the mean-square deviation from this solution due to random perturbation of the equation is

$$\mathrm{Var}[X_t^x] \;=\; \mathrm{Var}\left[\int_0^t \exp\left(\int_r^t \beta_s\,ds\right)\sigma_r\,dB_r\right] \;=\; \int_0^t \exp\left(2\int_r^t \beta_s\,ds\right)\sigma_r^2\,dr$$

by the Itô isometry.

Solutions as Gaussian processes

We now prove that the solution (X_t) of a linear s.d.e. with additive noise is a Gaussian process. We first observe that X_t is normally distributed for any $t \ge 0$.

Lemma 8.6. *For any deterministic function $h \in L^2(0,t)$, the Wiener-Itô integral $I_t = \int_0^t h_s\,dB_s$ is normally distributed with mean 0 and variance $\int_0^t h_s^2\,ds$.*

Proof. Suppose first that $h = \sum_{i=0}^{n-1} c_i \cdot I_{(t_i,t_{i+1}]}$ is a step function with $n \in \mathbb{N}$, $c_1,\ldots,c_n \in \mathbb{R}$, and $0 \le t_0 < t_1 < \ldots < t_n$. Then $I_t = \sum_{i=0}^{n-1} c_i \cdot (B_{t_{i+1}} - B_{t_i})$ is normally distributed with mean zero and variance

$$\mathrm{Var}[I_t] \;=\; \sum_{i=0}^{n-1} c_i^2 (t_{i+1} - t_i) \;=\; \int_0^t h_s^2\,ds.$$

In general, there exists a sequence $(h^{(n)})_{n\in\mathbb{N}}$ of step functions such that $h^{(n)} \to h$ in $L^2(0,t)$, and

$$I_t \;=\; \int_0^t h\,dB \;=\; \lim_{n\to\infty} \int_0^t h^{(n)}\,dB \qquad \text{in } L^2(\Omega,\mathcal{A},P).$$

Hence I_t is again normally distributed with mean zero and

$$\mathrm{Var}[I_t] \;=\; \lim_{n\to\infty} \mathrm{Var}\left[\int_0^t h^{(n)}\,dB\right] \;=\; \int_0^t h^2\,ds.$$

□

Theorem 8.7 (Wiener-Itô integrals are Gaussian processes). *Suppose that $h \in L^2_{loc}([0,\infty),\mathbb{R})$. Then $I_t = \int_0^t h_s\,dB_s$ is a continuous Gaussian process with*

$$E[I_t] \;=\; 0 \quad \textit{and} \quad \mathrm{Cov}[I_t, I_s] \;=\; \int_0^{t\wedge s} h_r^2\,ds \qquad \textit{for any } t,s \geq 0.$$

Proof. Let $0 \leq t_1 < \ldots < t_n$. To show that $(I_{t_1}, \ldots, I_{t_n})$ has a normal distribution it suffices to prove that any linear combination of the random variables $I_{t_1}, \ldots, I_{t_n}$ is normally distributed. This holds true since any linear combination is again an Itô integral with deterministic integrand:

$$\sum_{i=1}^n \lambda_i I_{t_i} \;=\; \int_0^{t_n} \sum_{i=1}^n \lambda_i \cdot I_{(0,t_i)}(s) h_s\,dB_s$$

for any $n \in \mathbb{N}$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Hence (I_t) is a Gaussian process with $E[I_t] = 0$ and

$$\begin{aligned}
\mathrm{Cov}[I_t, I_s] &= E[I_t I_s] \\
&= E\left[\int_0^\infty h_r \cdot I_{(0,t)}(r)\,dB_r \int_0^\infty h_r \cdot I_{(0,s)}(r)\,dB_r\right] \\
&= (h \cdot I_{(0,t)}, h \cdot I_{(0,s)})_{L^2(0,\infty)} \\
&= \int_0^{s\wedge t} h_r^2\,dr.
\end{aligned}$$

□

Example (Brownian motion). If $h \equiv 1$ then $I_t = B_t$. The Brownian motion (B_t) is a centered Gaussian process with $\mathrm{Cov}[B_t, B_s] = t \wedge s$.

More generally, by Theorem 8.7 and Theorem 8.5, any solution (X_t) of a linear SDE with additive noise and deterministic (or Gaussian) initial value is a continuous Gaussian process. In fact by (8.3.1), the marginals of (X_t) are affine functions of the corresponding marginals of a Wiener-Itô integral:

$$X_t^x = \frac{1}{h_t} \cdot \left(x + \int_0^t h_r \sigma_r \, dB_r \right) \quad \text{with} \quad h_r = \exp\left(- \int_0^r \beta_u \, du \right).$$

Hence all finite dimensional marginals of (X_t^x) are normally distributed with

$$E[X_t^x] = x/H_t \quad \text{and} \quad \mathrm{Cov}[X_t^x, X_s^x] = \frac{1}{h_t h_s} \cdot \int_0^{t \wedge s} h_r^2 \sigma_r^2 \, dr.$$

The Ornstein-Uhlenbeck process

In 1905, Einstein introduced a model for the movement of a "big" particle in a fluid. Suppose that V_t^{abs} is the absolute velocity of the particle, $\overline{V}_t$ is the mean velocity of the fluid molecules and $V_t = V_t^{\text{abs}} - \overline{V}_t$ is the velocity of the particle relative to the fluid. Then the velocity approximatively can be described as a solution to an s.d.e.

$$dV_t = -\gamma V_t \, dt + \sigma dB_t. \tag{8.3.3}$$

Here (B_t) is a Brownian motion in $\mathbb{R}^d$, $d = 3$, and γ, σ are strictly positive constants that describe the damping by the viscosity of the fluid and the magnitude of the random collisions. A solution to the s.d.e. (8.3.3) is called an **Ornstein-Uhlenbeck process**. Although it has first been introduced as a model for the velocity of physical Brownian motion, the Ornstein-Uhlenbeck process is a fundamental stochastic process that is almost as important as Brownian motion for mathematical theory and stochastic modeling. In particular, it is a continuous-time analogue of an AR(1) autoregressive process. Note that (8.3.3) is a system of d decoupled one-dimensional stochastic differential equations $dV_t^{(i)} = -\gamma V_t^{(i)} + \sigma dB_t^{(i)}$. Therefore, we will assume w.l.o.g. $d = 1$. By the considerations above, the one-dimensional Ornstein-Uhlenbeck process is a continuous

Gaussian process. The unique strong solution of the s.d.e. (8.3.3) with initial condition x is given explicitly by

$$V_t^x = e^{-\gamma t}\left(x + \sigma \int_0^t e^{\gamma s}\, dB_s\right). \tag{8.3.4}$$

In particular,

$$E[V_t^x] = e^{-\gamma t} x,$$

and

$$\begin{aligned}\mathrm{Cov}[V_t^x, V_s^x] &= e^{-\gamma(t+s)}\sigma^2 \int_0^{t\wedge s} e^{2\gamma r}\, dr \\ &= \frac{\sigma^2}{2\gamma}(e^{-\gamma|t-s|} - e^{-\gamma(t+s)}) \qquad \text{for any } t,s \geq 0.\end{aligned}$$

Note that as $t \to \infty$, the effect of the initial condition decays exponentially fast with rate γ. Similarly, the correlations between V_t^x and V_s^x decay exponentially as $|t-s| \to \infty$. The distribution at time t is

$$V_t^x \sim N\left(e^{-\gamma t}x,\ \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t})\right). \tag{8.3.5}$$

In particular, as $t \to \infty$

$$V_t^x \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2}{2\gamma}\right).$$

One easily verifies that $N(0, \sigma^2/2\gamma)$ is an *equilibrium* for the process: If $V_0 \sim N(0, \sigma^2/2\gamma)$ and (B_t) is independent of V_0 then

$$\begin{aligned}V_t &= e^{-\gamma t}V_0 + \sigma \int_0^t e^{\gamma(s-t)}\, dB_s \\ &\sim N\left(0, \frac{\sigma^2}{2\gamma}e^{-2\gamma t} + \sigma^2 \int_0^t e^{2\gamma(s-t)}\, ds\right) = N(0, \sigma^2/2\gamma)\end{aligned}$$

for any $t \geq 0$.

Theorem 8.8. *The Ornstein-Uhlenbeck process* (V_t^x) *is a time-homogeneous Markov process w.r.t. the filtration* $(\mathcal{F}_t^{B,P})$ *with stationary distribution* $N(0,\sigma^2/2\gamma)$ *and transition probabilities*

$$p_t(x,A) \;=\; P\left[e^{-\gamma t}x + \frac{\sigma}{\sqrt{2\gamma}}\sqrt{1-e^{-2\gamma t}}Z \in A\right], \qquad Z \sim N(0,1).$$

Proof. We first note that by (8.3.5),

$$V_t^x \;\sim\; e^{-\gamma t}x + \frac{\sigma}{\sqrt{2\gamma}}\sqrt{1-e^{-2\gamma t}}Z \qquad \text{for any } t \geq 0$$

with $Z \sim N(0,1)$. Hence,

$$E[f(V_t^x)] \;=\; (p_t f)(x)$$

for any non-negative measurable function $f : \mathbb{R} \to \mathbb{R}$. We now prove a *pathwise counterpart to the Markov property*: For $t, r \geq 0$, by (8.3.4)

$$\begin{aligned} V_{t+r}^x \;&=\; e^{-\gamma(t+r)}\left(x + \sigma\int_0^t e^{\gamma s}\,dB_s\right) + \sigma\int_0^{t+r} e^{\gamma(s-t-r)}\,dB_s \\ &=\; e^{-\gamma r}V_t^x + \sigma\int_0^r e^{\gamma(u-r)}\,d\overline{B}_u, \end{aligned} \tag{8.3.6}$$

where $\overline{B}_u := B_{t+u} - B_t$ is a Brownian motion that is independent of $\mathcal{F}_t^{B,P}$. Hence, the random variable $\sigma \cdot \int_0^r e^{\gamma(u-r)}\,d\overline{B}_u$ is also independent of $\mathcal{F}_t^{B,P}$ and, by (8.3.4), it has the same distribution as the Ornstein-Uhlenbeck process with initial condition 0:

$$\sigma \cdot \int_0^r e^{\gamma(u-r)}\,d\overline{B}_u \quad\sim\quad V_r^0.$$

Therefore, by (8.3.6), the conditional distribution of V_{t+r}^x given $\mathcal{F}_t^{B,P}$ coincides with the distribution of the process with initial V_t^x at time r:

$$\begin{aligned} E[f(V_{t+r}^x)\,|\,\mathcal{F}_t^{B,P}] \;&=\; E[f(e^{-\gamma r}V_t^x(\omega) + V_r^0)] \\ &=\; E[f(V_r^{V_t^x(\omega)})] \;=\; (p_r f)(V_t^x(\omega)) \qquad \text{for } P\text{-a.e. } \omega. \end{aligned}$$

This proves that (V_t^x) is a Markov process with transition kernels p_r, $r \geq 0$. ☐

Remark. The pathwise counterpart of the Markov property used in the proof above is called **cocycle property** of the stochastic flow $x \mapsto V_t^x$.

The Itô-Doeblin formula can now be used to identify the generator of the Ornstein-Uhlenbeck process: Taking expectation values, we obtain the forward equation

$$E[F(V_t^x)] \;=\; F(x) + \int_0^t E[(\mathscr{L}F)(V_s^x)]\,ds$$

for any function $F \in C_0^2(\mathbb{R})$ and $t \geq 0$, where

$$(\mathscr{L}F)(x) \;=\; \frac{1}{2}\sigma^2 f''(x) - \gamma x f'(x).$$

For the transition function this yields

$$(p_t F)(x) \;=\; F(x) + \int_0^t (p_s \mathscr{L}F)(x) \qquad \text{for any } x \in \mathbb{R},$$

whence

$$\lim_{t \searrow 0} \frac{(p_t f)(x) - f(x)}{t} \;=\; \lim_{t \searrow 0} \frac{1}{t} \int_0^t E[(\mathscr{L}f)(V_s^x)]\,ds \;=\; (\mathscr{L}f)(x)$$

by continuity and dominated convergence. This shows that the infinitesimal generator of the Ornstein-Uhlenbeck process is an extension of the operator $(\mathscr{L}, C_0^2(\mathbb{R}))$.

Change of time-scale

We will now prove that Wiener-Itô integrals can also be represented as Brownian motion with a coordinate transformation on the time axis. Hence solutions of one-dimensional linear SDE with additive noise are affine functions of time changed Brownian motions. We first note that a Wiener-Itô integral $I_t = \int_0^t h_r\,dB_r$ with $h \in L^2_{\mathrm{loc}}(0,\infty)$ is a continuous centered Gaussian process with covariance

$$\mathrm{Cov}[I_t, I_s] \;=\; \int_0^{t\wedge s} h_r^2\,dr \;=\; \tau(t) \wedge \tau(s)$$

where

$$\tau(t) := \int_0^t h_r^2 \, dr = \mathrm{Var}[I_t]$$

is the corresponding variance process. The variance process should be thought of as an "internal clock" for the process (I_t). Indeed, suppose $h > 0$ almost everywhere. Then τ is strictly increasing and continuous, and

$$\tau : [0, \infty) \rightarrow [0, \tau(\infty)) \qquad \text{is a homeomorphism.}$$

Transforming the time-coordinate by τ, we have

$$\mathrm{Cov}[I_{\tau^{-1}(t)}, I_{\tau^{-1}(s)}] = t \wedge s \qquad \text{for any } t, s \in [0, \tau(\infty)].$$

These are exactly the covariance of a Brownian motion. Since a continuous Gaussian process is uniquely determined by its expectations and covariances, we can conclude:

Theorem 8.9 (Wiener-Itô integrals as time changed Brownian motions). *The process $\widetilde{B}_s := I_{\tau^{-1}(s)}, \quad 0 \leq s < \tau(\infty)$, is a Brownian motion, and*

$$I_t = \widetilde{B}_{\tau(t)} \qquad \textit{for any } t \geq 0, \ P\textit{-almost surely.}$$

Proof. Since $(\widetilde{B}_s)_{0 \leq s < \tau(\infty)}$ has the same marginal distributions as the Wiener-Itô integral $(I_t)_{t \geq 0}$ (but at different times), $(\widetilde{B}_s)$ is again a continuous centered Gaussian process. Moreover, $\mathrm{Cov}[B_t, B_s] = t \wedge s$, so that (B_s) is indeed a Brownian motion. □

Note that the argument above is different from previous considerations in the sense that the Brownian motion $(\widetilde{B}_s)$ is constructed from the process (I_t) and not vice versa.
This means that we can not represent (I_t) as a time-change of a given Brownian motion (e.g. (B_t)) but we can only show that there exists a Brownian motion $(\widetilde{B}_s)$ such that I is a time-change of $\widetilde{B}$. This way of representing stochastic processes w.r.t. Brownian motions that are constructed from the process corresponds to the concept of weak solutions of stochastic differential equations, where driving Brownian motion is not given a

priori. We return to these ideas in Section 9, where we will also prove that continuous local martingales can be represented as time-changed Brownian motions.

Theorem 8.9 enables us to represent solution of linear SDE with additive noise by time-changed Brownian motions. We demonstrate this with an example: By the explicit formula (8.3.4) for the solution of the Ornstein-Uhlenbeck SDE, we obtain:

Corollary 8.10 (Mehler formula). *A one-dimensional Ornstein-Uhlenbeck process V_t^x with initial condition x can be represented as*

$$V_t^x = e^{-\gamma t}(x + \sigma \widetilde{B}_{\frac{1}{2\gamma}(e^{2\gamma t}-1)})$$

with a Brownian motion $(\widetilde{B}_t)_{t\geq 0}$ such that $\widetilde{B}_0 = 0$.

Proof. The corresponding time change for the Wiener-Itô integral is given by

$$\tau(t) = \int_0^t \exp(2\gamma s)\, ds = (\exp(2\gamma t) - 1)/2\gamma.$$

□

8.4 Brownian bridge

In many circumstances one is interested in conditioning diffusion process on taking a given value at specified times. A basic example is the Brownian bridge which is Brownian motion conditioned to end at a given point x after time t_0. We now present several ways to describe and characterize Brownian bridges. The first is based on the Wiener-Lévy construction and specific to Brownian motion, the second extends to Gaussian processes, whereas the final characterization of the bridge process as the solution of a time-homogeneous SDE can be generalized to other diffusion processes. From now on, we consider a one-dimensional Brownian motion $(B_t)_{0\leq t\leq 1}$ with $B_0 = 0$ that we would like to condition on taking a given value y at time 1

Wiener-Lévy construction

Recall that the Brownian motion (B_t) has the Wiener-Lévy representation

$$B_t(\omega) \;=\; Y(\omega)t + \sum_{n=0}^{\infty} \sum_{k=0} 2^n - 1 Y_{n,k}(\omega) e_{n,k}(t) \qquad \text{for } t \in [0,1] \tag{8.4.1}$$

where $e_{n,k}$ are the Schauder functions, and Y and $Y_{n,k}$ $(n \geq 0, k = 0, 1, 2, \ldots, 2^n - 1)$ are independent and standard normally distributed. The series in (8.4.1) converges almost surely uniformly on $[0,1]$, and the approximating partial sums are piecewise linear approximations of B_t. The random variables $Y = B_1$ and

$$X_t \;:=\; \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Y_{n,k} e_{n,k}(t) \;=\; B_t - tB_1$$

are independent. This suggests that we can construct the bridge by replacing $Y(\omega)$ by the constant value y. Let

$$X_t^y \;:=\; yt + X_t \;=\; B_t + (y - B_1) \cdot t,$$

and let μ_y denote the distribution of the process $(X_t^y)_{0 \leq t \leq 1}$ on $C([0,1])$. The next theorem shows that X_t^y is indeed a Brownian motion conditioned to end at y at time 1:

Theorem 8.11. *The map $y \mapsto \mu_y$ is a regular version of the conditional distribution of $(B_t)_{0 \leq t \leq 1}$ given B_1, i.e.,*

(1). μ_y is a probability measure on $C([0,1])$ for any $y \in \mathbb{R}$,

(2). $P[(B_t)_{0 \leq t \leq 1} \in A \mid B_1] \;=\; \mu_{B_1}[A]$ holds P-almost surely for any given Borel subset $A \subseteq C([0,1])$.

(3). If $F : C([0,1]) \to \mathbb{R}$ is a bounded and continuous function (w.r.t. the supremum norm on $C([0,1])$) then the map $y \mapsto \int F \, d\mu_y$ is continuous.

The last statement says that $<\mapsto \mu_y$ is a continuous function w.r.t. the topology of weak convergence.

Proof. By definition, μ_y is a probability measure for any $y \in \mathbb{R}$. Moreover, for any Borel set $A \subseteq C([0,1])$,

$$\begin{aligned} P[(B_t)_{0\le t\le 1} \in A \mid B_1](\omega) &= P[(X_t + tB_1) \in A \mid B_1](\omega) \\ &= P[(X_t + tB_1(\omega)) \in A] \;=\; P[(X_t^{B_t(\omega)}) \in A] \;=\; \mu_{B_1(\omega)}[A] \end{aligned}$$

for P-almost every ω by independence of (X_T) and B_1. Finally, if $F : C([0,1]) \to \mathbb{R}$ is continuous and bounded then

$$\int F\, d\mu_y \;=\; E[F((y_t + X_t)_{0\le t\le 1})]$$

is continuous in y by dominated convergence. □

Finite-dimensional distributions

We now compute the marginals of the Brownian bridge X_t^y:

Corollary 8.12. *For any $n \in \mathbb{N}$ and $0 < t_1 < \ldots < t_n < 1$, the distribution of $(X_{t_1}^y, \ldots, X_{t_n}^y)$ on $\mathbb{R}^n$ is absolutely continuous with density*

$$f_y(x_1,\ldots,x_n) \;=\; \frac{p_{t_1}(0,x_1)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n)p_{1-t_n}(x_n,y)}{p_1(0,y)}. \tag{8.4.2}$$

Proof. The distribution of $(B_{t_1},\ldots,B_{t_n},B_1)$ is absolutely continuous with density

$$f_{B_{t_1},\ldots,B_{t_n},B_1}(x_1,\ldots,x_n,y) = p_{t_1}(0,x_0)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n)p_{1-t_n}(x_n,y).$$

Since the distribution of $(X_{t_1}^y,\ldots,X_{t_n}^y)$ is a regular version of the conditional distribution of $(B_{t_1},\ldots,B_{t_n})$ given B_1, it is absolutely continuous with the conditional density

$$\begin{aligned} f_{B_{t_1},\ldots,B_{t_n}|B_1}(x_1,\ldots,x_n|y) &= \frac{f_{B_{t_1},\ldots,B_{t_n},B_1}(x_1,\ldots,x_n,y)}{\int\cdots\int f_{B_{t_1},\ldots,B_{t_n},B_1}(x_1,\ldots,x_n,y)\,dx_1\cdots dx_n} \\ &= f_y(x_1,\ldots,x_n). \end{aligned}$$

□

In general, any almost surely continuous process on $[0,1]$ with marginals given by (8.4.2) is called a **Brownian bridge from** 0 **to** y **in time** 1. A Brownian bridge from x to y in time t is defined correspondingly for any $x, y \in \mathbb{R}$ and any $t > 0$. In fact, this definition of the bridge process in terms of the marginal distributions carries over from Brownian motion to arbitrary Markov processes with strictly positive transition densities. In the case of the Brownian bridge, the marginals are again normally distributed:

Theorem 8.13 (**Brownian bridge as a Gaussian process**). *The Brownian bridge from* 0 *to* y *in time* 1 *is the (in distribution unique) continuous Gaussian process* $(X_t^y)_{t\in[0,1]}$ *with*

$$E[X_t^y] \;=\; ty \quad \textit{and} \quad \mathrm{Cov}[X_t^y, X_s^y] \;=\; t \wedge s - ts \qquad \textit{for any } s,t \in [0,1]. \quad (8.4.3)$$

Proof. A continuous Gaussian process is determined uniquely in distribution by its means and covariances. Therefore, it suffices to show that the bridge $X_t^y = B_t + (y - B_1)t$ defined above is a continuous Gaussian process such that (8.4.3) holds. This holds true: By (8.4.2), the marginals are normally distributed, and by definition, $t \mapsto X_t^y$ is almost surely continuous. Moreover,

$$\begin{aligned} E[X_t^y] &= E[B_t] + E[y - B_1] \cdot t \;=\; yt, \qquad \text{and} \\ \mathrm{Cov}[X_t^y, X_s^y] &= \mathrm{Cov}[B_t, B_s] - t \cdot \mathrm{Cov}[B_1, B_s] - s \cdot \mathrm{Cov}[B_t, B_1] + ts\,\mathrm{Var}[B_1] \\ &= t \wedge s - ts - st + ts \;=\; t \wedge s - ts. \end{aligned}$$

□

Remark (**Covariance as Green function, Cameron-Martin space**). The covariances of the Brownian bridge are given by

$$c(t,s) \;=\; \mathrm{Cov}[X_t^y, X_s^y] \;=\; \begin{cases} t \cdot (1-s) & \text{for } t \le s, \\ (1-t) \cdot s & \text{for } t \ge s. \end{cases}$$

The function $c(t,s)$ is the Green function of the operator d^2/dt^2 with Dirichlet boundary conditions on the interval $[0,1]$. This is related to the fact that the distribution of the

Brownian bridge from 0 to 0 can be viewed as a standard normal distribution on the space of continuous paths $\omega : [0,1] \to \mathbb{R}$ with $\omega(0) = \omega(1) = 0$ w.r.t. the Cameron-Martin inner product

$$(g,h)_H \;=\; \int_0^1 g'(s)h'(s)\,ds.$$

The second derivative d^2/dt^2 is the linear operator associated with this quadratic from.

SDE for the Brownian bridge

Our construction of the Brownian bridge by an affine transformation of Brownian motion has two disadvantages:

- It can not be carried over to more general diffusion processes with possibly non-linear drift and diffusion coefficients.
- The bridge $X_t^y = B_t + t(y - B_1)$ does not depend on (B_t) in an adapted way, because the terminal value B_1 is required to define X_t^y for any $t > 0$.

We will now show how to construct a Brownian bridge from a Brownian motion in an adapted way. The idea is to consider an SDE w.r.t. the given Brownian motion with a drift term that forces the solution to end at a given point at time 1. The size of the drift term will be large if the process is still far away from the given terminal point at a time close to 1. For simplicity we consider a bridge (X_t) from 0 to 0 in time 1. Brownian bridges with other end points can be constructed similarly. Since the Brownian bridge is a Gaussian process, we may hope that there is a linear stochastic differential equation with additive noise that has a Brownian bridge as a solution. We therefore try the Ansatz

$$dX_t \;=\; -\beta_t X_t\,dt + dB_t, \qquad X_0 = 0 \tag{8.4.4}$$

with a given continuous deterministic function $\beta_t, 0 \le t < 1$. By variation of constants, the solution of (8.4.4) is the Gaussian process $X_t, 0 \le t < 1$, given by

$$X_t \;=\; \frac{1}{h_t}\int_0^t h_r\,dB_r \quad \text{where} \quad h_t \;=\; \exp\left(\int_0^t \beta_s\,ds\right).$$

The process (X_t) is centered and has covariances

$$\mathrm{Cov}[X_t, X_s] \;=\; \frac{1}{h_t h_s} \int_0^{t\wedge s} h_r^2 \, dr.$$

Therefore, (X_t) is a Brownian bridge if and only if

$$\mathrm{Cov}[X_t, X_s] \;=\; t \cdot (1-s) \qquad \text{for any } t \le s,$$

i.e., if and only if

$$\frac{1}{t h_t} \int_0^t h_r^2 \, dr \;=\; h_s \cdot (1-s) \qquad \text{for any } 0 < t \le s. \tag{8.4.5}$$

The equation (8.4.5) holds if and only if h_t is a constant multiple of $1/1-t$, and in this case

$$\beta_t \;=\; \frac{d}{dt} \log h_t \;=\; \frac{h_t'}{h_t} \;=\; \frac{1}{1-t} \qquad \text{for } t \in [0,1].$$

Summarizing, we have shown:

Theorem 8.14. *If (B_t) is a Brownian motion then the process (X_t) defined by*

$$X_t \;=\; \int_0^t \frac{1-t}{1-r} \, dB_r \qquad \textit{for } t \in [0,1], \quad X_1 = 0,$$

is a Brownian bridge from 0 to 0 in time 1. It is the unique continuous process solving the SDE

$$dX_t \;=\; -\frac{X_t}{1-t} \, dt + \, dB_t \qquad \textit{for } t \in [0,1). \tag{8.4.6}$$

Proof. As shown above, $(X_t)_{t\in[0,1)}$ is a continuous centered Gaussian process with the covariances of the Brownian bridge. Hence its distribution on $C([0,1))$ coincides with that of the Brownian bridge from 0 to 0. In particular, this implies $\lim_{t\nearrow 1} X_t = 0$ almost surely, so the trivial extension from $[0,1)$ to $[0,1]$ defined by $X_1 = 0$ is a Brownian bridge. $\square$

If the Brownian bridge is replaced by a more general conditioned diffusion process, the Gaussian characterization does not apply. Nevertheless, it can still be shown by different means (the keyword is “h-transform”) that the bridge process solves an SDE generalizing (8.4.6), cf. **??** below.

8.5 Stochastic differential equations in $\mathbb{R}^{\mathrm{n}}$

We now explain how to generalize our considerations to systems of stochastic differential equations, or, equivalently, SDE in several dimensions. For the moment, we will not initiate a systematic study but rather consider some examples. The setup is the following: We are given a d-dimensional Brownian motion $B_t = (B_t^1, \ldots, B_t^d)$. The component processes $B_t^k, 1 \leq k \leq d$, are independent one-dimensional Brownian motions that drive the stochastic dynamics. We are looking for a stochastic process $X_t : \Omega \to \mathbb{R}^n$ solving an SDE of the form

$$dX_t \;=\; b(t, X_t)\, dt + \sum_{k=1}^{d} \sigma_k(t, X_t)\, dB_t^k. \tag{8.5.1}$$

Here n and d may be different, and $b, \sigma_1, \ldots, \sigma_d : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ are time-dependent continuous vector fields on $\mathbb{R}^n$. In matrix notation,

$$dX_t \;=\; b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{8.5.2}$$

where $\sigma(t, x) = (\sigma_1(t, x)\sigma_2(t, x) \cdots \sigma_d(t, x))$ is an $n \times d$-matrix.

Existence, uniqueness and stability

Assuming Lipschitz continuity of the coefficients, existence, uniqueness and stability of strong solutions of the SDE (8.5.2) can be shown by similar arguments as for ordinary differential equations.

Theorem 8.15 (Existence, uniqueness and stability under global Lipschitz conditions). *Suppose that b and σ satisfy a global Lipschitz condition of the following form: For any $t_0 \in \mathbb{R}$, there exists a constant $L \in \mathbb{R}_+$ such that*

$$|b(t,x)-b(t,\widetilde{x})|+||\sigma(t,x)-\sigma(t,\widetilde{x})|| \;\le\; L\cdot|x-\widetilde{x}| \quad \forall\, t \in [0,t_0],\ x,\widetilde{x} \in \mathbb{R}^n. \tag{8.5.3}$$

Then for any initial value $x \in \mathbb{R}^n$, the SDE (8.5.2) has a unique (up to equivalence) strong solution $(X_t)_{t\in[0,\infty)}$ such that $X_0 = x$ P-almost surely.

Furthermore, if (X_t) and $(\widetilde{X}_t)$ are two strong solutions with arbitrary initial conditions, then for any $t \in \mathbb{R}_+$, there exists a finite constant $C(t)$ such that

$$E\left[\sup_{s\in[0,t]} |X_s - \widetilde{X}_s|\right] \;\le\; C(t)\cdot E\left[|X_0 - \widetilde{X}_0|^2\right].$$

The proof of Theorem 8.15 is outlined in the exercises below. In Section 12.1, we will prove more general results that contain the assertion of the theorem as a special case. In particular, we will see that existence up to an explosion time and uniqueness of strong solutions still hold true if one assumes only a local Lipschitz condition.

The key step for proving stability and uniqueness is to control the deviation

$$\varepsilon_t \;:=\; E\left[\sup_{s\le t} |X_s - \widetilde{X}_s|^2\right]$$

between two solutions up to time t. Existence of strong solutions can then be shown by a Picard-Lindelöf approximation based on a corresponding norm:

Exercise (Proof of stability and uniqueness). Suppose that (X_t) and $(\widetilde{X}_t)$ are strong solutions of (8.5.2), and let $t_0 \in \mathbb{R}_+$. Apply Itô's isometry and Gronwall's inequality to show that if (8.5.3) holds, then there exists a finite constant $C \in \mathbb{R}_+$ such that for any $t \le t_0$,

$$\varepsilon_t \;\le\; C\cdot\left(\varepsilon_0 + \int_0^t \varepsilon_s\,ds\right), \quad \text{and} \tag{8.5.4}$$

$$\varepsilon_t \;\le\; C\cdot e^{Ct}\,\varepsilon_0. \tag{8.5.5}$$

Hence conclude that two strong solutions with the same initial value coincide almost surely.

Exercise (Existence of strong solutions). Define approximate solutions of (8.5.2) with initial value $x \in \mathbb{R}^n$ inductively by setting $X_t^0 := x$ for all t, and

$$X_t^{n+1} \;:=\; x \;+\; \int_0^t b(s, X_s^n)\, ds \;+\; \int_0^t \sigma(s, X_s^n)\, dB_s.$$

Let $\Delta_t^n := E[\sup_{s\leq t} |X_s^{n+1} - X_s^n|^2]$. Show that if (8.5.3) holds, then for any $t_0 \in \mathbb{R}_+$, there exists a finite constant $C(t_0)$ such that

$$\Delta_t^{n+1} \quad \leq \quad C(t_0) \int_0^t \Delta_s^n \, ds \qquad \text{for any } \; n \geq 0 \; \text{ and } \; t \leq t_0, \qquad \text{and}$$

$$\Delta_t^n \quad \leq \quad C(t_0)^n \, \frac{t^n}{n!} \, \Delta_t^0 \qquad \text{for any } \; n \in \mathbb{N} \; \text{ and } \; t \leq t_0.$$

Hence conclude that the limit $X_s = \lim_{n\to\infty} X_s^n$ exists uniformly for $s \in [0, t_0]$ with probability one, and X is a strong solution of (8.5.2) with $X_0 = x$.

Itô processes driven by several Brownian motions

Any solution to the SDE (8.5.1) is an Itô process pf type

$$X_t \;=\; \int_0^t G_s \, ds + \sum_{k=1}^d \int_0^t H_s^k \, dB_s^k \tag{8.5.6}$$

with continuous $(\mathcal{F}_t^{B,P})$ adapted stochastic processes $G_s, H_s^1, H_s^2, \ldots, H_s^d$. We now extend the stochastic calculus rules to such Itô processes that are driven by several independent Brownian motions. Let H_s and $\widetilde{H}_s$ be continuous $(\mathcal{F}_t^{B,P})$ adapted processes.

Lemma 8.16. *If (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$ then for any $1 \leq k, l \leq d$ and $a \in \mathbb{R}_+$, the covariation of the Itô integrals $t \mapsto \int_0^t H_s \, dB_s^k$ and $t \mapsto \int_0^t \widetilde{H}_s \, dB_s^l$ exists almost surely uniformly for $t \in [0, a]$ along a subsequence of (π_n), and*

$$\left[\int_0^\bullet H \, dB^k, \int_0^\bullet \widetilde{H} \, dB^l \right]_t \;=\; \int_0^t H \widetilde{H} \, d[B^k, B^l] \;=\; \delta_{kl} \int_0^t H_s \widetilde{H}_s \, ds.$$

The proof is an extension of the proof of Theorem 8.1(ii), where the assertion has been derived for $k = l$ and $H = \widetilde{H}$. The details are left as an exercise.

Similarly to the one-dimensional case, the lemma can be used to compute the covariation of Itô integrals w.r.t. arbitrary Itô processes. If X_s and Y_s are Itô processes as in (8.5.1), and K_s and L_s are adapted and continuous then we obtain

$$\left[\int_0^\bullet K\,dX, \int_0^\bullet L\,dY\right]_t = \int_0^t K_s L_s\, d[X,Y]_s$$

almost surely uniformly for $t \in [0,u]$, along an appropriate subsequence of (π_n).

Multivariate Itô-Doeblin formula

We now assume again that $(X_t)_{t\geq 0}$ is a solution of a stochastic differential equation of the form (8.5.1). By Lemma 8.16, we can apply Itô's formula to almost every sample path $t \mapsto X_t(\omega)$:

Theorem 8.17 (Itô-Doeblin). *Let $F \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$. Then almost surely,*

$$\begin{aligned} F(t, X_t) &= F(0, X_0) + \int_0^t (\sigma^\top \nabla_x F)(s, X_s) \cdot dB_s \\ &\quad + \int_0^t \left(\frac{\partial F}{\partial t} + \mathscr{L} F\right)(s, X_s)\, ds \qquad \text{for all } t \geq 0, \end{aligned}$$

where ∇_x denotes the gradient in the space variable, and

$$(\mathscr{L} F)(t,x) := \frac{1}{2} \sum_{i,j=1}^n a_{i,j}(t,x) \frac{\partial^2 F}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^n b_i(t,x) \frac{\partial F}{\partial x_i}(t,x)$$

with $a(t,x) := \sigma(t,x)\sigma(t,x)^\top \in \mathbb{R}^{n\times n}$.

Proof. If X is a solution to the SDE then

$$\begin{aligned}
[X^i, X^j]_t &= \sum_{k,l} \Big[\int \sigma^i_k(s,X)\, dB^k, \int \sigma^j_l(s,X)\, dB^l \Big]_t \\
&= \sum_{k,l} \int_0^t (\sigma^i_k \sigma^j_l)(s,X)\, d[B^k, B^l] \quad = \quad \int_0^t a^{ij}(s,X_s)\, ds
\end{aligned}$$

where $a^{ij} = \sum_k \sigma^i_k \sigma^j_k$, i.e.,

$$a(s,x) = \sigma(s,x)\sigma(s,x)^T \in \mathbb{R}^{n\times n}.$$

Therefore, Itô's formula applied to the process (t, X_t) yields

$$\begin{aligned}
dF(t,X) &= \frac{\partial F}{\partial t}(t,X)\, dt + \nabla_x F(t,X)\cdot dX + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 F}{\partial x^i \partial x^j}(t,X)\, d[X^i, X^j] \\
&= (\sigma^T \nabla_x F)(t,X)\cdot dB + \Big(\frac{\partial F}{\partial t} + \mathscr{L} F\Big)(t,X)\, dt,
\end{aligned}$$

for any $F \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$. □

The Itô-Doeblin formula shows that for any $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$, the process

$$M^F_s \;=\; F(s,X_s) - F(0,X_0) - \int_0^s \left(\frac{\partial F}{\partial t} + \mathscr{L} F\right)(t,X_t)\, dt$$

is a local martingale. If $\sigma^\top \nabla_x F$ is bounded then M^F is a global martingale.

Exercise (Drift and diffusion coefficients). Show that the processes

$$M^i_s \;=\; X^i_s - X^i_0 - \int_0^s b^i(s,X_s)\, ds, \qquad 1 \le i \le n,$$

are local martingales with covariations

$$[M^i, M^j]_s \;=\; a_{i,j}(s,X_s) \qquad \text{for any } s \ge 0,\ P\text{-almost surely.}$$

The vector field $b(s,x)$ is called the *drift vector field* of the SDE, and the coefficients $a_{i,j}(s,x)$ are called *diffusion coefficients*.

General Ornstein-Uhlenbeck processes

XXX to be included

Example (Stochastic oscillator).

Examples

Example (Physical Brownian motion with external force).

Example (Kalman-Bucy filter).

Example (Heston model for stochastic volatility).

Chapter 9

Change of measure

9.1 Local and global densities of probability measures

A thorough understanding of absolute continuity and relative densities of probability measures is crucial at many places in stochastic analysis. Martingale convergence yields an elegant approach to these issues including a proof of the Radon-Nikodym and the Lebesgue Decomposition Theorem. We first recall the definition of absolute continuity.

Absolute Continuity

Suppose that P and Q are probability measures on a measurable space $(\Omega, \mathcal{A})$, and $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{A}$.

Definition. *(1). The measure Q is called **absolutely continuous w.r.t.** P **on the σ-algebra** $\mathcal{F}$ if and only if $Q[A] = 0$ for any $A \in \mathcal{F}$ with $P[A] = 0$.*

*(2). The measures Q and P are called **singular on** $\mathcal{F}$ if and only if there exists $A \in \mathcal{F}$ such that $P[A] = 0$ and $Q[A^C] = 0$.*

We use the notations $Q \ll P$ for absolute continuity of Q w.r.t. P, $Q \approx P$ for mutual absolute continuity, and $Q \perp\!\!\!\perp P$ for singularity of Q and P. The definitions above extend to signed measures.

Example. The Dirac measure $\delta_{1/2}$ is obviously singular w.r.t. Lebesgue measure $\lambda_{(0,1]}$ on the Borel σ-algebra $\mathcal{B}((0,1])$. However, $\delta_{1/2}$ is absolutely continuous w.r.t. $\lambda_{(0,1]}$ on each of the σ-algebras $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the dyadic partitions $\mathcal{D}_n = \{(k \cdot 2^{-n}, (k+1)2^{-n}] \,:\, 0 \leq k < 2^n\}$, and $\mathcal{B}([0,1)) = \sigma(\bigcup \mathcal{D}_n)$.

The next lemma clarifies the term "absolute continuity."

Lemma 9.1. *The probability measure Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}$ if and only if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for $A \in \mathcal{F}$,*

$$P[A] < \delta \quad \Rightarrow \quad Q[A] < \varepsilon. \tag{9.1.1}$$

Proof. The "if" part is obvious. If $P[A] = 0$ and (9.1.1) holds for each $\varepsilon > 0$ with δ depending on ε then $Q[A] < \varepsilon$ for any $\varepsilon > 0$, and hence $Q[A] = 0$.
To prove the "only if" part, we suppose that there exists $\varepsilon > 0$ such that (9.1.1) does not hold for any $\delta > 0$. Then there exists a sequence (A_n) of events in $\mathcal{F}$ such that

$$Q[A_n] \geq \varepsilon \qquad \text{and} \qquad P[A_n] \leq 2^{-n}.$$

Hence, by the Borel-Cantelli-Lemma,

$$P[A_n \quad \text{infinitely often}] = 0,$$

whereas

$$Q[A_n \quad \text{infinitely often}] = Q\left[\bigcap_n \bigcup_{m \geq n} A_m\right] = \lim_{n\to\infty} Q\left[\bigcup_{m\geq n} A_m\right] \geq \varepsilon.$$

Therefore Q is not absolutely continuous w.r.t. P. □

Example (Absolute continuity on $\mathbb{R}$). A probability measure μ on a real interval is absolutely continuous w.r.t. Lebesgue measure if and only if the distribution function $F(t) = \mu[(-\infty, t]]$ satisfies:

For any $\varepsilon > 0$ there exists $\delta > 0$ such that for any $n \in \mathbb{N}$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$,

$$\sum_{i=1}^{n} |b_i - a_i| \;<\; \varepsilon \quad \Rightarrow \quad \sum_{i=1}^{n} |F(b_i) - F(a_i)| \;<\; \delta, \tag{9.1.2}$$

cf. e.g. [Billingsley: Probability and Measures].

Definition (Absolutely continuous functions). *A function $F : (a,b) \subset \mathbb{R} \to \mathbb{R}$ is called **absolutely continuous** iff (9.1.2) holds.*

The Radon-Nikodym Theorem states that absolute continuity is equivalent to the existence of a relative density.

Theorem 9.2 (Radon-Nikodym). *The probability measure Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}$ if and only if there exists a non-negative random variable $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ such that*

$$Q[A] \;=\; \int_A Z \, dP \qquad \textit{for any } A \in \mathcal{F}. \tag{9.1.3}$$

The relative density Z of Q w.r.t. P on $\mathcal{F}$ is determined by (9.1.3) uniquely up to modification on P-measure zero sets. It is also called the **Radon-Nikodym derivative** or the **likelihood ratio** of Q w.r.t. P on $\mathcal{F}$. We use the notation

$$Z \;=\; \left.\frac{dQ}{dP}\right|_{\mathcal{F}},$$

and omit the $\mathcal{F}$ when the choice of the σ-algebra is clear.

Example (Finitely generated σ-algebra). Suppose that the σ-algebra $\mathcal{F}$ is generated by finitely many disjoint atoms $B_1, \ldots, B_k$ with $\Omega = \bigcup B_i$. Then Q is absolutely continuous w.r.t. P if and only if for any i,

$$P[B_i] \;=\; 0 \quad \Longrightarrow \quad Q[B_i] = 0.$$

In this case, the relative density is given by

$$\frac{dQ}{dP}\bigg|_{\mathcal{F}} = \sum_{i\,:\,P[B_i]>0} \frac{Q[B_i]}{P[B_i]} \cdot I_{B_i}.$$

From local to global densities

Let $(\mathcal{F}_n)$ be a given filtration on $(\Omega, \mathcal{A})$.

Definition (Local absolutely continuity). *The measure Q is called **locally absolutely continuous** w.r.t. P and the filtration $(\mathcal{F}_n)$ if and only if Q is absolutely continuous w.r.t. P on the σ-algebra $\mathcal{F}_n$ for each n.*

Example (Dyadic partitions). Any probability measure on the unit interval $[0,1]$ is locally absolutely continuous w.r.t. Lebesgue measure on the filtration $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the dyadic partitions of the unit interval. The Radon-Nikodym derivative on $\mathcal{F}_n$ is the dyadic difference quotient defined by

$$\frac{d\mu}{d\lambda}\bigg|_{\mathcal{F}_n}(x) = \frac{\mu[((k-1)\cdot 2^{-n}, k\cdot 2^{-n})]}{\lambda[((k-1)\cdot 2^{-n}, k\cdot 2^{-n})]} = \frac{F(k\cdot 2^{-n}) - F((k-1)\cdot 2^{-n})}{2^{-n}} \tag{9.1.4}$$

for $x \in ((k-1)2^{-n}, k2^{-n}]$.

Example (Product measures). If $Q = \bigotimes_{i=1}^{\infty} \nu$ and $P = \bigotimes_{i=1}^{\infty} \mu$ are infinite products of probability measures ν and μ, and ν is absolutely continuous w.r.t. μ with density ϱ, then Q is locally absolutely continuous w.r.t. P on the filtration

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$$

generated by the coordinate maps $X_i(\omega) = \omega_i$. The local relative density is

$$\frac{dQ}{dP}\bigg|_{\mathcal{F}_n} = \prod_{i=1}^{n} \varrho(X_i)$$

However, if $\nu \neq \mu$, then Q is not absolutely continuous w.r.t. P on $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$, since by the LLN, $n^{-1}\sum_{i=1}^{n} I_A(X_i)$ converges Q almost surely to $\nu[A]$ and P-almost surely to $\mu[A]$.

Now suppose that Q is locally absolutely continuous w.r.t. P on a filtration $(\mathcal{F}_n)$ with relative densities

$$Z_n = \frac{dQ}{dP}\bigg|_{\mathcal{F}_n}.$$

The L^1 martingale convergence theorem can be applied to study the existence of a global density on the σ-algebra

$$\mathcal{F}_\infty = \sigma(\bigcup \mathcal{F}_n).$$

Let $Z_\infty := \limsup Z_n$.

Theorem 9.3 (Convergence of local densities, Lebesgue decomposition).

(1). The sequence (Z_n) of successive relative densities is an $(\mathcal{F}_n)$-martingale w.r.t. P. In particular, (Z_n) converges P-almost surely to Z_∞, and Z_∞ is integrable w.r.t. P.

(2). The following statements are equivalent:

(a) (Z_n) is uniformly integrable w.r.t. P.

(b) Q is absolutely continuous w.r.t. P on $\mathcal{F}_\infty$.

(c) $Q[A] = \int\limits_A Z_\infty \, dP$ for any P on $\mathcal{F}_\infty$.

(3). In general, the decomposition $Q = Q_a + Q_s$ holds with

$$Q_a[A] = \int\limits_A Z_\infty \, dP, \qquad Q_s[A] = Q[A \cap \{Z_\infty = \infty\}]. \tag{9.1.5}$$

Q_a and Q_s are positive measure with $Q_a \ll P$ and $Q_s \perp\!\!\!\perp P$.

The decomposition $Q = Q_a + Q_s$ into an absolutely continuous and a singular part is called the **Lebesgue decomposition** of the measure Q w.r.t. P on the σ-algebra $\mathcal{F}_\infty$.

Proof. (1). For $n \geq 0$, the density Z_n is in $\mathcal{L}^1(\Omega, \mathcal{F}_n, P)$, and

$$E_P[Z_n \,;\, A] = Q[A] = E_P[Z_{n+1} \,;\, A] \qquad \text{for any } A \in \mathcal{F}_n.$$

Hence $Z_n = E_P[Z_{n+1} \mid \mathcal{F}_n]$, i.e., (Z_n) is a martingale w.r.t. P. Since $Z_n \geq 0$, the martingale converges P-almost surely, and the limit is integrable.

(2). (a) $\Rightarrow$ (c): If (Z_n) is uniformly integrable w.r.t. P, then

$$Z_n \;=\; E_P[Z_\infty \mid \mathcal{F}_n] \qquad P\text{-almost surely for any } n,$$

by the L^1 convergence theorem. Hence for $A \in \mathcal{F}_n$,

$$Q[A] \;=\; E_P[Z_n\,;\,A] \;=\; E_P[Z_\infty\,;\,A].$$

This shows that $Q[A] = E_P[Z_\infty\,;\,A]$ holds for any $A \in \bigcup \mathcal{F}_n$, and thus for any $A \in \mathcal{F}_\infty = \sigma(\bigcup \mathcal{F}_n)$.

(c) $\Rightarrow$ (b) is evident.

(b) $\Rightarrow$ (a): If $Q \ll P$ on $\mathcal{F}_\infty$ then Z_n converges also Q-almost surely to a finite limit Z_∞. Hence for $n_0 \in \mathbb{N}$ and $c > 1$,

$$\begin{aligned}\sup_n E_P[\,|Z_n|\,;\,|Z_n| \geq c] \;&=\; \sup_n E_P[Z_n\,;\,Z_n \geq c] \;=\; \sup_n Q[Z_n \geq c]\\ &\leq\; \max_{n<n_0} Q[Z_n \geq c] + \sup_{n \geq n_0} Q[Z_n \geq c]\\ &\leq\; \max_{n<n_0} Q[Z_n \geq c] + Q[Z_\infty \geq c-1] + \sup_{n \geq n_0} Q[|Z_n - Z_\infty| \geq 1].\end{aligned}$$

Given $\varepsilon > 0$, the last summand is smaller than $\varepsilon/3$ for n_0 sufficiently large, and the other two summands on the right hand side are smaller than $\varepsilon/3$ if c is chosen sufficiently large depending on n_0. Hence (Z_n) is uniformly integrable w.r.t. P.

(3). In general, $Q_a[A] \;=\; E_P[Z_\infty\,;\,A]$ is a positive measure on $\mathcal{F}_\infty$ with $Q_a \leq Q$, since for $n \geq 0$ and $A \in \mathcal{F}_n$,

$$Q_a[A] \;=\; E_P[\liminf_{k\to\infty} Z_k\,;\,A] \;\leq\; \liminf_{k\to\infty} E_P[Z_k\,;\,A] \;=\; E_P[Z_n\,;\,A] \;=\; Q[A]$$

by Fatou's Lemma and the martingale property. It remains to show that

$$Q_a[A] \;=\; Q[A \cap \{Z_\infty < \infty\}] \qquad \text{for any } A \in \mathcal{F}_\infty. \tag{9.1.6}$$

If (9.1.6) holds, then $Q = Q_a + Q_s$ with Q_s defined by (9.1.5). In particular, Q_s is then singular w.r.t. P, since $P[Z_\infty = \infty] = 0$ and $Q_s[Z_\infty = \infty] = 0$, whereas

Q_a is absolutely continuous w.r.t. P by definition.
Since $Q_a \leq Q$, it suffices to verify (9.1.6) for $A = \Omega$. Then

$$(Q - Q_a)[A \cap \{Z_\infty < \infty\}] \;=\; (Q - Q_a)[Z_\infty < \infty] \;=\; 0,$$

and therefore

$$Q[A \cap \{Z_\infty < \infty\}] \;=\; Q_a[A \cap \{Z_\infty < \infty\}] \;=\; Q_a[A]$$

for any $A \in \mathcal{F}_\infty$.
To prove (9.1.6) for $A = \Omega$ we observe that for $c \in (0, \infty)$,

$$\begin{aligned} Q\left[\limsup_{n\to\infty} Z_n < c\right] \;&\leq\; \limsup_{n\to\infty} Q[Z_n < c] \;=\; \limsup_{n\to\infty} E_P[Z_n \,;\, Z_n < c] \\ &\leq\; E_P\left[\limsup_{n\to\infty} Z_n \cdot I_{\{Z_n < c\}}\right] \;\leq\; E_P[Z_\infty] \;=\; Q_a[\Omega] \end{aligned}$$

by Fatou's Lemma. As $c \to \infty$, we obtain

$$Q[Z_\infty < \infty] \;\leq\; Q_a[\Omega] \;=\; Q_a[Z_\infty < \infty] \;\leq\; Q[Z_\infty < \infty]$$

and hence (9.1.6) with $A = \Omega$. This completes the proof

□

As a first consequence of Theorem 9.3, we prove the Radon-Nikodym Theorem on a separable σ-algebra $\mathcal{A}$. Let P and Q be probability measures on $(\Omega, \mathcal{A})$ with $Q \ll P$.

Proof of the Radon-Nikodym Theorem for separable σ-algebras. We fix a filtration $(\mathcal{F}_n)$ consisting of finitely generated σ-algebras $\mathcal{F}_n \subseteq \mathcal{A}$ with $\mathcal{A} = \sigma(\bigcup \mathcal{F}_n)$. Since Q is absolutely continuous w.r.t. P, the local densities $Z_n = dQ/dP|_{\mathcal{F}_n}$ on the finitely generated σ-algebras $\mathcal{F}_n$ exist, cf. the example above. Hence by Theorem 9.3,

$$Q[A] \;=\; \int_A Z_\infty \, dP \qquad \text{for any } A \in \mathcal{A}.$$

□

The approach above can be generalized to probability measures that are not absolutely continuous:

Exercise (Lebesgue decomposition, Lebesgue densities). Let P and Q be arbitrary (not necessarily absolutely continuous) probability measures on $(\Omega, \mathcal{A})$. A **Lebesgue density** of Q w.r.t. P is a random variable $Z : \Omega \to [0, \infty]$ such that $Q = Q_a + Q_s$ with

$$Q_a[A] \;=\; \int_A Z\,dP, \quad Q_s[A] \;=\; Q[A \cap \{Z = \infty\}] \qquad \text{for any } A \in \mathcal{A}.$$

The goal of the exercise is to prove that a Lebesgue density exists if the σ-algebra $\mathcal{A}$ is separable.

(1). Show that if Z is a Lebesgue density of Q w.r.t. P then $1/Z$ is a Lebesgue density of P w.r.t. Q. Here $1/\infty := 0$ and $1/0 := \infty$.

From now on suppose that the σ-algebra is separable with $\mathcal{A} = \sigma(\bigcup \mathcal{F}_n)$ where $(\mathcal{F}_n)$ is a filtration consisting of σ-algebras generated by finitely many atoms.

(1). Write down Lebesgue densities Z_n of Q w.r.t. P on each $\mathcal{F}_n$. Show that

$$Q[Z_n = \infty \text{ and } Z_{n+1} < \infty] \;=\; 0 \qquad \text{for any } n,$$

and conclude that (Z_n) is a non-negative supermartingale under P, and $(1/Z_n)$ is a non-negative supermartingale under Q.

(2). Prove that the limit $Z_\infty = \lim Z_n$ exists both P-almost surely and Q-almost surely, and $P[Z_\infty < \infty] = 1$ and $Q[Z_\infty > 0] = 1$.

(3). Conclude that Z_∞ is a Lebesgue density of P w.r.t. Q on $\mathcal{A}$, and $1/Z_\infty$ is a Lebesgue density of Q w.r.t. P on $\mathcal{A}$.

Derivatives of monotone functions

Suppose that $F : [0, 1] \to \mathbb{R}$ is a monotone and right-continuous function. After an appropriate linear transformation we may assume that F is non decreasing with $F(0) = 0$ and $F(1) = 1$. Let μ denote the probability measure with distribution function F. By the example above, the Radon-Nikodym derivative of μ w.r.t. Lebesgue measure on the σ-algebra $\mathcal{F}_n = \sigma(\mathcal{D}_n)$ generated by the n-th dyadic partition of the unit interval is given by the dyadic difference quotients (9.1.4) of F. By Theorem 9.3, we obtain a version of Lebesgue's Theorem on derivatives of monotone functions:

Corollary 9.4 (Lebesgue's Theorem). *Suppose that $F : [0,1] \to \mathbb{R}$ is monotone (and right continuous). Then the dyadic derivative*

$$F'(t) \;=\; \lim_{n\to\infty} \frac{d\mu}{d\lambda}\bigg|_{\mathcal{F}_n}(t)$$

exists for almost every t and F' is an integrable function on $(0,1)$. Furthermore, if F is absolutely continuous then

$$F(s) \;=\; \int_0^s F'(t)\,dt \qquad \text{for all } s \in [0,1]. \tag{9.1.7}$$

Remark. Right continuity is only a normalization and can be dropped from the assumptions. Moreover, the assertion extends to function of finite variation since these can be represented as the difference of two monotone functions, cf. **??** below. Similarly, (9.1.7) also holds for absolutely continuous functions that are not monotone. See e.g. [Elstrodt: Maß- und Integrationstheorie] for details.

Absolute continuity of infinite product measures

Suppose that $\Omega = \bigtimes_{i=1}^{\infty} S_i$, and

$$Q \;=\; \bigotimes_{i=1}^{\infty} \nu_i \qquad \text{and} \qquad P \;=\; \bigotimes_{i=1}^{\infty} \mu_i$$

are products of probability measures ν_i and μ_i defined on measurable spaces $(S_i, \mathcal{S}_i)$. We assume that ν_i and μ_i are mutually absolutely continuous for every $i \in \mathbb{N}$. Denoting by $X_k : \Omega \to S_k$ the evaluation of the k-th coordinate, the product measures are mutually absolutely continuous on each of the σ-algebras

$$\mathcal{F}_n \;=\; \sigma(X_1, \ldots, X_n), \qquad n \in \mathbb{N},$$

with relative densities

$$\frac{dQ}{dP}\bigg|_{\mathcal{F}_n} \;=\; Z_n \quad \text{and} \quad \frac{dP}{dQ}\bigg|_{\mathcal{F}_n} \;=\; 1/Z_n,$$

where

$$Z_n \;=\; \prod_{i=1}^{n} \frac{d\nu_i}{d\mu_i}(X_i) \quad \in \; (0,\infty) \qquad P\text{-almost surely.}$$

In particular, (Z_n) is a martingale under P, and $(1/Z_n)$ is a martingale under Q. Let $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$ denote the product σ-algebra.

Theorem 9.5 (Kakutani's dichotomy). *The infinite product measures Q and P are either singular or mutually absolutely continuous with relative density Z_∞. More precisely, the following statements are equivalent:*

(1). $Q \ll P$ on $\mathcal{F}_\infty$.

(2). $Q \approx P$ on $\mathcal{F}_\infty$.

(3). $\prod\limits_{i=1}^{\infty} \int \sqrt{\frac{d\nu_i}{d\mu_i}} d\mu_i \;>\; 0.$

(4). $\sum\limits_{i=1}^{\infty} d_H^2(\nu_i,\mu_i) \;<\; \infty.$

Here the squared Hellinger distance $d_H^2(\nu_i,\mu_i)$ of mutually absolutely continuous probability measures ν and μ is defined by

$$\begin{aligned} d_H^2 \;&=\; \frac{1}{2}\int \left(\sqrt{\frac{d\nu}{d\mu}} - 1\right)^2 d\mu \;=\; \frac{1}{2}\int \left(\sqrt{\frac{d\mu}{d\nu}} - 1\right)^2 d\nu \\ &=\; 1 - \int \sqrt{\frac{d\nu}{d\mu}}\, d\mu \;=\; 1 - \int \sqrt{\frac{d\mu}{d\nu}}\, d\nu. \end{aligned}$$

Remark. (1). If mutual absolutely continuity holds then the relative densities on $\mathcal{F}_\infty$ are

$$\frac{dQ}{dP} \;=\; \lim_{n\to\infty} Z_n \quad P\text{-almost surely, and} \quad \frac{dP}{dQ} \;=\; \lim_{n\to\infty} \frac{1}{Z_n} \quad Q\text{-almost surely.}$$

(2). If ν and μ are absolutely continuous w.r.t. a measure dx then

$$d_H^2(\nu,\mu) \;=\; \frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx \;=\; 1 - \int \sqrt{f(x)g(x)}\, dx.$$

Proof. (1) $\iff$ (3): For $i \in \mathbb{N}$ let $Y_i := \frac{d\nu_i}{d\mu_i}(X_i)$. Then the random variables Y_i are independent under both P and Q with $E_P[Y_i] = 1$, and

$$Z_n \;=\; Y_1 \cdot Y_2 \cdots Y_n.$$

By Theorem 9.3, the measure Q is absolutely continuous w.r.t. P if and only if the martingale (Z_n) is uniformly integrable. To obtain a sharp criterion for uniform integrability we switch from L^1 to L^2, and consider the non-negative martingale

$$M_n \;=\; \frac{\sqrt{Y_1}}{\beta_1} \cdot \frac{\sqrt{Y_2}}{\beta_2} \cdots \frac{\sqrt{Y_n}}{\beta_n} \qquad \text{with } \beta_i \;=\; E_P[\sqrt{Y_i}] \;=\; \int \sqrt{\frac{d\nu_i}{d\mu_i}}\, d\mu_i$$

under the probability measure P. Note that for $n \in \mathbb{N}$,

$$E[M_n^2] \;=\; \prod_{i=1}^n E[Y_i]/\beta_i^2 \;=\; 1 \Big/ \left(\prod_{i=1}^n \beta_i\right)^2 .$$

If (3) holds then (M_n) is bounded in $L^2(\Omega, \mathcal{A}, P)$. Therefore, by Doob's L^2 inequality, the supremum of M_n is in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, i.e.,

$$E[\sup |Z_n|\,] \;=\; E[\sup M_n^2] \;<\; \infty.$$

Thus (Z_n) is uniformly integrable and $Q \ll P$ on $\mathcal{F}_\infty$.

Conversely, if (3) does not hold then

$$Z_n \;=\; M_N^2 \cdot \prod_{i=1}^n \beta_i \;\longrightarrow\; 0 \qquad P\text{-almost surely,}$$

since M_n converges to a finite limit by the martingale convergence theorem. Therefore, the absolute continuous part Q_a vanishes by Theorem 9.3 (3), i.e., Q is singular w.r.t. P.

(3) $\iff$ (4): For reals $\beta_i \in (0,1)$, the condition $\prod\limits_{i=1}^{\infty} \beta_i > 0$ is equivalent to $\sum\limits_{i=1}^{\infty}(1-\beta_i) < \infty$. For β_i as above, we have

$$1 - \beta_i \;=\; 1 - \int \sqrt{\frac{d\nu_i}{d\mu_i}}\, d\mu_i \;=\; d_H^2(\nu_i, \mu_i).$$

(2) $\Rightarrow$ (1) is obvious.

(4) $\Rightarrow$ (2): Condition (4) is symmetric in ν_i and μ_i. Hence, if (4) holds then both $Q \ll P$ and $P \ll Q$. □

Example (Gaussian products). Let $P = \bigotimes_{i=1}^{\infty} N(0,1)$ and $Q = \bigotimes_{i=1}^{\infty} N(a_i,1)$ where $(a_i)_{i\in\mathbb{N}}$ is a sequence of reals. The relative density of the normal distributions $\nu_i := N(a_i,1)$ and $\mu := N(0,1)$ is

$$\frac{d\nu_i}{d\mu}(x) \;=\; \frac{\exp(-(x-a_i)^2)/2}{\exp(-x^2/2)} \;=\; \exp(a_i x - a_i^2/2),$$

and

$$\int \sqrt{\frac{d\nu_i}{d\mu}}\, d\mu \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x^2 - a_i x + a_i^2/2)\right) dx \;=\; \exp(-a_i^2/8).$$

Therefore, by condition (3) in Theorem 9.5,

$$Q \ll P \iff Q \approx P \iff \sum_{i=1}^{\infty} a_i^2 < \infty.$$

Hence mutual absolute continuity holds for the infinite products if and only if the sequence (a_i) is contained in ℓ^2, and otherwise Q and P are singular.

Remark (Relative entropy). (1). In the singular case, the exponential rate of degeneration of the relative densities on the σ-algebras $\mathcal{F}_n$ is related to the relative entropies

$$H(\nu_i \mid \mu_i) \;=\; \int \frac{d\nu_i}{d\mu_i} \log \frac{d\nu_i}{d\mu_i}\, d\mu_i \;=\; \int \log \frac{d\nu_i}{d\mu_i}\, d\nu_i.$$

For example in the i.i.d. case $\mu_i \equiv \mu$ and $\nu_i \equiv \nu$, we have

$$\frac{1}{n}\log Z_n \;=\; \frac{1}{n}\sum_{i=1}^{n} \log \frac{d\nu}{d\mu}(X_i) \;\longrightarrow\; H(\nu \mid \mu) \qquad Q\text{-a.s., and}$$

$$-\frac{1}{n}\log Z_n \;=\; \frac{1}{n}\log Z^{-1} \;\longrightarrow\; H(\mu \mid \nu) \qquad P\text{-a.s.}$$

as $n \to \infty$ by the Law of Large Numbers.

In general, $\log Z_n - \sum_{i=1}^{n} H(\nu_i|\mu_i)$ is a martingale w.r.t. Q, and $\log Z_n + \sum_{i=1}^{n} H(\nu_i|\mu_i)$ is a martingale w.r.t. P.

(2). The relative entropy is related to the squared Hellinger distance by the inequality

$$\frac{1}{2} H(\nu \mid \mu) \;\geq\; d_H^2(\nu \mid \mu),$$

which follows from the elementary inequality

$$\frac{1}{2}\log x^{-1} \;=\; -\log\sqrt{x} \;\geq\; 1 - \sqrt{x} \qquad \text{for } x > 0.$$

9.2 Translations of Wiener measure

We now return to stochastic processes in continuous time. We endow the continuous path space $C([0,\infty),\mathbb{R}^d)$ with the σ-algebra generated by the evolution maps $X_t(\omega) = \omega(t)$, and with the filtration

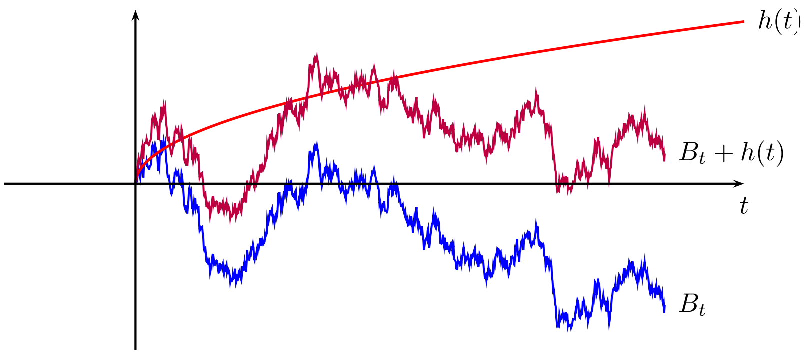
$$\mathcal{F}_t^X \;=\; \sigma(X_s \mid s \in [0,t]), \qquad t \geq 0.$$

Note that $\mathcal{F}_t^X$ consists of all sets of type

$$\big\{\omega \in C([0,\infty),\mathbb{R}^d) \;:\; \omega|_{[0,t]} \in \Gamma\big\} \qquad \text{with } \Gamma \in \mathcal{B}(C([0,t],\mathbb{R}^d)).$$

In many situations one is interested in the distribution on path space of a process

$$B_t^h \;=\; B_t + h(t)$$



obtained by translating a Brownian motion (B_t) by a deterministic function $h : [0,\infty) \to \mathbb{R}^d$. In particular, it is important to know if the distribution of (B_t^h) has a density w.r.t. the Wiener measure on the σ-algebras $\mathcal{F}_t^X$, and how to compute the densities if they exist.

Example. (1). Suppose we would like to evaluate the probability that $\sup\limits_{s\in[0,t]} |B_s - g(s)| < \varepsilon$ for a given $t > 0$ and a given function $g \in C([0,\infty),\mathbb{R}^d)$ asymptotically as $\varepsilon \searrow 0$. One approach is to study the distribution of the translated process $B_t - g(t)$ near 0.

(2). Similarly, computing the passage probability $P[B_s \geq a+bs \quad \text{for some } s \in [0,t]]$ to a line $s \mapsto a + bs$ for a one-dimensional Brownian motion is equivalent to computing the passage probability to the point a for the translated process $B_t - bt$.

(3). A solution to a stochastic differential equation

$$dY_t \;=\; dB_t + b(t, Y_t)dt$$

is a translation of the Brownian motion $B_t - B_0$ by the stochastic process $H_t = Y_0 + \int_0^t b(s, Y_s)\, ds$. Again, in the simplest case (when $b(t, y)$ only depends on t), H_t is a deterministic function.

The Cameron-Martin Theorem

Let (B_t) denote a continuous Brownian motion with $B_0 = 0$, and let $h \in C([0,\infty), \mathbb{R}^d)$. The distribution

$$\mu_h \;:=\; P \circ (B + h)^{-1}$$

of the translated process $B_t^h = B_t + h(t)$ is the image of Wiener measure μ_0 under the translation map

$$\tau_h \;:\; C([0,\infty), \mathbb{R}^d) \;\longrightarrow\; C([0,\infty), \mathbb{R}^d), \quad \tau_h(x) \;=\; x + h.$$

Recall that Wiener measure is a Gaussian measure on the infinite dimensional space $C([0,\infty), \mathbb{R}^d)$. The next exercise discusses translations of Gaussian measures in $\mathbb{R}^n$:

Exercise (Translations of normal distributions). Let $C \in \mathbb{R}^{n\times n}$ be a symmetric non-negative definite matrix, and let $h \in \mathbb{R}^n$. the image of the normal distribution $N(0, C)$ under the translation map $x \mapsto x + h$ on $\mathbb{R}^n$ is the normal distribution $N(h, C)$.

(1). Show that if C is non-degenerate then $N(h, C) \approx N(0, C)$ with relative density

$$\frac{dN(h,C)}{dN(0,C)}(x) \;=\; e^{(h,x)-\frac{1}{2}(h,h)} \qquad \text{for } x \in \mathbb{R}^n, \tag{9.2.1}$$

where $(g, h) = (g, C^{-1}, h)$ for $g, h \in \mathbb{R}^n$.

(2). Prove that in general, $N(h, C)$ is absolutely continuous w.r.t. $N(0, C)$ if and only if h is orthogonal to the kernel of C w.r.t. the Euclidean inner product.

On $C([0,\infty), \mathbb{R}^d)$, we can usually not expect the existence of a global density of the translated measures μ_h w.r.t. μ_0. The Cameron-Martin Theorem states that for $t \geq 0$, a relative density on $\mathcal{F}_t^X$ exists if and only if h is contained in the corresponding Cameron-Martin space:

Theorem 9.6 (**Cameron, Martin**). *For $h \in C([0,\infty), \mathbb{R}^d)$ and $t \in \mathbb{R}_+$ the translated measure $\mu_h = \mu \circ \tau_h^{-1}$ is absolutely continuous w.r.t. Wiener measure μ_0 on $\mathcal{F}_t^X$ if and only if h is an absolutely continuous function on $[0, t]$ with $h(0) = 0$ and $\int_0^t |h'(s)|^2 ds < \infty$. In this case, the relative density is given by*

$$\left.\frac{d\mu_h}{d\mu_0}\right|_{\mathcal{F}_t^X} = \exp\left(\int_0^t h'(s)\, dX_s - \frac{1}{2}\int_0^t |h'(s)|^2\, ds\right). \tag{9.2.2}$$

where $\int_0^t h'(s)\, dX_s$ is the Itô integral w.r.t. the canonical Brownian motion (X, μ_0).

Before giving a rigorous proof let us explain heuristically why the result should be true. Clearly, absolute continuity does not hold if $h(0) \neq 0$, since then the translated paths do not start at 0. Now suppose $h(0) = 0$, and fix $t \in (0,\infty)$. Absolute continuity on $\mathcal{F}_t^X$ means that the distribution μ_h^t of $(B_s^h)_{0 \leq s \leq t}$ on $C([0,t], \mathbb{R}^d)$ is absolutely continuous w.r.t. Wiener measure μ_0^t on this space. The measure μ_0^t, however, is a kind of infinite dimensional standard normal distribution w.r.t. the inner product

$$(x, y)_H = \int_0^t x'(s) \cdot y'(s)\, ds$$

on functions $x, y : [0, t] \to \mathbb{R}^d$ vanishing at 0, and the translated measure μ_h^t is a Gaussian measure with mean h and the same covariances.

Choosing an orthonormal basis $(e_i)_{i \in \mathbb{N}}$ w.r.t. the H-inner product (e.g. Schauder functions), we can identify μ_0^t and μ_h^t with the product measures $\bigotimes_{i=1}^{\infty} N(0, 1)$ and $\bigotimes_{i=1}^{\infty} N(\alpha_i, 1)$

respectively where $\alpha_i = (h, e_i)_H$ is the i-th coefficient of h in the basis expansion. Therefore, μ_h^t should be absolutely continuous w.r.t. μ_0^t if and only if

$$(h, h)_H = \sum_{i=1}^{\infty} \alpha_i^2 < \infty,$$

i.e., if and only if h is absolutely continuous with $h' \in \mathcal{L}^2(0, t)$.

Moreover, in analogy to the finite-dimensional case (9.2.1), we would expect informally a relative density of the form

$$\text{“}\ \frac{d\mu_h^t}{d\mu_0^t}(x) = e^{(h,x)_H - \frac{1}{2}(h,h)_H} = \exp\left(\int_0^t h'(s) \cdot x'(s)\, ds - \frac{1}{2}\int_0^t |h'(s)|^2\, ds\right)\ \text{”}$$

Since μ_0^t-almost every path $x \in C([0, \infty), \mathbb{R}^d)$ is not absolutely continuous, this expression does not make sense. Nevertheless, using finite dimensional approximations, we can derive the rigorous expression (9.2.2) for the relative density where the integral $\int_0^t h'x'\, ds$ is replaced by the almost surely well-defined stochastic integral $\int_0^t h'\, dx$:

Proof of Theorem 9.6. We assume $t = 1$. The proof for other values of t is similar. Moreover, as explained above, it is enough to consider the case $h(0) = 0$.

(1). *Local densities:* We first compute the relative densities when the paths are only evaluated at dyadic time points. Fix $n \in \mathbb{N}$, let $t_i = i \cdot 2^{-n}$, and let

$$\delta_i x = x_{t_{i+1}} - x_{t_i}$$

denote the i-th dyadic increment. Then the increments $\delta_i B^h$ $(i = 0, 1, \ldots, 2^n - 1)$ of the translated Brownian motion are independent random variables with distributions

$$\delta_i B^h = \delta_i B + \delta_i h \sim N(\delta_i h, (\delta t) \cdot I_d), \quad \delta t = 2^{-n}.$$

Consequently, the marginal distribution of $(B_{t_1}^h, B_{t_2}^h, \ldots, B_{t_{2^n}}^h)$ is a normal distribution with density w.r.t. Lebesgue measure proportional to

$$\exp\left(-\sum_{i=0}^{2^n-1} \frac{|\delta_i x - \delta_i h|^2}{2\delta t}\right), \qquad x = (x_{t_1}, x_{t_2}, \ldots, x_{t_{2^n}}) \in \mathbb{R}^{2^n d}.$$

Since the normalization constant does not depend on h, the joint distribution of $(B^h_{t_1}, B^h_{t_2}, \dots, B^h_{t_{2^n}})$ is absolutely continuous w.r.t. that of $(B_{t_1}, B_{t_2}, \dots, B_{t_{2^n}})$ with relative density

$$\exp\left(\sum \frac{\delta_i h}{\delta t} \cdot \delta_i x - \frac{1}{2} \sum \left|\frac{\delta_i h}{\delta t}\right|^2 \delta t\right). \tag{9.2.3}$$

Consequently, μ_h is always absolutely continuous w.r.t. μ_0 on each of the σ-algebras

$$\mathcal{F}_n \;=\; \sigma(X_{i\cdot 2^{-n}} : i = 0, 1, \dots, 2^n - 1), \qquad n \in \mathbb{N},$$

with relative densities

$$Z_n \;=\; \exp\left(\sum_{i=0}^{2^n-1} \frac{\delta_i h}{\delta t} \cdot \delta_i X - \frac{1}{2} \sum_{i=0}^{2^n-1} \left|\frac{\delta_i h}{\delta t}\right|^2 \delta t\right). \tag{9.2.4}$$

(2). *Limit of local densities:* Suppose that h is absolutely continuous with

$$\int_0^1 |h'(t)|^2 \, dt \;<\; \infty.$$

We now identify the limit of the relative densities Z_n as $n \to \infty$.
First, we note that

$$\sum_{i=0}^{2^n-1} \left|\frac{\delta_i h}{\delta t}\right|^2 \delta t \quad\longrightarrow\quad \int_0^1 |h'(t)|^2 \, dt \qquad \text{as } n \to \infty.$$

In fact, the sum on the right hand side coincides with the squared L^2 norm

$$\int_0^1 \left| dh/dt|_{\sigma(\mathcal{D}_n)} \right|^2 \, dt$$

of the dyadic derivative

$$\left.\frac{dh}{dt}\right|_{\sigma(\mathcal{D}_n)} \;=\; \sum_{i=0}^{2^n-1} \frac{\delta_i h}{\delta t} \cdot I_{((i-1)\cdot 2^{-n}, i\cdot 2^{-n}]}$$

on the σ-algebra generated by the intervals $((i-1)\cdot 2^{-n}, i \cdot 2^{-n}]$. If h is absolutely continuous with $h' \in L^2(0,1)$ then $\left.\frac{dh}{dt}\right|_{\sigma(\mathcal{D}_n)} \to h'(t)$ in $L^2(0,1)$ by the L^2 martingale convergence theorem.

Furthermore, by Itô's isometry,

$$\sum_{i=0}^{2^n-1} \frac{\delta_i h}{\delta t} \cdot \delta_i X \;\to\; \int_0^1 h'(s)\, dX_s \qquad \text{in } L^2(\mu_0) \text{ as } n \to \infty. \tag{9.2.5}$$

Indeed, the sum on the right-hand side is the Itô integral of the step function $\left.\frac{dh}{dt}\right|_{\sigma(\mathcal{D}_n)}$ w.r.t. X, and as remarked above, these step functions converge to h' in $L^2(0,1)$. Along a subsequence, the convergence in (9.2.5) holds μ_0-almost surely, and hence by (9.2.4),

$$\lim_{n\to\infty} Z_n \;=\; \exp\left(\int_0^1 h'(s)\, dX_s - \frac{1}{2}\int_0^1 |h'(s)|^2\, ds\right) \qquad \mu_0\text{-a.s.} \tag{9.2.6}$$

(3). *Absolute continuity on $\mathcal{F}_1^{\mathbf{X}}$:* We still assume $h' \in L^2(0,1)$. Note that $\mathcal{F}_1^X = \sigma(\bigcup \mathcal{F}_n)$. Hence for proving that μ_h is absolutely continuous w.r.t. μ_0 on $\mathcal{F}_1^X$ with density given by (9.2.6), it suffices to show that $\limsup Z_n < \infty$ μ_h-almost surely (i.e., the singular part in the Lebesgue decomposition of μ_h w.r.t. μ_0 vanishes). Since $\mu_h = \mu_0 \circ \tau_h^{-1}$, the process

$$W_t \;=\; X_t - h(t) \qquad \text{is a Brownian motion w.r.t. } \mu_h,$$

and by (9.2.3) and (9.2.4),

$$Z_n \;=\; \exp\left(\sum_{i=0}^{2^n-1} \frac{\delta_i h}{\delta t} \cdot \delta_i W + \frac{1}{2}\sum_{i=0}^{2^n-1} \left|\frac{\delta_i h}{\delta t}\right|^2 \delta t\right).$$

Note that the minus sign in front of the second sum has turned into a plus by the translation! Arguing similarly as above, we see that along a subsequence, (Z_n) converges μ_h-almost surely to a finite limit:

$$\lim Z_n \;=\; \exp\left(\int_0^1 h'(s)\, dW_s + \frac{1}{2}\int_0^1 |h'(s)|^2\, ds\right) \qquad \mu_h\text{-a.s.}$$

Hence $\mu_h \ll \mu_0$ with density $\lim Z_n$.

(4). *Singularity on* $\mathcal{F}_1^X$*:* Conversely, let us suppose now that h is not absolutely continuous or h' is not in $L^2(0,1)$. Then

$$\sum_{i=0}^{2^n-1} \left|\frac{\delta_i h}{\delta_i t}\right|^2 \delta t \;=\; \int_0^1 \left|\frac{dh}{dt}\right|^2_{\sigma(\mathcal{D}_n)} dt \quad\longrightarrow\quad \infty \qquad \text{as } n\to\infty.$$

Since

$$\left\|\sum_{i=0}^{2^n-1} \frac{\delta_i h}{\delta_t}\cdot \delta_i X\right\|_{L^2(\mu_0)} \;=\; \left(\sum_{i=0}^{2^n-1}\left(\frac{\delta_i h}{\delta t}\right)^2 \delta t\right)^{1/2},$$

we can conclude by (9.2.4) that

$$\lim Z_n \;=\; 0 \qquad \mu_0\text{-almost surely,}$$

i.e., μ_h is singular w.r.t. μ_0.

□

In Section 11.5, we will give an alternative proof of the Cameron-Martin Theorem.

Passage times for Brownian motion with constant drift

We now consider a one-dimensional Brownian motion with constant drift β, i.e., a process

$$Y_t \;=\; B_t + \beta t, \qquad t \geq 0,$$

where B_t is a Brownian motion starting at 0 and $\beta \in \mathbb{R}$. We will apply the Cameron-Martin Theorem to compute the distributions of the first passage times

$$T_a^Y \;=\; \min\{t \geq 0 \,:\, Y_t = a\}, \qquad a > 0.$$

Note that T_a^Y is also the first passage time to the line $t \mapsto a - \beta t$ for the Brownian motion (B_t).

Theorem 9.7. *For $a > 0$ and $\beta \in \mathbb{R}$, the restriction of the distribution of T_a^Y to $(0, \infty)$ is absolutely continuous with density*

$$f_{a,\beta}(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{(a-\beta t)^2}{2t}\right).$$

In particular,

$$P[T_a^Y < \infty] = \int_0^\infty f_{a,\beta}(s)\, ds.$$

Proof. Let $h(t) = \beta t$. By the Cameron-Martin Theorem, the distribution μ_h of (Y_t) is absolutely continuous w.r.t. Wiener measure on $\mathcal{F}_t^X$ with density

$$Z_t = \exp(\beta \cdot X_t - \beta^2 t/2).$$

Therefore, denoting by $T_a = \inf\{t \geq 0 : X_t = a\}$ the passage time of the canonical process, we obtain

$$\begin{aligned} P[T_a^Y \leq t] &= \mu_h[T_a \leq t] = E_{\mu_0}[Z_t \,;\, T_a \leq t] \\ &= E_{\mu_0}[Z_{T_a} \,;\, T_a \leq t] = E_{\mu_0}[\exp(\beta a - \frac{1}{2}\beta^2 T_a) \,;\, T_a \leq t] \\ &= \int_0^t \exp(\beta a - \beta^2 s/2)\, f_{T_a}(s)\, ds \end{aligned}$$

by the optional sampling theorem. The claim follows by inserting the explicit expression for f_{T_a} derived in Corollary 1.25. □

9.3 Girsanov transform

We will now extend the results in the previous section 9.2 considerably. To this end, we will consider locally absolutely continuous changes of measure with local densities of type

$$Z_t = \exp\left(\int_0^t G_s \cdot dX_s - \frac{1}{2}\int_0^t |G_s|^2\, ds\right),$$

where (G_s) is an adapted process. Recall that the densities in the Cameron-Martin-Theorem took this form with the *deterministic* function $G_s = h'(s)$. We start with a general discussion about changing measure on filtered probability spaces that will be useful in other contexts as well.

Change of measure on filtered probability spaces

Let $(\mathcal{F}_t)$ be a filtration on a measurable space $(\Omega, \mathcal{A})$, and fix $t_0 \in (0, \infty)$. We consider two probability measures P and Q on $(\Omega, \mathcal{A})$ that are mutually absolutely continuous on the σ-algebra $\mathcal{F}_{t_0}$ with relative density

$$Z_{t_0} \quad = \quad \frac{dP}{dQ}\Big|_{\mathcal{F}_{t_0}} \quad > \quad 0 \qquad Q\text{-almost surely.}$$

Then P and Q are also mutually absolutely continuous on each of the σ-algebras $\mathcal{F}_t$, $t \leq t_0$, with Q- and P-almost surely strictly positive relative densities

$$Z_t \quad = \quad \frac{dP}{dQ}\Big|_{\mathcal{F}_t} \quad = \quad E_Q\left[Z_{t_0}\middle|\mathcal{F}_t\right] \qquad \text{and} \qquad \frac{dQ}{dP}\Big|_{\mathcal{F}_t} \quad = \quad \frac{1}{Z_t}.$$

The process $(Z_t)_{t\leq t_0}$ is a martingale w.r.t. Q, and, correspondingly, $(1/Z_t)_{t\leq t_0}$ is a martingale w.r.t. P. From now on, we always choose a right continuous version of these martingales.

Lemma 9.8. *1) For any $0 \leq s \leq t \leq t_0$, and for any $\mathcal{F}_t$-measurable random variable $X : \Omega \to [0, \infty]$,*

$$E_P[X|\mathcal{F}_s] \;=\; \frac{E_Q[XZ_t|\mathcal{F}_s]}{E_Q[Z_t|\mathcal{F}_s]} \;=\; \frac{E_Q[XZ_t|\mathcal{F}_s]}{Z_s} \qquad P\text{-a.s. and } Q\text{-a.s.} \quad (9.3.1)$$

2) Suppose that $(M_t)_{t\leq t_0}$ is an $(\mathcal{F}_t)$ adapted right continuous stochastic process. Then

(i) M is a martingale w.r.t. P $\quad\Leftrightarrow\quad$ $M \cdot Z$ is a martingale w.r.t. Q,

(ii) M is a local martingale w.r.t. P $\quad\Leftrightarrow\quad$ $M \cdot Z$ is a local martingale w.r.t. Q.

Proof. 1) The right hand side of (9.3.1) is $\mathcal{F}_s$-measurable. Moreover, for any $A \in \mathcal{F}_s$,

$$\begin{aligned} E_P[E_Q[XZ_t|\mathcal{F}_s]/Z_s\,;\,A] &= E_Q[E_Q[XZ_t|\mathcal{F}_s]\,;\,A] \\ &= E_Q[XZ_t\,;\,A] \quad = \quad E_Q[X\,;\,A]. \end{aligned}$$

2) (i) is a direct consequence of 1). Moreover, by symmetry, it is enough to prove the implication "$\Leftarrow$" in (ii). Hence suppose that $M \cdot Z$ is a local Q-martingale with localizing sequence (T_n). We show that M^{T_n} is a P-martingale, i.e.,

$$E_P[M_{t\wedge T_n}\,;\,A] \;=\; E_P[M_{s\wedge T_n}\,;\,A] \quad \text{for any } A \in \mathcal{F}_s,\;\; 0 \le s \le t \le t_0. \tag{9.3.2}$$

To verify (9.3.2), we first note that

$$E_P[M_{t\wedge T_n}\,;\,A\cap\{T_n \le s\}] \quad = \quad E_P[M_{s\wedge T_n}\,;\,A\cap\{T_n \le s\}] \tag{9.3.3}$$

since $t \wedge T_n = T_n = s \wedge T_n$ on $\{T_n \le s\}$. Moreover, one verifies from the definition of the σ-algebra $\mathcal{F}_{s\wedge T_n}$ that for any $A \in \mathcal{F}_s$, the event $A \cap \{T_n > s\}$ is contained in $\mathcal{F}_{s\wedge T_n}$, and hence in $\mathcal{F}_{t\wedge T_n}$. Therefore,

$$\begin{aligned} E_P[M_{t\wedge T_n}\,;\,A\cap\{T_n > s\}] \;&=\; E_Q[M_{t\wedge T_n}\,Z_{t\wedge T_n}\,;\,A\cap\{T_n > s\}] \\ =\; E_Q[M_{s\wedge T_n}\,Z_{s\wedge T_n}\,;\,A\cap\{T_n > s\}]] \;&=\; E_P[M_{s\wedge T_n}\,;\,A\cap\{T_n > s\}] \end{aligned} \tag{9.3.4}$$

by the martingale property for $(MZ)^{T_n}$, the optional sampling theorem, and the fact that $P \ll Q$ on $\mathcal{F}_{t\wedge T_n}$ with relative density $Z_{t\wedge T_n}$. (9.3.2) follows from (9.3.3) and (9.3.4). □

Girsanov's Theorem

We now return to our original problem of identifying the change of measure induced by a random translation of the paths of a Brownian motion. Suppose that (X_t) is a Brownian motion in $\mathbb{R}^d$ with $X_0 = 0$ w.r.t. the probability measure Q and the filtration $(\mathcal{F}_t)$, and fix $t_0 \in [0, \infty)$. Let

$$L_t \quad = \quad \int_0^t G_s \cdot dX_s, \quad t \ge 0,$$

with $G \in \mathcal{L}^2_{\text{a,loc}}(\mathbb{R}_+, \mathbb{R}^d)$. Then $[L]_t = \int_0^t |G_s|^2 \, ds$, and hence

$$Z_t \quad = \quad \exp\Big(\int_0^t G_s \cdot dX_s - \frac{1}{2} \int_0^t |G_s|^2 \, ds \Big) \tag{9.3.5}$$

is the exponential of L. In particular, since L is a local martingale w.r.t. Q, Z is a non-negative local martingale, and hence a supermartingale w.r.t. Q. It is a Q-martingale for $t \leq t_0$ if and only if $E_Q[Z_{t_0}] = 1$:

Exercise (Martingale property for exponentials). Let $(Z_t)_{t \in [0,t_0]}$ on $(\Omega, \mathcal{A}, Q)$ be a non-negative local martingale satisfying $Z_0 = 1$.

a) Show that Z is a supermartingale.

b) Prove that Z is a martingale if and only if $E_Q[Z_{t_0}] = 1$.

In order to use Z_{t_0} for changing the underlying probability measure on $\mathcal{F}_{t_0}$ we have to assume the martingale property:

Assumption. $(Z_t)_{t \leq t_0}$ is a martingale w.r.t. Q.

Theorem 9.10 below implies that the assumption is satisfied if

$$E\left[\exp\left(\frac{1}{2} \int_0^t |G_s|^2 \, ds \right) \right] \; < \; \infty.$$

If the assumption holds then we can consider a probability measure P on $\mathcal{A}$ with

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_{t_0}} \quad = \quad Z_{t_0} \qquad Q\text{-a.s.} \tag{9.3.6}$$

Note that P and Q are mutually absolutely continuous on $\mathcal{F}_t$ for any $t \leq t_0$ with

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_t} \quad = \quad Z_t \qquad \text{and} \qquad \frac{dQ}{dP}\Big|_{\mathcal{F}_t} \quad = \quad \frac{1}{Z_t}$$

both P- and Q-almost surely. We are now ready to prove one of the most important results of stochastic analysis:

Theorem 9.9 (Maruyama 1954, Girsanov 1960). *Suppose that X is a d-dimensional Brownian motion w.r.t. Q and $(Z_t)_{t\le t_0}$ is defined by (9.3.5) with $G \in \mathcal{L}^2_{\mathrm{a,loc}}(\mathbb{R}_+,\mathbb{R}^d)$. If $E_Q[Z_{t_0}] = 1$ then the process*

$$B_t \quad := \quad X_t - \int_0^t G_s\,ds, \qquad t \le t_0,$$

is a Brownian motion w.r.t. any probability measure P on $\mathcal{A}$ satisfying (9.3.6).

Proof. By the extension of Lévy's characterization of Brownian motion to the multi-dimensional case, it suffices to show that $(B_t)_{t\le t_0}$ is an $\mathbb{R}^d$-valued P-martingale with $[B^i,B^j]_t = \delta_{ij}t$ P-almost surely for any $i,j \in \{1,\dots,d\}$, cf. Theorem 11.2 below. Furthermore, by Lemma 9.8, and since P and Q are mutually absolutely continuous on $\mathcal{F}_{t_0}$, this holds true provided $(B_tZ_t)_{t\le t_0}$ is an $\mathbb{R}^d$ valued local martingale under Q, and $[B^i,B^j] = \delta_{ij}t$ Q-almost surely. The identity for the covariations holds since (B_t) differs from the Q-Brownian motion (X_t) only by a continuous finite variation process. To show that $B\cdot Z$ is a local Q-martingale, we apply Itô's formula: For $1\le i\le d$,

$$\begin{aligned} d(B^i\,Z) &= B^i\,dZ + Z\,dB^i + d[B^i,Z] \\ &= B^iZG\cdot dX + Z\,dX^i - Z\,G^i dt + ZG^i\,dt, \end{aligned} \tag{9.3.7}$$

where we have used that

$$d[B^i,Z] \quad = \quad ZG\cdot d[B^i,X] \quad = \quad ZG^i\,dt \qquad Q\text{-almost surely.}$$

The right-hand side of (9.3.7) is a stochastic integral w.r.t. the Q-Brownian motion X, and hence a local Q-martingale. □

The theorem shows that if X is a Brownian motion w.r.t. Q, and Z defined by (9.3.5) is a Q-martingale, then X satisfies

$$dX_t \quad = \quad G_t\,dt\,+\,dB_t.$$

with a P-Brownian motion B. This can be used to construct weak solutions of stochastic differential equations by changing the underlying probability measure, see Section 11.3

below. For instance, if we choose $G_t = b(X_t)$ then the Q-Brownian motion (X_t) is a solution to the SDE

$$dX_t \quad = \quad b(X_t)\, dt \,+\, dB_t,$$

where B is a Brownian motion under the modified probability measure P.

Furthermore, Girsanov's Theorem generalizes the Cameron-Martin Theorem to non-deterministic adapted translations

$$X_t(\omega) \quad \longrightarrow \quad X_t(\omega) - H_t(\omega), \qquad H_t \;=\; \int_0^t G_s\, ds,$$

of a Brownian motion X.

Remark (Assumptions in Girsanov's Theorem).
1) Absolute continuity and adaptedness of the "translation process" $H_t = \int_0^t G_s\, ds$ are essential for the assertion of Theorem 9.9.
2) The assumption $E_Q[Z_{t_0}] = 1$ ensuring that $(Z_t)_{t\le t_0}$ is a Q-martingale is not always satisfied $-$ a sufficient condition is given in Theorem 9.10 below. If (Z_t) is not a martingale w.r.t. Q it can still be used to define a positive measure P_t with density Z_t w.r.t. Q on each σ-algebra $\mathcal{F}_t$. However, in this case, $P_t[\Omega] < 1$. The sub-probability measures P_t correspond to a transformed process with finite life-time.

Novikov's condition

To verify the assumption in Girsanov's theorem, we now derive a sufficient condition for ensuring that the exponential

$$Z_t \quad = \quad \exp\big(L_t - 1/2\,[L]_t\big)$$

of a continuous local $(\mathcal{F}_t)$ martingale (L_t) is a martingale. Recall that Z is always a non-negative local martingale, and hence a supermartingale w.r.t. $(\mathcal{F}_t)$.

Theorem 9.10 (Novikov 1971). *Let $t_0 \in \mathbb{R}_+$. If $E[\exp\big([L]_{t_0}/2\big)] < \infty$ then $(Z_t)_{t\le t_0}$ is an $(\mathcal{F}_t)$ martingale.*

We only prove the theorem under the slightly more restrictive condition

$$E\left[\exp(p[L]_t/2)\right] \;<\; \infty \qquad \text{for some } p > 1. \tag{9.3.8}$$

This simplifies the proof considerably, and the condition is sufficient for many applications. For a proof in the general case and under even weaker assumptions see e.g. [37].

Proof. Let $(T_n)_{n\in\mathbb{N}}$ be a localizing sequence for the martingale Z. Then $(Z_{t\wedge T_n})_{t\geq 0}$ is a martingale for any n. To carry over the martingale property to the process $(Z_t)_{t\in[0,t_0]}$, it is enough to show that the random variables $Z_{t\wedge T_n}$, $n \in \mathbb{N}$, are uniformly integrable for each fixed $t \leq t_0$. However, for $c > 0$ and $p, q \in (1, \infty)$ with $p^{-1} + q^{-1} = 1$, we have

$$\begin{aligned}
&E[Z_{t\wedge T_n}\;;\;Z_{t\wedge T_n} \geq c] \\
&\quad = \quad E\big[\exp\big(L_{t\wedge T_n} - \frac{p}{2}[L]_{t\wedge T_n}\big)\ \exp\big(\frac{p-1}{2}[L]_{t\wedge T_n}\big)\;;\;Z_{t\wedge T_n} \geq c\big] \qquad (9.3.9)\\
&\quad \leq \quad E\big[\exp\big(pL_{t\wedge T_n} - \frac{p^2}{2}[L]_{t\wedge T_n}\big)\big]^{1/p} \cdot E\big[\exp\big(q\cdot\frac{p-1}{2}[L]_{t\wedge T_n}\big)\;;\;Z_{t\wedge T_n} \geq c\big]^{1/q} \\
&\quad \leq \quad E\big[\exp\big(\frac{p}{2}[L]_t\big)\;;\;Z_{t\wedge T_n} \geq c\big]^{1/q}
\end{aligned}$$

for any $n \in \mathbb{N}$. Here we have used Hölder's inequality and the fact that $\exp\big(pL_{t\wedge T_n} - \frac{p^2}{2}[L]_{t\wedge T_n}\big)$ is an exponential supermartingale. If $\exp\big(\frac{p}{2}[L]_t\big)$ is integrable then the right hand side of (9.3.9) converges to 0 uniformly in n as $c \to \infty$, because

$$P[Z_{t\wedge T_n} \geq 0] \quad \leq \quad c^{-1}\,E[Z_{t\wedge T_n}] \quad \leq \quad c^{-1} \quad \longrightarrow \quad 0$$

uniformly in n as $c \to \infty$. Hence $\{Z_{t\wedge T_n} : n \in \mathbb{N}\}$ is indeed uniformly integrable, and thus $(Z_t)_{t\in[0,t_0]}$ is a martingale. □

Example (Bounded drifts). If $L_t = \int_0^t G_s \cdot dX_s$ with a Brownian motion (X_t) and an adapted process (G_t) that is uniformly bounded on $[0, t]$ for any finite t then the quadratic variation $[L]_t = \int_0^t |G_s|^2\, ds$ is also bounded for finite t. Hence $\exp(L - \frac{1}{2}[L])$ is an $(\mathcal{F}_t)$ martingale for $t \in [0, \infty)$.

Example (Option pricing in continuous time II: Risk-neutral measure). XXX to be included

9.4 Itô's Representation Theorem and Option Pricing

We now prove two basic representation theorems for functionals and martingales that are adapted w.r.t. the filtration generated by a Brownian motion. Besides their intrinsic interest, such representation theorems are relevant e.g. for the theory of financial markets, and for stochastic filtering. Throughout this section, (B_t) denotes a Brownian motion starting at 0 on a probability space $(\Omega, \mathcal{A}, P)$, and

$$\mathcal{F}_t \quad = \quad \sigma(B_s : s \in [0,t])^P, \qquad t \geq 0,$$

is the completed filtration generated by (B_t). It is crucial that the filtration does not contain additional information. By the factorization lemma, this implies that $\mathcal{F}_t$ measurable random variables $F : \Omega \to \mathbb{R}$ are almost surely functions of the Brownian path $(B_s)_{s \leq t}$. Indeed, we will show that such functions can be represented as stochastic integrals.

Representation theorems for functions and martingales

The first version of Itô's Representation Theorem states that random variables that are measurable w.r.t. the σ-algebra $\mathcal{F}_1 = \mathcal{F}_1^{B,P}$ can be represented as stochastic integrals:

Theorem 9.11 (Itô). *For any function $F \in \mathcal{L}^2(\Omega, \mathcal{F}_1, P)$ there exists a unique process $G \in L^2_a(0,1)$ such that*

$$F \quad = \quad E[F] + \int_0^1 G_s \cdot dB_s \qquad P\text{-almost surely.} \tag{9.4.1}$$

An immediate consequence of Theorem 9.11 is a corresponding representation for martingales **w.r.t. the Brownian filtration** $\mathcal{F}_t = \mathcal{F}_t^{B,P}$:

Corollary 9.12 (Itô representation for martingales). *For any right-continuous L^2-bounded $(\mathcal{F}_t)$ martingale $(M_t)_{t\in[0,1]}$ there exists a unique process $G \in L_a^2(0,1)$ such that*

$$M_t \quad = \quad M_0 + \int_0^t G_s \cdot dB_s \qquad \textit{for any } t \in [0,1], \qquad P\textit{-a.s.}$$

The corollary is of fundamental importance in financial mathematics where it is related to completeness of financial markets. It also proves the remarkable fact that **every martingale w.r.t. the Brownian filtration has a continuous modification**! Of course, this result can not be true w.r.t. a general filtration.

We first show that the corollary follows from Theorem 9.11, and then we prove the theorem:

Proof of Corollary 9.12. If $(M_t)_{t\in[0,1]}$ is an L^2 bounded $(\mathcal{F}_t)$ martingale then $M_1 \in \mathcal{L}^2(\Omega, \mathcal{F}_1, P)$, and

$$M_t \quad = \quad E[M_1|\mathcal{F}_t] \qquad \text{a.s.} \quad \text{for any } t \in [0,1].$$

Hence, by Theorem 9.11, there exists a unique process $G \in L_a^2(0,1)$ such that

$$M_1 \quad = \quad E[M_1] + \int_0^1 G \cdot dB \quad = \quad M_0 + \int_0^1 G \cdot dB \qquad \text{a.s.,}$$

and thus

$$M_t \quad = \quad E[M_1|\mathcal{F}_t] \quad = \quad M_0 + \int_0^t G \cdot dB \qquad \text{a.s. for any } t \in [0,1].$$

Since both sides in the last equation are almost surely right continuous, the identity actually holds simultaneously for all $t \in [0,1]$ with probability 1. $\square$

Proof of Theorem 9.11. **Uniqueness.** Suppose that (9.4.1) holds for two processes $G, \widetilde{G} \in L_a^2(0,1)$. Then

$$\int_0^1 G \cdot dB \quad = \quad \int_0^1 \widetilde{G} \cdot dB,$$

and hence, by Itô's isometry,

$$||G - \widetilde{G}||_{L^2(P\otimes\lambda)} \quad = \quad \Big|\Big| \int (G - \widetilde{G}) \cdot dB \Big|\Big|_{L^2(P)} \quad = \quad 0.$$

Hence $G_t(\omega) = \widetilde{G}_t(\omega)$ for almost every (t, ω).

Existence. We prove the existence of a representation as in (9.4.1) in several steps − starting with "simple" functions F.

1. Suppose that $F = \exp(ip \cdot (B_t - B_s))$ for some $p \in \mathbb{R}^d$ and $0 \le s \le t \le 1$. By Itô's formula,

$$\exp(ip \cdot B_t + \frac{1}{2}|p|^2 t) \quad = \quad \exp(ip \cdot B_s + \frac{1}{2}|p|^2 s) + \int_s^t \exp\big(ip \cdot B_r + \frac{1}{2}|p|^2 r\big) ip \cdot dB_r.$$

Rearranging terms, we obtain an Itô representation for F with a bounded adapted integrand G.

2. Now suppose that $F = \prod\limits_{k=1}^{n} F_k$ where $F_k = \exp\big(ip_k \cdot (B_{t_k} - B_{t_{k-1}})\big)$ for some $n \in \mathbb{N}$, $p_1, \ldots, p_n \in \mathbb{R}^d$, and $0 \le t_0 \le t_1 \le \cdots \le t_n \le 1$. Denoting by G_k the bounded adapted process in the Itô representation for F_k, we have

$$F \quad = \quad \prod_{k=1}^{n} \Big(E[F_k] + \int_{t_k}^{t_{k+1}} G^k \cdot dB \Big).$$

We show that the right hand side can be written as the sum of $\prod_{k=1}^n E[F_k]$ and a stochastic integral w.r.t. B. For this purpose, it suffices to verify that the product of two stochastic integrals $X_t = \int_0^t G \cdot dB$ and $Y_t = \int_0^t H \cdot dB$ with bounded adapted processes G and H is the stochastic integral of a process in $L^2_a(0, 1)$ provided $\int_0^1 G_t \cdot H_t \, dt = 0$. This holds true, since by the product rule,

$$X_1 Y_1 \quad = \quad \int_0^1 X_t H_t \cdot dB_t + \int_0^1 Y_t G_t \cdot dB_t + \int_0^1 G_t \cdot H_t \, dt,$$

and $XH + YG$ is square-integrable by Itô's isometry.

3. Clearly, an Itô representation also holds for any linear combination of functions as in Step 2.

4. To prove an Itô representation for arbitrary functions in $\mathcal{L}^2(\Omega, \mathcal{F}_1, P)$, we first note

that the linear combinations of the functions in Step 2 form a *dense* subspace of the Hilbert space $L^2(\Omega, \mathcal{F}_1, P)$. Indeed, if ϕ is an element in $L^2(\Omega, \mathcal{F}_1, P)$ that is orthogonal to this subspace then

$$E\Big[\phi \prod_{k=1}^{n} \exp(ip_k \cdot (B_{t_k} - B_{t_{k-1}}))\Big] \quad = \quad 0$$

for any $n \in \mathbb{N}$, $p_1, \ldots, p_n \in \mathbb{R}^d$ and $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n \leq 1$. By Fourier inversion, this implies

$$E[\phi \mid \sigma(B_{t_k} - B_{t_{k-1}} : 1 \leq k \leq n)] \quad = \quad 0 \qquad \text{a.s.}$$

for any $n \in \mathbb{N}$ and $0 \leq t_0 \leq \cdots \leq t_n \leq 1$, and hence $\phi = 0$ a.s. by the Martingale Convergence Theorem.

Now fix an arbitrary function $F \in L^2(\Omega, \mathcal{F}_1, P)$. Then by Step 3, there exists a sequence (F_n) of functions in $L^2(\Omega, \mathcal{F}_1, P)$ converging to F in L^2 that have a representation of the form

$$F_n - E[F_n] \quad = \quad \int_0^1 G^{(n)} \cdot dB \tag{9.4.2}$$

with processes $G^{(n)} \in L^2_a(0,1)$. As $n \to \infty$,

$$F_n - E[F_n] \quad \longrightarrow \quad F - E[F] \qquad \text{in } \ L^2(P).$$

Hence, by (9.4.2) and Itô's isometry, $(G^{(n)})$ is a Cauchy sequence in $L^2(P \otimes \lambda_{(0,1)})$. Denoting by G the limit process, we obtain the representation

$$F - E[F] \quad = \quad \int_0^1 G \cdot dB$$

by taking the L^2 limit on both sides of (9.4.2). □

Application to option pricing

XXX to be included

Application to stochastic filtering

XXX to be included

Chapter 10

Lévy processes and Poisson point processes

A widely used class of possible discontinuous driving processes in stochastic differential equations are Lévy processes. They include Brownian motion, Poisson and compound Poisson processes as special cases. In this chapter, we outline basics from the theory of Lévy processes, focusing on prototypical examples of Lévy processes and their construction. For more details we refer to the monographs of Applebaum [5] and Bertoin [8].

Apart from simple transformations of Brownian motion, Lévy processes do not have continuous paths. Instead, we will assume that the paths are **càdlàg (continue à droite, limites à gauche)**, i.e., right continuous with left limits. This can always be assured by choosing an appropriate modification. We now summarize a few notations and facts about càdlàg functions that are frequently used below. If $x : I \to \mathbb{R}$ is a càdlàg function defined on a real interval I, and s is a point in I except the left boundary point, then we denote by

$$x_{s-} \quad = \quad \lim_{\varepsilon \downarrow 0} x_{s-\varepsilon}$$

the left limit of x at s, and by

$$\Delta x_s \quad = \quad x_s - x_{s-}$$

the size of the jump at s. Note that the function $s \mapsto x_{s-}$ is left continuous with right limits. Moreover, x is continuous if and only if $\Delta x_s = 0$ for all s. Let $\mathcal{D}(I)$ denote the linear space of all càdlàg functions $x : I \to \mathbb{R}$.

Exercise (Càdlàg functions). Prove the following statements:

1) If I is a compact interval, then for any function $x \in \mathcal{D}(I)$, the set

$$\{s \in I : |\Delta x_s| > \varepsilon\}$$

is finite for any $\varepsilon > 0$. Conclude that any function $x \in \mathcal{D}([0, \infty))$ has at most countably many jumps.

2) A càdlàg function defined on a compact interval is bounded.
3) A uniform limit of a sequence of càdlàg functions is again càdlàg .

10.1 Lévy processes

Lévy processes are $\mathbb{R}^d$-valued stochastic processes with stationary and independent increments. More generally, let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{A}, P)$.

Definition. *An $(\mathcal{F}_t)$ Lévy process is an $(\mathcal{F}_t)$ adapted càdlàg stochastic process $X_t : \Omega \to \mathbb{R}^d$ such that w.r.t. P,*

(a) $X_{s+t} - X_s$ *is independent of* $\mathcal{F}_s$ *for any* $s, t \geq 0$, *and*

(b) $X_{s+t} - X_s \quad \sim \quad X_t - X_0 \qquad$ *for any* $s, t \geq 0$.

Any Lévy process (X_t) is also a Lévy process w.r.t. the filtration $(\mathcal{F}_t^X)$ generated by the process. Often continuity in probability is assumed instead of càdlàg sample paths. It can then be proven that a càdlàg modification exists, cf. [36, Ch.I Thm.30].

Remark (Lévy processes in discrete time are Random Walks). A discrete-time process $(X_n)_{n=0,1,2,...}$ with stationary and independent increments is a Random Walk: $X_n = X_0 + \sum_{j=1}^n \eta_j$ with i.i.d. increments $\eta_j = X_j - X_{j-1}$.

Remark (Lévy processes and infinite divisibility). The increments $X_{s+t} - X_s$ of a Lévy process are **infinitely divisible** random variables, i.e., for any $n \in \mathbb{N}$ there exist i.i.d. random variables $Y_1, \dots, Y_n$ such that $X_{s+t} - X_s$ has the same distribution as $\sum\limits_{i=1}^n Y_i$. Indeed, we can simply choose $Y_i = X_{s+it/n} - X_{s+i(t-1)/n}$. The Lévy-Khinchin formula gives a characterization of all distributions of infinitely divisible random variables, cf. e.g. [5]. The simplest examples of infinitely divisible distributions are normal and Poisson distributions.

Characteristic exponents

We now restrict ourselves w.l.o.g. to Lévy processes with $X_0 = 0$. The distribution of the sample paths is then uniquely determined by the distributions of the increments $X_t - X_0 = X_t$ for $t \geq 0$. Moreover, by stationarity and independence of the increments we obtain the following representation for the characteristic functions $\phi_t(p) = E[\exp(ip \cdot X_t)]$:

Theorem 10.1 (Characteristic exponent). *If $(X_t)_{t\geq 0}$ is a Lévy process with $X_0 = 0$ then there exists a continuous function $\psi : \mathbb{R}^d \to \mathbb{C}$ with $\psi(0) = 0$ such that*

$$E[e^{ip\cdot X_t}] = e^{-t\psi(p)} \qquad \text{for any } t \geq 0 \text{ and } p \in \mathbb{R}^d. \tag{10.1.1}$$

Moreover, if (X_t) has finite first or second moments, then ψ is $\mathcal{C}^1$, $\mathcal{C}^2$ respectively, and

$$E[X_t] = it\nabla\psi(0) \quad , \quad \mathrm{Cov}[X_t^k, X_t^l] = t\,\frac{\partial^2\psi}{\partial p_k \partial p_l}(0) \tag{10.1.2}$$

for any $k, l = 1, \dots, d$ and $t \geq 0$.

Proof. Stationarity and independence of the increments implies the identity

$$\begin{aligned} \phi_{t+s}(p) = E[\exp(ip\cdot X_{t+s})] &= E[\exp(ip\cdot X_s)]\cdot E[\exp(ip\cdot(X_{t+s}-X_s))] \\ &= \phi_t(p)\cdot\phi_s(p) \end{aligned} \tag{10.1.3}$$

for any $p\in\mathbb{R}^d$ and $s,t\geq 0$. For a given $p\in\mathbb{R}^d$, right continuity of the paths and dominated convergence imply that $t\mapsto\phi_t(p)$ is right-continuous. Since

$$\phi_{t-\varepsilon}(p) \quad = \quad E[\exp(ip\cdot(X_t-X_\varepsilon))],$$

the function $t\mapsto\phi_t(p)$ is also left continuous, and hence continuous. By (10.1.3) and since $\phi_0(p)=1$, we can now conclude that for each $p\in\mathbb{R}^d$, there exists $\psi(p)\in\mathbb{C}$ such that (10.1.1) holds. Arguing by contradiction we then see that $\psi(0)=0$ and ψ is continuous, since otherwise ϕ_t would not be continuous for all t.
Moreover, if X_t is (square) integrable then ϕ_t is $\mathcal{C}^1$ (resp. $\mathcal{C}^2$), and hence ψ is also $\mathcal{C}^1$ (resp. $\mathcal{C}^2$). The formulae in (10.1.2) for the first and second moment now follow by computing the derivatives w.r.t. p at $p=0$ in (10.1.1). □

The function ψ is called the **characteristic exponent** of the Lévy process.

Basic examples

We now consider first examples of continuous and discontinuous Lévy processes.

Example (Brownian motion and Gaussian Lévy processes). A d-dimensional Brownian motion (B_t) is by definition a continuous Lévy process with

$$B_t-B_s\sim N(0,(t-s)I_d) \qquad \text{for any } 0\leq s<t.$$

Moreover, $X_t=\sigma B_t+bt$ is a Lévy process with normally distributed marginals for any $\sigma\in\mathbb{R}^{d\times d}$ and $b\in\mathbb{R}^d$. Note that these Lévy processes are precisely the driving processes in SDE considered so far. The characteristic exponent of a Gaussian Lévy process is given by

$$\psi(p) \quad = \quad \frac{1}{2}|\sigma^T p|^2 - ib\cdot p \quad = \quad \frac{1}{2}p\cdot ap - ib\cdot p \qquad \text{with } a=\sigma\sigma^T.$$

First examples of discontinuous Lévy processes are Poisson and, more generally, compound Poisson processes.

Example (Poisson processes). The most elementary example of a pure jump Lévy process in continuous time is the Poisson process. It takes values in $\{0, 1, 2, \dots\}$ and jumps up one unit each time after an exponentially distributed waiting time. Explicitly, a Poisson process $(N_t)_{t\geq 0}$ with intensity $\lambda > 0$ is given by

$$N_t \quad = \quad \sum_{n=1}^{\infty} I_{\{S_n \leq t\}} \quad = \quad \sharp\{n \in \mathbb{N} \,:\, S_n \leq t\} \tag{10.1.4}$$

where $S_n = T_1 + T_2 + \cdots + T_n$ with independent random variables $T_i \sim \text{Exp}(\lambda)$. The increments $N_t - N_s$ of a Poisson process over disjoint time intervals are independent and Poisson distributed with parameter $\lambda(t-s)$, cf. [13, Satz 10.12]. Note that by (10.1.4), the sample paths $t \mapsto N_t(\omega)$ are càdlàg. In general, any Lévy process with

$$X_t - X_s \quad \sim \quad \text{Poisson}\,(\lambda(t-s)) \qquad \text{for any } 0 \leq s \leq t$$

is called a **Poisson process with intensity $\boldsymbol{\lambda}$**, and can be represented as above. The characteristic exponent of a Poisson process with intensity λ is

$$\psi(p) \quad = \quad \lambda(1 - e^{ip}).$$

The paths of a Poisson process are increasing and hence of finite variation. The **compensated Poisson process**

$$M_t \quad := \quad N_t - E[N_t] \quad = \quad N_t - \lambda t$$

is an $(\mathcal{F}_t^N)$ martingale, yielding the semimartingale decomposition

$$N_t \quad = \quad M_t + \lambda t$$

with the continuous finite variation part λt. On the other hand, there is the alternative trivial semimartingale decomposition $N_t = 0 + N_t$ with vanishing martingale part. This demonstrates that without an additional regularity condition, the semimartingale decomposition of discontinuous processes is not unique. A compensated Poisson process is a Lévy process which has both a continuous and a pure jump part.

Exercise (Martingales of Poisson processes). Prove that the compensated Poisson process $M_t = N_t - \lambda t$ and the process $M_t^2 - \lambda t$ are $(\mathcal{F}_t^N)$ martingales.

Any linear combination of independent Lévy processes is again a Lévy process:

Example (Superpositions of Lévy processes). If (X_t) and (X_t') are independent Lévy processes with values in $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ then $\alpha X_t + \beta X_t'$ is a Lévy process with values in $\mathbb{R}^n$ for any constant matrices $\alpha \in \mathbb{R}^{n\times d}$ and $\beta \in \mathbb{R}^{n\times d'}$. The characteristic exponent of the superposition is

$$\psi_{\alpha X+\beta X'}(p) \quad = \quad \psi_X(\alpha^T p) + \psi_Y(\beta^T p).$$

For example, linear combinations of independent Brownian motions and Poisson processes are again Lévy processes.

Compound Poisson processes

Next we consider general Lévy processes with paths that are constant apart from a finite number of jumps in finite time. We will see that such processes can be represented as compound Poisson processes. A compound Poisson process is a continuous time Random Walk defined by

$$X_t = \sum_{j=1}^{N_t} \eta_j \qquad , \quad t \geq 0,$$

with a Poisson process (N_t) of intensity $\lambda > 0$ and with independent identically distributed random variables $\eta_j : \Omega \to \mathbb{R}^d$ $(j \in \mathbb{N})$ that are independent of the Poisson process as well. The process (X_t) is again a pure jump process with jump times that do not accumulate. It has jumps of size y with intensity

$$\nu(dy) \quad = \quad \lambda\, \pi(dy),$$

where π denotes the joint distribution of the random variables η_j.

Lemma 10.2. *A compound Poisson process is a Lévy process with characteristic exponent*

$$\psi(p) = \int (1 - e^{ip\cdot y})\, \nu(dy). \tag{10.1.5}$$

Proof. Let $0 = t_0 < t_1 < \cdots < t_n$. Then the increments

$$X_{t_k} - X_{t_{k-1}} = \sum_{j=N_{t_{k-1}}+1}^{N_{t_k}} \eta_j \quad , \quad k = 1, 2, \ldots, n \, , \tag{10.1.6}$$

are conditionally independent given the σ-algebra generated by the Poisson process $(N_t)_{t\geq 0}$. Therefore, for $p_1, \ldots, p_n \in \mathbb{R}^d$,

$$\begin{aligned} E\big[\exp\big(i\sum_{k=1}^n p_k \cdot (X_{t_k} - X_{t_{k-1}})\big) \mid (N_t)\big] &= \prod_{k=1}^n E[\exp(ip_k \cdot (X_{t_k} - X_{t_{k-1}}) \mid (N_t)] \\ &= \prod_{k=1}^n \phi(p_k)^{N_{t_k} - N_{t_{k-1}}}, \end{aligned}$$

where ϕ denotes the characteristic function of the jump sizes η_j. By taking the expectation on both sides, we see that the increments in (10.1.6) are independent and stationary, since the same holds for the Poisson process (N_t). Moreover, by a similar computation,

$$\begin{aligned} E[\exp(ip \cdot X_t)] &= E[E[\exp(ip \cdot X_t) \mid (N_s)]] = E[\phi(p)^{N_t}] \\ &= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \phi(p)^k = e^{\lambda t(\phi(p)-1)} \end{aligned}$$

for any $p \in \mathbb{R}^d$, which proves (10.1.5). □

The paths of a compound Poisson process are of finite variation and càdlàg. One can show that every pure jump Lévy process with finitely many jumps in finite time is a compound Poisson process , cf. Theorem 10.15 below.

Exercise (Martingales of compound Poisson processes). Show that the following processes are martingales:

(a) $M_t = X_t - bt$ where $b = \int y \, \nu(dy)$ provided $\eta_1 \in \mathcal{L}^1$,

(b) $|M_t|^2 - at$ where $a = \int |y|^2 \, \nu(dy)$ provided $\eta_1 \in \mathcal{L}^2$.

We have shown that a compound Poisson process with jump intensity measure $\nu(dy)$ is a Lévy process with characteristic exponent

$$\psi_\nu(p) = \int (1 - e^{ip\cdot y}) \nu(dy) \quad , \quad p \in \mathbb{R}^d. \tag{10.1.7}$$

Since the distribution of a Lévy process on the space $\mathcal{D}([0,\infty),\mathbb{R}^d)$ of càdlàg paths is uniquely determined by its characteristic exponent, we can prove conversely:

Lemma 10.3. *Suppose that ν is a finite positive measure on $\mathcal{B}\big(\mathbb{R}^d\setminus\{0\}\big)$ with total mass $\lambda=\nu(\mathbb{R}^d\setminus\{0\})$, and (X_t) is a Lévy process with $X_0=0$ and characteristic exponent ψ_ν, defined on a complete probability space $(\Omega,\mathcal{A},\mathcal{P})$. Then there exists a sequence $(\eta_j)_{j\in\mathbb{N}}$ of i.i.d. random variables with distribution $\lambda^{-1}\nu$ and an independent Poisson Process (N_t) with intensity λ on $(\Omega,\mathcal{A},P)$ such that almost surely,*

$$X_t \quad = \quad \sum_{j=1}^{N_t}\eta_j \quad . \tag{10.1.8}$$

Proof. Let $(\widetilde{\eta}_j)$ be an arbitrary sequence of i.i.d. random variables with distribution $\lambda^{-1}\nu$, and let $(\widetilde{N_t})$ be an independent Poisson process of intensity $\nu(\mathbb{R}^d\setminus\{0\})$, all defined on a probability space $(\widetilde{\Omega},\widetilde{\mathcal{A}},\widetilde{P})$. Then the compound Poisson process $\widetilde{X_t}=\sum_{j=1}^{\widetilde{N_t}}\widetilde{\eta}_j$ is also a Lévy process with $\widetilde{X_0}=0$ and characteristic exponent ψ_ν. Therefore, the finite dimensional marginals of (X_t) and $(\widetilde{X_t})$, and hence the distributions of (X_t) and $(\widetilde{X_t})$ on $\mathcal{D}([0,\infty),\mathbb{R}^d)$ coincide. In particular, almost every path $t\mapsto X_t(\omega)$ has only finitely many jumps in a finite time interval, and is constant inbetween. Now set $S_0=0$ and let

$$S_j \quad = \quad \inf\{s>S_{j-1}\ :\ \Delta X_s\neq 0\} \qquad \text{for } j\in\mathbb{N}$$

denote the successive jump-times of (X_t). Then (S_j) is a sequence of non-negative random variables on $(\Omega,\mathcal{A},P)$ that is almost surely finite and strictly increasing with $\lim S_j=\infty$. Defining $\eta_j:=\Delta X_{S_j}$ if $S_j<\infty$, $\eta_j=0$ otherwise, and

$$N_t \quad := \quad \big|\{s\in(0,t]\ :\ \Delta X_s\neq 0\}\big| \quad = \quad \big|\{j\in\mathbb{N}\ :\ S_j\leq t\}\big|,$$

as the successive jump sizes and the number of jumps up to time t, we conclude that almost surely, (N_t) is finite, and the representation (10.1.8) holds. Moreover, for any $j\in\mathbb{N}$ and $t\geq 0$, η_j and N_t are measurable functions of the process $(X_t)_{t\geq 0}$. Hence the joint distribution of all these random variables coincides with the joint distribution of the random variables $\widetilde{\eta}_j$ $(j\in\mathbb{N})$ and $\widetilde{N_t}$ $(t\geq 0)$, which are the corresponding measurable functions of the process $(\widetilde{X_t})$. We can therefore conclude that $(\eta_j)_{j\in\mathbb{N}}$ is a sequence of i.i.d. random variables with distributions $\lambda^{-1}\nu$ and (N_t) is an independent Poisson process with intensity λ. $\square$

The lemma motivates the following formal definition of a compound Poisson process:

Definition. *Let ν be a finite positive measure on $\mathbb{R}^d$, and let $\psi_\nu : \mathbb{R}^d \to \mathbb{C}$ be the function defined by (10.1.7).*

1) *The unique probability measure π_ν on $\mathcal{B}(\mathbb{R}^d)$ with characteristic function*

$$\int e^{ip\cdot y}\,\pi_\nu(dy) \quad = \quad \exp(-\psi_\nu(p)) \qquad \forall\, p \in \mathbb{R}^d$$

is called the ***compound Poisson distribution with intensity measure ν***.

2) *A Lévy process (X_t) on $\mathbb{R}^d$ with $X_{s+t} - X_s \sim \pi_{t\nu}$ for any $s, t \geq 0$ is called a* ***compound Poisson process with jump intensity measure (Lévy measure) ν***.

The compound Poisson distribution π_ν is the distribution of $\sum_{j=1}^{K} \eta_j$ where K is a Poisson random variable with parameter $\lambda = \nu(\mathbb{R}^d)$ and (η_j) is a sequence of i.i.d. random variables with distribution $\lambda^{-1}\nu$. By conditioning on the value of K , we obtain the explicit series representation

$$\pi_\nu \quad = \quad \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \nu^{*k},$$

where ν^{*k} denotes the k-fold convolution of ν.

Examples with infinite jump intensity

The Lévy processes considered so far have only a finite number of jumps in a finite time interval. However, by considering limits of Lévy processes with finite jump intensity, one also obtains Lévy processes that have infinitely many jumps in a finite time interval. We first consider two important classes of examples of such processes:

Example (Inverse Gaussian subordinators). Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion with $B_0 = 0$ w.r.t. a right continuous filtration $(\mathcal{F}_t)$, and let

$$T_s \quad = \quad \inf\{t \geq 0 \ : \ B_t = s\}$$

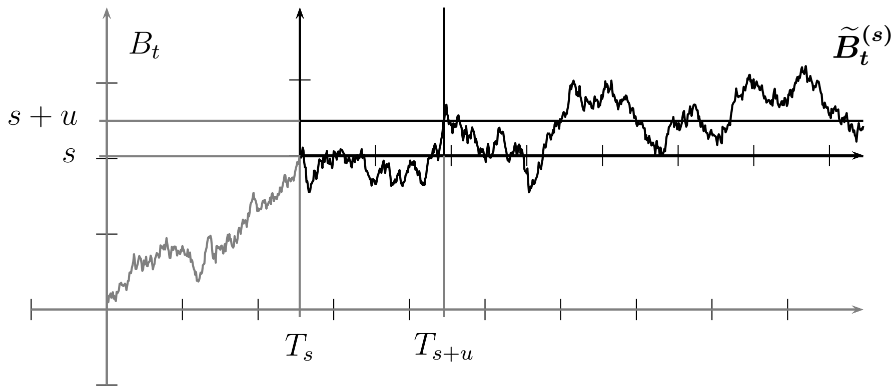
denote the first passage time to a level $s \in \mathbb{R}$. Then $(T_s)_{s\geq 0}$ is an increasing stochastic process that is adapted w.r.t. the filtration $(\mathcal{F}_{T_s})_{s\geq 0}$. For any ω, $s \mapsto T_s(\omega)$ is the generalized left-continuous inverse of the Brownian path $t \mapsto B_t(\omega)$. Moreover, by the strong Markov property, the process

$$\widetilde{B}^{(s)}_t \quad := \quad B_{T_s+t} - B_{T_s} \qquad , \ t \geq 0,$$

is a Brownian motion independent of $\mathcal{F}_{T_s}$ for any $s \geq 0$, and

$$T_{s+u} \quad = \quad T_s + \widetilde{T}^{(s)}_u \qquad \text{for } s, u \geq 0, \tag{10.1.9}$$

where $\widetilde{T}^{(s)}_u = \inf\left\{t \geq 0 \ : \ \widetilde{B}^{(s)}_t = u\right\}$ is the first passage time to u for the process $\widetilde{B}^{(s)}$.



By (10.1.9), the increment $T_{s+u} - T_s$ is independent of $\mathcal{F}_{T_s}$, and, by the reflection principle,

$$T_{s+u} - T_s \quad \sim \quad T_u \quad \sim \quad \frac{u}{\sqrt{2\pi}}\, x^{-3/2} \exp\left(-\frac{u^2}{2x}\right) I_{(0,\infty)}(x)\, dx.$$

Hence (T_s) is an increasing process with stationary and independent increments. The process (T_s) is left-continuous, but it is not difficult to verify that

$$T_{s+} \ = \ \lim_{\varepsilon\downarrow 0} T_{s+\varepsilon} \ = \ \inf\left\{t \geq 0 \ : \ \widetilde{B}^{(s)}_t > u\right\}$$

is a càdlàg modification, and hence a Lévy process. (T_{s+}) is called **"The Lévy subordinator"**, where "subordinator" stands for an increasing Lévy process. We will see below that subordinators are used for random time transformations ("subordination") of other Lévy processes.

More generally, if $X_t = \sigma B_t + bt$ is a Gaussian Lévy process with coefficients $\sigma > 0$, $b \in \mathbb{R}$, then the right inverse

$$T_s^X \quad = \quad \inf\{t \geq 0 \,:\, X_t = s\} \quad , \quad s \geq 0,$$

is called an **Inverse Gaussian subordinator**.

Exercise (Sample paths of Inverse Gaussian processes). Prove that the process $(T_s)_{s\geq 0}$ is increasing and *purely discontinuous*, i.e., with probability one, (T_s) is not continuous on any non-empty open interval $(a, b) \subset [0, \infty)$.

Example (Stable processes). Stable processes are Lévy processes that appear as scaling limits of Random Walks. Suppose that $S_n = \sum_{j=1}^n \eta_j$ is a Random Walk in $\mathbb{R}^d$ with i.i.d. increments η_j. If the random variables η_j are square-integrable with mean zero then Donsker's invariance principle (the "*functional central limit theorem*") states that the diffusively rescaled process $(k^{-1/2} S_{\lfloor kt \rfloor})_{t\geq 0}$ converges in distribution to $(\sigma B_t)_{t\geq 0}$ where (B_t) is a Brownian motion in $\mathbb{R}^d$ and σ is a non-negative definite symmetric $d\times d$ matrix. However, the functional central limit theorem does not apply if the increments η_j are not square integrable ("**heavy tails**"). In this case, one considers limits of rescaled Random Walks of the form $X_t^{(k)} = k^{-1/\alpha} S_{\lfloor kt \rfloor}$ where $\alpha \in (0, 2]$ is a fixed constant. It is not difficult to verify that if $(X_t^{(k)})$ converges in distribution to a limit process (X_t) then (X_t) is a Lévy process that is invariant under the rescaling, i.e.,

$$k^{-1/\alpha} X_{kt} \quad \sim \quad X_t \qquad \text{for any } k \in (0,\infty) \text{ and } t \geq 0. \tag{10.1.10}$$

Definition. *Let $\alpha \in (0, 2]$. A Lévy process (X_t) satisfying (10.1.10) is called **(strictly) α-stable**.*

The reason for the restriction to $\alpha \in (0,2]$ is that for $\alpha > 2$, an α-stable process does not exist. This will become clear by the proof of Theorem 10.4 below. There is a broader class of Lévy processes that is called α-stable in the literature, cf. e.g. [28]. Throughout these notes, by an **α-stable process** we always mean a strictly α-stable process as defined above.

For $b \in \mathbb{R}$, the deterministic process $X_t = bt$ is a 1-stable Lévy process. Moreover, a Lévy process X in $\mathbb{R}^1$ is 2-stable if and only if $X_t = \sigma B_t$ for a Brownian motion (B_t) and a constant $\sigma \in [0,\infty)$. Characteristic exponents can be applied to classify all α-stable processes:

Theorem 10.4 (Characterization of stable processes). *For $\alpha \in (0,2]$ and a Lévy process (X_t) in $\mathbb{R}^1$ with $X_0 = 0$ the following statements are equivalent:*

(i) *(X_t) is strictly α-stable.*

(ii) *$\psi(cp) = c^\alpha \psi(p)$ for any $c \geq 0$ and $p \in \mathbb{R}$.*

(iii) *There exists constants $\sigma \geq 0$ and $\mu \in \mathbb{R}$ such that*

$$\psi(p) \quad = \quad \sigma^\alpha |p|^\alpha (1 + i\mu\,\mathrm{sgn}(p)).$$

Proof. $(i) \Leftrightarrow (ii)$. The process (X_t) is strictly α-stable if and only if $X_{c^\alpha t} \sim cX_t$ for any $c, t \geq 0$, i.e., if and only if

$$e^{-t\psi(cp)} \quad = \quad E\Big[e^{ipcX_t}\Big] = E\Big[e^{ipX_{c^\alpha t}}\Big] \quad = \quad e^{-c^\alpha t\psi(p)}$$

for any $c, t \geq 0$ and $p \in \mathbb{R}$.

$(ii) \Leftrightarrow (iii)$. Clearly, Condition (ii) holds if and only if there exist complex numbers z_+ and z_- such that

$$\psi(p) \quad = \quad \begin{cases} z_+|p|^\alpha & \text{for } p \geq 0, \\ z_-|p|^\alpha & \text{for } p \leq 0. \end{cases}$$

Moreover, since $\phi_t(p) = \exp(-t\psi(p))$ is a characteristic function of a probability measure for any $t \geq 0$, the characteristic exponent ψ satisfies $\psi(-p) = \overline{\psi(p)}$ and $\Re(\psi(p)) \geq 0$. Therefore, $z_- = \overline{z}_+$ and $\Re(z_+) \geq 0$. □

Example (Symmetric α-stable processes). A Lévy process in $\mathbb{R}^d$ with characteristic exponent

$$\psi(p) \quad = \quad \sigma^\alpha |p|^\alpha$$

for some $\sigma \geq 0$ and $a \in (0, 2]$ is called a **symmetric α-stable process**. We will see below that a symmetric α-stable process is a Markov process with generator $-\sigma^\alpha(-\Delta)^{\alpha/2}$. In particular, Brownian motion is a symmetric 2-stable process.

10.2 Martingales and Markov property

For Lévy processes, one can identify similar fundamental martingales as for Brownian motion. Furthermore, every Lévy process is a strong Markov process.

Martingales of Lévy processes

The notion of a martingale immediately extends to complex or vector valued processes by a componentwise interpretation. As a consequence of Theorem 10.1 we obtain:

Corollary 10.5. *If (X_t) is a Lévy process with $X_0 = 0$ and characteristic exponent ψ, then the following processes are martingales:*

(i) $\exp(ip \cdot X_t + t\psi(p))$ *for any $p \in \mathbb{R}^d$,*

(ii) $M_t = X_t - bt$ *with $b = i\nabla\psi(0)$, provided* $X_t \in \mathcal{L}^1$ $\forall t \geq 0$.

(iii) $M_t^j M_t^k - a^{jk} t$ *with* $a^{jk} = \frac{\partial^2\psi}{\partial p_j \partial p_k}(0)$ $(j, k = 1, \ldots, d)$, *provided* $X_t \in \mathcal{L}^2$ $\forall\, t \geq 0$.

Proof. We only prove (ii) and (iii) for $d = 1$ and leave the remaining assertions as an exercise to the reader. If $d = 1$ and (X_t) is integrable then for $0 \leq s \leq t$,

$$E[X_t - X_s \mid \mathcal{F}_s] \quad = \quad E[X_t - X_s] \quad = \quad i(t-s)\psi'(0)$$

by independence and stationarity of the increments and by (10.1.2). Hence $M_t = X_t - it\psi'(0)$ is a martingale. Furthermore,

$$M_t^2 - M_s^2 = (M_t + M_s)\,(M_t - M_s) = 2M_s(M_t - M_s) + (M_t - M_s)^2.$$

If (X_t) is square integrable then the same holds for (M_t), and we obtain

$$\begin{aligned} E[M_t^2 - M_s^2 \mid \mathcal{F}_s] &= E[(M_t - M_s)^2 \mid \mathcal{F}_s] = \mathrm{Var}[M_t - M_s \mid \mathcal{F}_s] \\ &= \mathrm{Var}[X_t - X_s \mid \mathcal{F}_s] = \mathrm{Var}[X_t - X_s] = \mathrm{Var}[X_{t-s}] = (t-s)\psi''(0) \end{aligned}$$

Hence $M_t^2 - t\psi''(0)$ is a martingale. □

Note that Corollary 10.5 (ii) shows that an integrable Lévy process is a *semimartingale* with martingale part M_t and continuous finite variation part bt. The identity (10.1.1) can be used to classify all Lévy processes, c.f. e.g. [5]. In particular, we will prove below that by Corollary 10.5, any continuous Lévy process with $X_0 = 0$ is of the type $X_t = \sigma B_t + bt$ with a d-dimensional Brownian motion (B_t) and constants $\sigma \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$.

Lévy processes as Markov processes

The independence and stationarity of the increments of a Lévy process immediately implies the Markov property:

Theorem 10.6 (Markov property). *A Lévy process (X_t, P) is a time-homogeneous Markov process with translation invariant transition functions*

$$p_t(x, B) \;=\; \mu_t(B - x) \;=\; p_t(a + x, a + B) \quad \forall\, a \in \mathbb{R}^d, \tag{10.2.1}$$

where $\mu_t = P \circ (X_t - X_0)^{-1}$.

Proof. For any $s, t \geq 0$ and $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} P[X_{s+t} \in B \mid \mathcal{F}_s](\omega) &= P[X_s + (X_{s+t} - X_s) \in B \mid \mathcal{F}_s](\omega) \\ &= P[X_{s+t} - X_s \in B - X_s(\omega)] \\ &= P[X_t - X_0 \in B - X_s(\omega)] \\ &= \mu_t(B - X_s(\omega)). \end{aligned}$$

□

Remark (Feller property). The transition semigroup of a Lévy process has the **Feller property**, i.e., if $f : \mathbb{R}^d \to \mathbb{R}$ is a continuous function vanishing at infinity then the same holds for $p_t f$ for any $t \geq 0$. Indeed,

$$(p_t f)(x) = \int f(x+y)\, \mu_t(dy)$$

is continuous by dominated convergence, and, similarly, $(p_t f)(x) \to 0$ as $|x| \to \infty$.

Exercise (Strong Markov property for Lévy processes). Let (X_t) be an $(\mathcal{F}_t)$ Lévy process, and let T be a finite stopping time. Show that $Y_t = X_{T+t} - X_T$ is a process that is independent of $\mathcal{F}_T$, and X and Y have the same law.

Hint: Consider the sequence of stopping times defined by $T_n = (k+1)2^{-n}$ if $k2^{-n} \leq T < (k+1)2^{-n}$. Notice that $T_n \downarrow T$ as $n \to \infty$. In a first step show that for any $m \in \mathbb{N}$ and $t_1 < t_2 < \ldots < t_m$, any bounded continuous function f on $\mathbb{R}^m$, and any $A \in \mathcal{F}_T$ we have

$$E\left[f(X_{T_n+t_1} - X_{T_n}, \ldots, X_{T_n+t_m} - X_{T_n}) I_A\right] = E\left[f(X_{t_1}, \ldots, X_{t_m})\right] \, P[A].$$

Exercise (A characterization of Poisson processes). Let $(X_t)_{t \geq 0}$ be a Lévy process with $X_0 = 0$ a.s. Suppose that the paths of X are piecewise constant, increasing, all jumps of X are of size 1, and X is not identically 0. Prove that X is a Poisson process.

Hint: Apply the Strong Markov property to the jump times $(T_i)_{i=1,2,\ldots}$ of X to conclude that the random variables $U_i := T_i - T_{i-1}$ are i.i.d. (with $T_0 := 0$). Then, it remains to show that U_1 is an exponential random variable with some parameter $\lambda > 0$.

The marginals of a Lévy process $((X_t)_{t\geq 0}, P)$ are completely determined by the characteristic exponent ψ. In particular, one can obtain the transition semigroup and its infinitesimal generator from ψ by Fourier inversion. Let $\mathcal{S}(\mathbb{R}^d)$ denote the Schwartz space consisting of all functions $f \in C^\infty(\mathbb{R}^d)$ such that $|x|^k \partial^\alpha f(x)$ goes to 0 as $|x| \to \infty$ for any $k \in \mathbb{N}$ and derivatives of f of arbitary order $\alpha \in \mathbb{Z}_+^d$. Recall that the Fourier transform maps $\mathcal{S}(\mathbb{R}^d)$ one-to-one onto $\mathcal{S}(\mathbb{R}^d)$.

Corollary 10.7 (Transition semigroup and generator of a Lévy process).

(1). For any $f \in \mathcal{S}(\mathbb{R}^d)$ and $t \geq 0$,

$$p_t f \;=\; (e^{-t\psi}\hat{f})\check{}$$

*where $\hat{f}(p) = (2\pi)^{-d/2}\int e^{-ip\cdot x} f(x)\,dx$ and $\check{g}(x) = (2\pi)^{-d/2}\int e^{ip\cdot x} g(p)\,dp$ denote the **Fourier transform** and the **inverse Fourier transform** of functions $f, g \in \mathcal{L}^1(\mathbb{R}^d)$.*

(2). The Schwartz space $\mathcal{S}(\mathbb{R}^d)$ is contained in the domain of the generator L of the Feller semigroup induced by $(p_t)_{t\geq 0}$ on the Banach space $\hat{C}(\mathbb{R}^d)$ of continuous functions vanishing at infinity, and the generator is the pseudo-differential operator given by

$$Lf \;=\; (-\psi\hat{f})\check{}\,. \tag{10.2.2}$$

Proof. (1). Since $(p_t f)(x) = E[f(X_t + x)]$, we conclude by Fubini that

$$\begin{aligned}
(\widehat{p_t f})(p) &= (2\pi)^{-\frac{d}{2}} \int e^{-ip\cdot x}(p_t f)(x)\,dx \\
&= (2\pi)^{-\frac{d}{2}} \cdot E\left[\int e^{-ip\cdot x} f(X_t + x)\,dx\right] \\
&= E\left[e^{ip\cdot X_t}\right] \cdot \hat{f}(p) \\
&= e^{-t\psi(p)}\hat{f}(p)
\end{aligned}$$

for any $p \in \mathbb{R}^d$. The claim follows by the Fourier inversion theorem, noting that $\left|e^{-t\psi}\right| \leq 1$.

(2). For $f \in \mathcal{S}(\mathbb{R}^d)$, $\hat{f}$ is in $\mathcal{S}(\mathbb{R}^d)$ as well. The Lévy-Khinchin formula that we will state below gives an explicit representation of all possible Lévy exponents which shows in particular that $\psi(p)$ is growing at most polynomially as $|p| \to \infty$. Since

$$\left|\frac{e^{-t\psi}\hat{f} - \hat{f}}{t} + \psi\hat{f}\right| = \left|\frac{e^{-t\psi} - 1}{t} + \psi\right| \cdot |\hat{f}|, \qquad \text{and}$$

$$\frac{e^{-t\psi} - 1}{t} + \psi = -\frac{1}{t}\int_0^t \psi\left(e^{-s\psi} - 1\right)\, ds = \frac{1}{t}\int_0^t\int_0^s \psi^2 e^{-r\psi}\, dr\, ds,$$

we obtain

$$\left|\frac{e^{-t\psi}\hat{f} - \hat{f}}{t} + \psi\hat{f}\right| \leq t \cdot |\psi^2| \cdot |\hat{f}| \ \in \mathcal{L}^1(\mathbb{R}^d),$$

and, therefore,

$$\begin{aligned} &\frac{(p_t f)(x) - f(x)}{t} - (-\psi\hat{f})\check{}(x) \\ = \ & (2\pi)^{-\frac{d}{2}} \int e^{ip\cdot x} \left(\frac{1}{t}\left(e^{-t\psi(p)}\hat{f}(p) - \hat{f}(p)\right) \ + \ \psi(p)\hat{f}(p)\right) dp \quad \longrightarrow \quad 0 \end{aligned}$$

as $t \downarrow 0$ uniformly in x. This shows $f \in \mathrm{Dom}(L)$ and $Lf = (-\psi\hat{f})\check{}$.

□

By the theory of Markov processes, the corollary shows in particular that a Lévy process (X_t, P) solves the martingale problem for the operator $(L, \mathcal{S}(\mathbb{R}^d))$ defined by (14.1.10).

Examples. 1) For a Gaussian Lévy processes as considered above, $\psi(p) = \frac{1}{2}p\cdot ap - ib\cdot p$ where $a := \sigma\sigma^T$. Hence the generator is given by

$$Lf \ = \ -(\psi\hat{f})\check{} \ = \ \frac{1}{2}\nabla\cdot(a\nabla f) - b\cdot\nabla f, \qquad \text{for } f \in \mathcal{S}(\mathbb{R}^n).$$

2) For a Poisson process (N_t), $\psi(p) = \lambda(1 - e^{ip})$ implies

$$(Lf)(x) \ = \ \lambda(f(x+1) - f(x)).$$

3) For the compensated Poisson process $M_t = N_t - \lambda t$,

$$(Lf)(x) \ = \ \lambda(f(x+1) - f(x) - f'(x)).$$

4) For a symmetric α-stable process with characteristic exponent $\psi(p) = \sigma^\alpha \cdot |p|^\alpha$ for some $\sigma \geq 0$ and $\alpha \in (0, 2]$, the generator is a fractional power of the Laplacian:

$$Lf = -(\psi \hat{f})\check{} = -\sigma^\alpha (-\Delta)^{\alpha/2} f .$$

We remark that for $\alpha > 2$, the operator L does not satisfy the positive maximum principle. Therefore, in this case L does not generate a transition semigroup of a Markov process.

10.3 Poisson random measures and Poisson point processes

A compensated Poisson process has only finitely many jumps in a finite time interval. General Lévy jump processes may have a countably infinite number of (small) jumps in finite time. In the next section, we will construct such processes from their jumps. As a preparation we will now study Poisson random measures and Poisson point processes that encode the jumps of Lévy processes. The jump part of a Lévy process can be recovered from these counting measure valued processes by integration, i.e., summation of the jump sizes. We start with the observation that the jump times of a Poisson process form a Poisson random measure on $\mathbb{R}_+$.

The jump times of a Poisson process

For a different point of view on Poisson processes let

$$\mathcal{M}_c^+(S) = \left\{ \sum \delta_{y_i} \ : \ (y_i) \text{ finite or countable sequence in } S \right\}$$

denote the set of all counting measures on a set S. A Poisson process $(N_t)_{t\geq 0}$ can be viewed as the distribution function of a random counting measure, i.e., of a random variable $N : \Omega \to \mathcal{M}_c^+([0, \infty))$.

Definition. *Let ν be a σ-finite measure on a measurable space $(S, \mathcal{S})$. A collection of random variables $N(B)$, $B \in \mathcal{S}$, on a probability space $(\Omega, \mathcal{A}, P)$ is called a* ***Poisson random measure (or spatial Poisson process) of intensity ν*** *if and only if*

(i) $B \mapsto N(B)(\omega)$ *is a counting measure for any* $\omega \in \Omega$,

(ii) if $B_1, \ldots, B_n \in \mathcal{S}$ *are disjoint then the random variables* $N(B_1), \ldots, N(B_n)$ *are independent,*

(iii) $N(B)$ *is Poisson distributed with parameter* $\nu(B)$ *for any* $B \in \mathcal{S}$ *with* $\nu(B) < \infty$.

A Poisson random measure N with finite intensity ν can be constructed as the empirical measure of a Poisson distributed number of independent samples from the normalized measure $\nu/\nu(S)$:.

$$N = \sum_{j=1}^{K} \delta_{X_j} \quad \text{with } X_j \sim \nu/\nu(s) \text{ i.i.d.,} \quad K \sim \text{Poisson}(\nu(S)) \text{ independent.}$$

If the intensity measure ν does not have atoms then almost surely, $N(\{x\}) \in \{0, 1\}$ for any $x \in S$, and $N = \sum_{x \in A} \delta_x$ for a random subset A of S. For this reason, a Poisson random measure is often called a Poisson point process, but we will use this terminology differently below.

A real-valued process $(N_t)_{t \geq 0}$ is a Poisson process of intensity $\lambda > 0$ if and only if $t \mapsto N_t(\omega)$ is the distribution function of a Poisson random measure $N(dt)(\omega)$ on $\mathcal{B}([0, \infty))$ with intensity measure $\nu(dt) = \lambda\, dt$. The Poisson random measure $N(dt)$ can be interpreted as the derivative of the Poisson process:

$$N(dt) = \sum_{s:\ \Delta N_s \neq 0} \delta_s(dt).$$

In a stochastic differential equation of type $dY_t = \sigma(Y_{t-})\, dN_t$, $N(dt)$ is the driving **Poisson noise**.

The following assertion about Poisson processes is intuitively clear from the interpretation of a Poisson process as the distribution function of a Poisson random measure.

Compound Poisson processes enable us to give a simple proof of the second part of the theorem:

Theorem 10.8 (**Superpositions and subdivisions of Poisson processes**). *Let K be a countable set.*

1) *Suppose that $(N_t^{(k)})_{t\geq 0}$, $k\in K$, are independent Poisson processes with intensities λ_k. Then*

$$N_t \quad = \quad \sum_{k\in K} N_t^{(k)} \qquad , \quad t\geq 0,$$

is a Poisson process with intensity $\lambda = \sum \lambda_k$ provided $\lambda<\infty$.

2) *Conversely, if $(N_t)_{t\geq 0}$ is a Poisson process with intensity $\lambda > 0$, and $(C_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random variables $C_n : \Omega \mapsto K$ that is also independent of (N_t), then the processes*

$$N_t^{(k)} \quad = \quad \sum_{j=1}^{N_t} I_{\{C_j = k\}} \quad , \quad t\geq 0,$$

are independent Poisson processes of intensities $q_k\lambda$, where $q_k = \mathcal{P}[C_1 = k]$.

The subdivision in the second assertion can be thought of as colouring the points in the support of the corresponding Poisson random measure $N(dt)$ independently with random colours C_j, and decomposing the measure into parts $N^{(k)}(dt)$ of equal colour.

Proof. The first part is rather straightforward, and left as an exercise. For the second part, we may assume w.l.o.g. that K is finite. Then the process $\vec{N}_t : \Omega \to \mathbb{R}^K$ defined by

$$\vec{N}_t \quad := \quad \left(N_t^{(k)}\right)_{k\in K} = \sum_{j=1}^{N_t} \eta_j \quad \text{with} \quad \eta_j = \left(I_{\{k\}}(C_j)\right)_{k\in K}$$

is a compound Poisson process on $\mathbb{R}^K$, and hence a Lévy process. Moreover, by the proof of Lemma 10.2, the characteristic function of $\vec{N}_t$ for $t\geq 0$ is given by

$$E\left[\exp\left(ip\cdot\vec{N}_t\right)\right] = \exp\left(\lambda t(\phi(p)-1)\right), \quad p\in\mathbb{R}^K,$$

where

$$\phi(p) = E\left[\exp(ip \cdot \eta_1)\right] = E\left[\exp\left(i\sum_{k\in K} p_k I_{\{k\}}(C_1)\right)\right] = \sum_{k\in K} q_k e^{ip_k}.$$

Noting that $\sum q_k = 1$, we obtain

$$E[\exp(ip \cdot \vec{N}_t)] = \prod_{k\in K} \exp(\lambda t q_k(e^{ip_k} - 1)) \qquad \text{for any } p \in \mathbb{R}^K \text{ and } t \geq 0.$$

The assertion follows, because the right hand side is the characteristic function of a Lévy process in $\mathbb{R}^K$ whose components are independent Poisson processes with intensities $q_k\lambda$. □

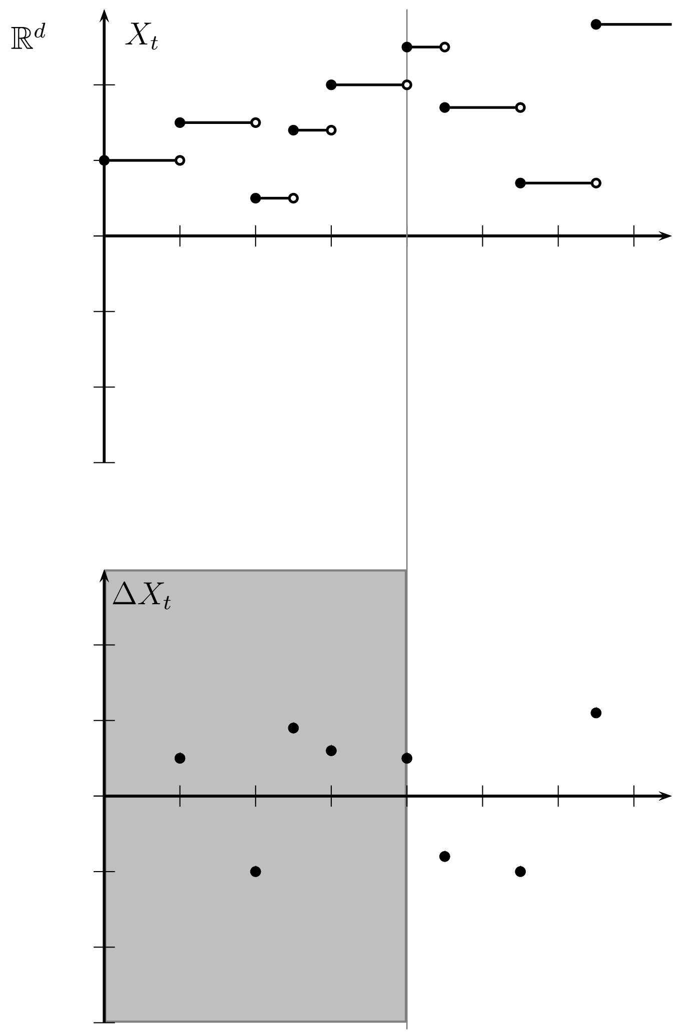
The jumps of a Lévy process

We now turn to general Lévy processes. Note first that a Lévy process (X_t) has only countably many jumps, because the paths are càdlàg. The jumps can be encoded in the counting measure-valued stochastic process $N_t : \Omega \to \mathcal{M}_c^+(\mathbb{R}^d \setminus \{0\})$,

$$N_t(dy) \quad = \quad \sum_{\substack{s\leq t\\ \Delta X_s\neq 0}} \delta_{\Delta X_s}(dy), \quad t \geq 0,$$

or, equivalently, in the random counting measure $N : \Omega \to \mathcal{M}_c^+\big(\mathbb{R}_+ \times (\mathbb{R}^d \setminus \{0\})\big)$ defined by

$$N(dt\, dy) \quad = \quad \sum_{\substack{s\leq t\\ \Delta X_s\neq 0}} \delta_{(s,\Delta X_s)}(dt\, dy).$$



The process $(N_t)_{t\geq 0}$ is increasing and adds a Dirac mass at y each time the Lévy process has a jump of size y. Since (X_t) is a Lévy process, (N_t) also has stationary and independent increments:

$$N_{s+t}(B) - N_s(B) \sim N_t(B) \qquad \text{for any } s, t \geq 0 \text{ and } B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}).$$

Hence for any set B with $N_t(B) < \infty$ a.s. for all t, the integer valued stochastic process $(N_t(B))_{t\geq 0}$ is a Lévy process with jumps of size $+1$. By an exercise in Section 10.1, we can conclude that $(N_t(B))$ is a Poisson process. In particular, $t \mapsto \mathbb{E}[N_t(B)]$ is a linear function.

Definition. *The* ***jump intensity measure*** *of a Lévy process (X_t) is the σ-finite measure ν on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$ determined by*

$$\mathbb{E}[N_t(B)] \;=\; t \cdot \nu(B) \qquad \forall\, t \geq 0,\; B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}). \tag{10.3.1}$$

It is elementary to verify that for any Lévy process, there is a unique measure ν satisfying (10.3.1). Moreover, since the paths of a Lévy process are càdlàg, the measures N_t and ν are finite on $\{y \in \mathbb{R}^d : |y| \geq \varepsilon\}$ for any $\varepsilon > 0$.

Example (Jump intensity of stable processes). The jump intensity measure of strictly α-stable processes in $\mathbb{R}^1$ can be easily found by an informal argument. Suppose we rescale in space and time by $y \to cy$ and $t \to c^\alpha t$. If the jump intensity is $\nu(dy) = f(y)\,dy$, then after rescaling we would expect the jump intensity $c^\alpha f(cy)c\,dy$. If scale invariance holds then both measures should agree, i.e., $f(y) \propto |y|^{-1-\alpha}$ both for $y > 0$ and for $y < 0$ respectively. Therefore, the jump intensity measure of a strictly α-stable process on $\mathbb{R}^1$ should be given by

$$\nu(dy) \quad = \quad \big(c_+ I_{(0,\infty)}(y) + c_- I_{(-\infty,0)}(y)\big)\, |y|^{-1-\alpha}\, dy \tag{10.3.2}$$

with constants $c_+, c_- \in [0, \infty)$.

If (X_t) is a pure jump process with finite jump intensity measure (i.e., finitely many jumps in a finite time interval) then it can be recovered from (N_t) by adding up the jump sizes:

$$X_t - X_0 \quad = \quad \sum_{s \leq t} \Delta X_s \quad = \quad \int y\; N_t(dy).$$

In the next section, we are conversely going to construct more general Lévy jump processes from the measure-valued processes encoding the jumps. As a first step, we are going to define formally the counting-measure valued processes that we are interested in.

Poisson point processes

Let $(S, \mathcal{S}, \nu)$ be a σ-finite measure space.

Definition. *A collection $N_t(B)$, $t \geq 0$, $B \in \mathcal{S}$, of random variables on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is called a* ***Poisson point process of intensity ν*** *if and only if*

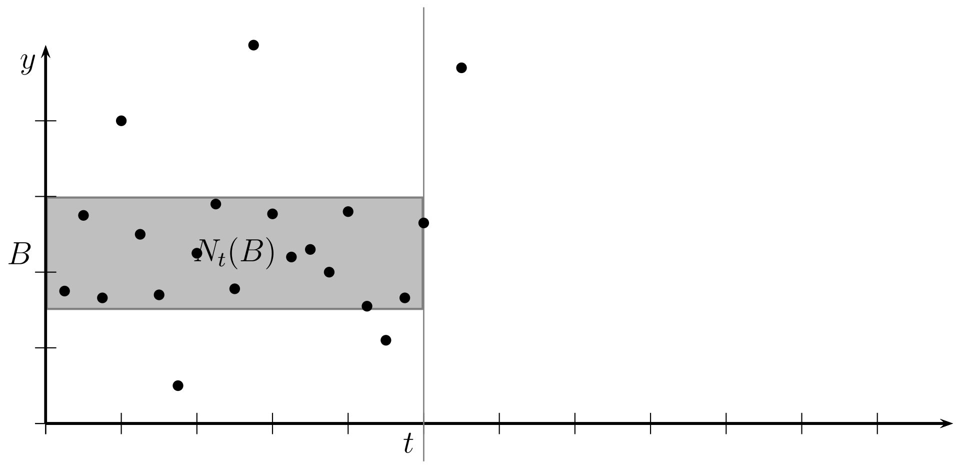
(i) $B \mapsto N_t(B)(\omega)$ *is a counting measure on S for any $t \geq 0$ and $\omega \in \Omega$,*

(ii) if $B_1, \ldots, B_n \in \mathcal{S}$ are disjoint then $(N_t(B_1))_{t\geq 0}, \ldots, (N_t(B_n))_{t\geq 0}$ are independent stochastic processes and

(iii) $(N_t(B))_{t\geq 0}$ *is a Poisson process of intensity $\nu(B)$ for any $B \in \mathcal{S}$ with $\nu(B) < \infty$.*

A Poisson point process adds random points with intensity $\nu(dt)\ dy$ in each time instant dt. It is the distribution function of a Poisson random measure $N(dt\ dy)$ on $\mathbb{R}^+ \times \mathcal{S}$ with intensity measure $dt\ \nu(dy)$, i.e.

$$N_t(B) \quad = \quad N((0,t] \times B) \qquad \text{for any } t \geq 0 \text{ and } B \in \mathcal{S}.$$



The distribution of a Poisson point process is uniquely determined by its intensity measure: If (N_t) and $(\widetilde{N_t})$ are Poisson point processes with intensity ν then

$$(N_t(B_1), \ldots, N_t(B_n))_{t\geq 0} \quad \sim \quad (\widetilde{N_t}(B_1), \ldots, (\widetilde{N_t}(B_n))_{t\geq 0}$$

for any finite collection of disjoint sets $B_1, \ldots, B_n \in \mathcal{S}$, and, hence, for any finite collection of measurable arbitrary sets $B_1, \ldots, B_n \in \mathcal{S}$.

Applying a measurable map to the points of a Poisson point process yields a new Poisson point process:

Exercise (Mapping theorem). Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be measurable spaces and let $f : S \to T$ be a measurable function. Prove that if (N_t) is a Poisson point process with intensity measure ν then the image measures $N_t \circ f^{-1}$, $t \geq 0$, form a Poisson point process on T with intensity measure $\nu \circ f^{-1}$.

An advantage of Poisson point processes over Lévy processes is that the passage from finite to infinite intensity (of points or jumps respectively) is not a problem on the level of Poisson point processes because the resulting sums trivially exist by positivity:

Theorem 10.9 (Construction of Poisson point processes).

1) *Suppose that ν is a finite measure with total mass $\lambda = \nu(S)$. Then*

$$N_t \quad = \quad \sum_{j=1}^{K_t} \delta_{\eta_j}$$

is a Poisson point process of intensity ν provided the random variables η_j are independent with distribution $\lambda^{-1}\nu$, and (K_t) is an independent Poisson process of intensity λ.

2) *If $(N_t^{(k)})$, $k \in \mathbb{N}$, are independent Poisson point processes on $(S, \mathcal{S})$ with intensity measures ν_k then*

$$\overline{N_t} \quad = \quad \sum_{k=1}^{\infty} N_t^{(k)}$$

is a Poisson point process with intensity measure $\nu = \sum \nu_k$.

The statements of the theorem are consequences of the subdivision and superposition properties of Poisson processes. The proof is left as an exercise.

Conversely, one can show that any Poisson point process with finite intensity measure ν can be almost surely represented as in the first part of Theorem 10.9, where $K_t = N_t(S)$. The proof uses uniqueness in law of the Poisson point process, and is similar to the proof of Lemma 10.3.

Construction of compound Poisson processes from PPP

We are going to construct Lévy jump processes from Poisson point processes. Suppose first that (N_t) is a Poisson point process on $\mathbb{R}^d \setminus \{0\}$ with **finite** intensity measure ν. Then the support of N_t is almost surely finite for any $t \geq 0$. Therefore, we can define

$$X_t \quad = \quad \int_{\mathbb{R}^d \setminus \{0\}} y\, N_t(dy) \quad = \quad \sum_{y \in \text{supp}(N_t)} y\, N_t(\{y\}),$$

Theorem 10.10. *If $\nu(\mathbb{R}^d \setminus \{0\}) < \infty$ then $(X_t)_{t \geq 0}$ is a compound Poisson process with jump intensity ν. More generally, for any Poisson point process with finite intensity measure ν on a measurable space $(S, \mathcal{S})$ and for any measurable function $f : S \to \mathbb{R}^n$, $n \in \mathbb{N}$, the process*

$$N_t(f) \quad := \quad \int f(y) N_t(dy) \quad , \quad t \geq 0,$$

is a compound Poisson process with intensity measure $\nu \circ f^{-1}$.

Proof. By Theorem 10.9 and by the uniqueness in law of a Poisson point process with given intensity measure, we can represent (N_t) almost surely as $N_t = \sum_{j=1}^{K_t} \delta_{\eta_j}$ with i.i.d. random variables $\eta_j \sim \nu/\nu(S)$ and an independent Poisson process (K_t) of intensity $\nu(S)$. Thus,

$$N_t(f) \quad = \quad \int f(y) N_t(dy) \quad = \quad \sum_{j=1}^{K_t} f(\eta_j) \quad \text{almost surely.}$$

Since the random variables $f(\eta_j)$, $j \in \mathbb{N}$, are i.i.d. and independent of (K_t) with distribution $\nu \circ f^{-1}$, $(N_t(f))$ is a compound Poisson process with this intensity measure. □

As a direct consequence of the theorem and the properties of compound Poisson processes derived above, we obtain:

Corollary 10.11 (Martingales of Poisson point processes). *Suppose that (N_t) is a Poisson point process with finite intensity measure ν. Then the following processes are martingales w.r.t. the filtration $\mathcal{F}_t^N = \sigma(N_s(B) \mid 0 \leq s \leq t,\ B \in \mathcal{S})$:*

(i) $\widetilde{N_t}(f) \ = \ N_t(f) - t \int f d\nu$ *for any* $f \in \mathcal{L}^1(\nu)$,

(ii) $\widetilde{N_t}(f)\widetilde{N_t}(g) - t \int fg\, d\nu$ *for any* $f, g \in \mathcal{L}^2(\nu)$,

(iii) $\exp\left(ipN_t(f) + t \int (1 - e^{ipf})\, d\nu\right)$ *for any measurable* $f : S \to \mathbb{R}$ *and* $p \in \mathbb{R}$.

Proof. If f is in $\mathcal{L}^p(\nu)$ for $p = 1, 2$ respectively, then

$$\int |x|^p\ \nu \circ f^{-1}(dx) = \int |f(y)|^p\ \nu(dy) < \infty,$$

$$\int x\ \nu \circ f^{-1}(dx) \quad = \int f\ d\nu, \quad \text{and} \quad \int xy\ \nu \circ (fg)^{-1}(dxdy) = \int fg\ d\nu \quad .$$

Therefore (i) and (ii) (and similarly also (iii)) follow from the corresponding statements for compound Poisson processes. □

With a different proof and an additional integrability assumption, the assertion of Corollary 10.11 extends to σ-finite intensity measures:

Exercise (Expectation values and martingales for Poisson point processes with infinite intensity). Let (N_t) be a Poisson point process with σ-finite intensity ν.

a) By considering first elementary functions, prove that for $t \geq 0$, the identity

$$E\left[\int f(y) N_t(dy)\right] \quad = \quad t \int f(y) \nu(dy)$$

holds for any measurable function $f : S \to [0, \infty]$. Conclude that for $f \in \mathcal{L}^1(\nu)$, the integral $N_t(f) = \int f(y) N_t(dy)$ exists almost surely and defines a random variable in $L^1(\Omega, \mathcal{A}, \mathcal{P})$.

b) Proceeding similarly as in a), prove the identities

$$
\begin{aligned}
E[N_t(f)] &= t\int f\,d\nu && \text{for any } f\in\mathcal{L}^1(\nu),\\
\mathrm{Cov}[N_t(f),N_t(g)] &= t\int fg\,d\nu && \text{for any } f,g\in\mathcal{L}^1(\nu)\cap\mathcal{L}^2(\nu),\\
E[\exp(ipN_t(f))] &= \exp(t\int(e^{ipf}-1)\,d\nu) && \text{for any } f\in\mathcal{L}^1(\nu).
\end{aligned}
$$

c) Show that the processes considered in Corollary 10.11 are again martingales provided $f\in\mathcal{L}^1(\nu)$, $f,g\in\mathcal{L}^1(\nu)\cap\mathcal{L}^2(\nu)$ respectively.

If (N_t) is a Poisson point process with intensity measure ν then the signed measure valued stochastic process

$$\widetilde{N}_t(dy) \quad := \quad N_t(dy)-t\,\nu(dy) \quad , \quad t\geq 0,$$

is called a **compensated Poisson point process** . Note that by Corollary 10.11 and the exercise,

$$\widetilde{N_t}(f) \quad = \quad \int f(y)\widetilde{N_t}(dy)$$

is a martingale for any $f\in\mathcal{L}^1(\nu)$, i.e., $(\widetilde{N_t})$ is a **measure-valued martingale**.

10.4 Stochastic integrals w.r.t. Poisson point processes

Let $(S,\mathcal{S},\nu)$ be a σ-finite measure space, and let $(\mathcal{F}_t)$ be a filtration on a probability space $(\Omega,\mathcal{A},P)$. Our main interest is the case $S=\mathbb{R}^d$. Suppose that $(N_t(dy))_{t\geq 0}$ is an $(\mathcal{F}_t)$ Poisson point process on $(S,\mathcal{S})$ with intensity measure ν. As usual, we denote by $\widetilde{N}_t=N_t-t\nu$ the compensated Poisson point process, and by $N(dt\,dy)$ and $\widetilde{N}(dt\,dy)$ the corresponding uncompensated and compensated Poisson random measure on $\mathbb{R}_+\times S$. Recall that for $A,B\in\mathcal{S}$ with $\nu(A)<\infty$ and $\nu(B)<\infty$, the processes $\widetilde{N}_t(A)$, $\widetilde{N}_t(B)$, and $\widetilde{N}_t(A)\widetilde{N}_t(B)-t\nu(A\cap B)$ are martingales. Our goal is to define stochastic integrals of type

$$(G_\bullet N)_t \quad = \quad \int_{(0,t]\times S} G_s(y)\,N(ds\,dy), \tag{10.4.1}$$

$$(G_\bullet \widetilde{N})_t \quad = \quad \int_{(0,t]\times S} G_s(y)\,\widetilde{N}(ds\,dy) \tag{10.4.2}$$

respectively for predictable processes $(\omega, s, y) \mapsto G_s(y)(\omega)$ defined on $\Omega \times \mathbb{R}_+ \times S$. In particular, choosing $G_s(y)(\omega) = y$, we will obtain Lévy processes with possibly infinite jump intensity from Poisson point processes. If the measure ν is finite and has no atoms, the process $G_\bullet N$ is defined in an elementary way as

$$(G_\bullet N)_t \;=\; \sum_{(s,y)\in \text{supp}(N),\; s\leq t} G_s(y).$$

Definition. *The* **predictable σ-algebra** *on $\Omega \times \mathbb{R}_+ \times S$ is the σ-algebra $\mathcal{P}$ generated by all sets of the form $A \times (s,t] \times B$ with $0 \leq s \leq t$, $A \in \mathcal{F}_s$ and $B \in \mathcal{S}$. A stochastic process defined on $\Omega \times \mathbb{R}_+ \times S$ is called $(\mathcal{F}_\mathbf{t})$* **predictable** *iff it is measurable w.r.t. $\mathcal{P}$.*

It is not difficult to verify that *any adapted left-continuous process is predictable*:

Exercise. Prove that $\mathcal{P}$ is the σ-algebra generated by all processes $(\omega, t, y) \mapsto G_t(y)(\omega)$ such that G_t is $\mathcal{F}_t \times \mathcal{S}$ measurable for any $t \geq 0$ and $t \mapsto G_t(y)(\omega)$ is left-continuous for any $y \in S$ and $\omega \in \Omega$.

Example. If (N_t) is an $(\mathcal{F}_t)$ Poisson process then the left limit process $G_t(y) = N_{t-}$ is predictable, since it is left-continuous. However, $G_t(y) = N_t$ is not predictable. This is intuitively convincing since the jumps of a Poisson process can not be "predicted in advance". A rigorous proof of the non-predictability, however, is surprisingly difficult and seems to require some background from the general theory of stochastic processes, cf. e.g. [7].

Elementary integrands

We denote by $\mathcal{E}$ the vector space consisting of all **elementary predictable processes** G of the form

$$G_t(y)(\omega) \quad = \quad \sum_{i=0}^{n-1}\sum_{k=1}^{m} Z_{i,k}(\omega)\, I_{(t_i,t_{i+1}]}(t)\, I_{B_k}(y) \tag{10.4.3}$$

with $m, n \in \mathbb{N}$, $0 \le t_0 < t_1 < \dots < t_n$, $B_1, \dots, B_m \in \mathcal{S}$ disjoint with $\nu(B_k) < \infty$, and $Z_{i,k} : \Omega \to \mathbb{R}$ bounded and $\mathcal{F}_{t_i}$-measurable. For $G \in \mathcal{E}$, the stochastic integral $G_\bullet N$ is a well-defined Lebesgue integral given by

$$(G_\bullet N)_t \quad = \quad \sum_{i=0}^{n-1} \sum_{k=1}^{m} Z_{i,k} \left(N_{t_{i+1} \wedge t}(B_k) - N_{t_i \wedge t}(B_k) \right), \tag{10.4.4}$$

Notice that the summands vanish for $t_i \ge t$ and that $G_\bullet N$ is an $(\mathcal{F}_t)$ adapted process with càdlàg paths.

Stochastic integrals w.r.t. Poisson point processes have properties reminiscent of those known from Itô integrals based on Brownian motion:

Lemma 10.12 (Elementary properties of stochastic integrals w.r.t. PPP). *Let $G \in \mathcal{E}$. Then the following assertions hold:*

1) For any $t \ge 0$,

$$E\left[(G_\bullet N)_t\right] \;=\; E\left[\int_{(0,t]\times S} G_s(y)\, ds\, \nu(dy)\right].$$

2) The process $G_\bullet \widetilde{N}$ defined by

$$(G_\bullet \widetilde{N})_t \;=\; \int_{(0,t]\times S} G_s(y)\, N(ds\, dy) \;-\; \int_{(0,t]\times S} G_s(y)\, ds\, \nu(dy)$$

is a square integrable $(\mathcal{F}_t)$ martingale with $(G_\bullet \widetilde{N})_0 = 0$.

3) For any $t \ge 0$, $G_\bullet \widetilde{N}$ satisfies the Itô isometry

$$E\left[(G_\bullet \widetilde{N})_t^2\right] \;=\; E\left[\int_{(0,t]\times S} G_s(y)^2\, ds\, \nu(dy)\right].$$

4) The process $(G_\bullet \widetilde{N})_t^2 - \int_{(0,t]\times S} G_s(y)^2\, ds\, \nu(dy)$ is an $(\mathcal{F}_t)$ martingale.

Proof. 1) Since the processes $(N_t(B_k))$ are Poisson processes with intensities $\nu(B_k)$, we obtain by conditioning on $\mathcal{F}_{t_i}$:

$$\begin{aligned} E\left[(G_\bullet N)_t\right] &= \sum_{i,k:t_i<t} E\left[Z_{i,k}\left(N_{t_{i+1}\wedge t}(B_k) - N_{t_i}(B_k)\right)\right] \\ &= \sum_{i,k} E\left[Z_{i,k}\left(t_{i+1}\wedge t - t_i \wedge t\right)\, \nu(B_k)\right] \\ &= E\left[\int_{(0,t]\times S} G_s(y)\, ds\, \nu(dy)\right]. \end{aligned}$$

2) The process $G_\bullet\widetilde{N}$ is bounded and hence square integrable. Moreover, it is a martingale, since by 1), for any $0 \le s \le t$ and $A \in \mathcal{F}_s$,

$$\begin{aligned}
E\left[(G_\bullet N)_t - (G_\bullet N)_s;\, A\right] &= E\left[\int_{(0,t]\times S} I_A\, G_r(y)\, I_{(s,t]}(r)\, N(dr\, dy)\right] \\
&= E\left[\int_{(0,t]\times S} I_A\, G_r(y)\, I_{(s,t]}(r)\, dr\, \nu(dy)\right] \\
&= E\left[\int_{(0,t]\times S} G_r(y)\, dr\, \nu(dy) - \int_{(0,s]\times S} G_r(y)\, dr\, \nu(dy)\,;\, A\right] \\
&= E\left[\int_{(0,t]\times S} G_s(y)\, ds\, \nu(dy)\right].
\end{aligned}$$

3) We have $(G_\bullet\widetilde{N})_t = \sum_{i,k} Z_{i,k}\, \Delta_i \widetilde{N}(B_k)$, where

$$\Delta_i \widetilde{N}(B_k) \;:=\; \widetilde{N}_{t_{i+1}\wedge t}(B_k) - \widetilde{N}_{t_i \wedge t}(B_k)$$

are increments of independent compensated Poisson point processes. Noticing that the summands vanish if $t_i \ge t$, we obtain

$$\begin{aligned}
E\left[(G_\bullet\widetilde{N})_t^2\right] &= \sum_{i,j,k,l} E\left[Z_{i,k}Z_{j,l}\Delta_i\widetilde{N}(B_k)\Delta_j\widetilde{N}(B_l)\right] \\
&= 2\sum_{k,l}\sum_{i<j} E\left[Z_{i,k}Z_{j,l}\Delta_i\widetilde{N}(B_k)\, E[\Delta_j\widetilde{N}(B_l)|\mathcal{F}_{t_j}]\right] \\
&\quad + \sum_{k,l}\sum_{i} E\left[Z_{i,k}Z_{i,l}\, E[\Delta_i\widetilde{N}(B_k)\Delta_i\widetilde{N}(B_l)|\mathcal{F}_{t_i}]\right] \\
&= \sum_k\sum_i E[Z_{i,k}^2\, \Delta_i t]\, \nu(B_k) \;=\; E\left[\int_{(0,t]\times S} G_s(y)^2\, ds\, \nu(dy)\right].
\end{aligned}$$

Here we have used that the coefficients $Z_{i,k}$ are $\mathcal{F}_{t_i}$ measurable, and the increments $\Delta_i\widetilde{N}(B_k)$ are independent of $\mathcal{F}_{t_i}$ with covariance $E[\Delta_i\widetilde{N}(B_k)\Delta_i\widetilde{N}(B_l)] = \delta_{kl}\nu(B_k)\Delta_i t$.

4) now follows similarly as 2), and is left as an exercise to the reader. □

Lebesgue integrals

If the integrand $G_t(y)$ is non-negative, then the integrals (10.4.1) and (10.4.2) are well-defined Lebesgue integrals for every ω. By Lemma 10.12 and monotone convergence, the identity

$$E\left[(G_\bullet N)_t\right] \;=\; E\left[\int_{(0,t]\times S} G_s(y)\,ds\,\nu(dy)\right] \tag{10.4.5}$$

holds for any predictable $G \geq 0$.

Now let $u \in (0,\infty]$, and suppose that $G : \Omega \times (0,u) \times S \to \mathbb{R}$ is predictable and integrable w.r.t. the product measure $P \otimes \lambda_{(0,u)} \otimes \nu$. Then by (10.4.5),

$$E\left[\int_{(0,u]\times S} |G_s(y)|\,N(ds\,dy)\right] \;=\; E\left[\int_{(0,u]\times S} |G_s(y)|\,ds\,\nu(dy)\right] \;<\; \infty.$$

Hence the processes $G^+_\bullet N$ and $G^-_\bullet N$ are almost surely finite on $[0,u]$, and, correspondingly $G_\bullet N = G^+_\bullet N - G^-_\bullet N$ is almost surely well-defined as a Lebesgue integral, and it satisfies the identity (10.4.5).

Theorem 10.13. *Suppose that $G \in \mathcal{L}^1(P\otimes\lambda_{(0,u)}\otimes\nu)$ is predictable. Then the following assertions hold:*

1) *$G_\bullet N$ is an $(\mathcal{F}_t^P)$ adapted stochastic process satisfying (10.4.5).*

2) *The compensated process $G_\bullet \widetilde{N}$ is an $(\mathcal{F}_t^P)$ martingale.*

3) *The sample paths $t \mapsto (G_\bullet N)_t$ are càdlàg with almost surely finite variation*

$$V_t^{(1)}(G_\bullet N) \;\leq\; \int_{(0,t]\times S} |G_s(y)|\,N(ds\;dy).$$

Proof. 1) extends by a monotone class argument from elementary predictable G to general non-negative predictable G, and hence also to integrable predictable G.

2) can be verified similarly as in the proof of Lemma 10.12.

3) We may assume w.l.o.g. $G \geq 0$, otherwise we consider $G^+_\bullet N$ and $G^-_\bullet N$ separately. Then, by the Monotone Convergence Theorem,

$$\begin{aligned} (G_\bullet N)_{t+\varepsilon} - (G_\bullet N)_t &= \int_{(t,t+\varepsilon]\times S} G_s(y)\, N(ds\, dy) \;\to\; 0, \qquad \text{and} \\ (G_\bullet N)_t - (G_\bullet N)_{t-\varepsilon} &\to \int_{\{t\}\times S} G_s(y)\, N(ds\, dy) \end{aligned}$$

as $\varepsilon \downarrow 0$. This shows that the paths are càdlàg. Moreover, for any partition π of $[0,u]$,

$$\begin{aligned} \sum_{r\in\pi} |(G_\bullet N)_{r'} - (G_\bullet N)_r| &= \sum_{r\in\pi} \left| \int_{(r,r']\times S} G_s(y)\, N(ds\, dy) \right| \\ &\leq \int_{(0,u]\times S} |G_s(y)|\, N(ds\, dy) \;<\; \infty \qquad \text{a.s.} \end{aligned}$$

□

Remark (Watanabe characterization). It can be shown that a counting measure valued process (N_t) is an $(\mathcal{F}_t)$ Poisson point process if and only if (10.4.5) holds for any non-negative predictable process G.

Itô integrals w.r.t. compensated Poisson point processes

Suppose that $(\omega, s, y) \mapsto G_s(y)(\omega)$ is a predictable process in $\mathcal{L}^2(P \otimes \lambda_{(0,u)} \otimes \nu)$ for some $u \in (0, \infty]$. If G is not integrable w.r.t. the product measure, then the integral $G_\bullet N$ does not exist in general. Nevertheless, under the square integrability assumption, the integral $G_\bullet \widetilde{N}$ w.r.t. the compensated Poisson point process exists as a square integrable martingale. Note that square integrability does not imply integrability if the intensity measure ν is not finite.

To define the stochastic integral $G_\bullet \widetilde{N}$ for square integrable integrands G we use the Itô isometry. Let

$$\mathcal{M}^2_d([0,u]) \quad = \quad \left\{ M \in \mathcal{M}^2([0,u]) \mid t \mapsto M_t(\omega) \text{ càdlàg for any } \omega \in \Omega \right\}$$

denote the space of all square-integrable càdlàg martingales w.r.t. the completed filtration $(\mathcal{F}^P_t)$. Recall that the L^2 maximal inequality

$$E\Big[\sup_{t\in[0,u]} |M_t|^2 \Big] \quad \leq \quad \left(\frac{2}{2-1} \right)^2 E[|M_u|^2]$$

holds for any right-continuous martingale in $\mathcal{M}^2([0,u])$. Since a uniform limit of càdlàg functions is again càdlàg, this implies that the space $M^2_d([0,u])$ of equivalence classes of indistinguishable martingales in $\mathcal{M}^2_d([0,u])$ is a **closed** subspace of the Hilbert space $M^2([0,u])$ w.r.t. the norm

$$||M||_{M^2([0,u])} \quad = \quad E[|M_u|^2]^{1/2}.$$

Lemma 10.12, 3), shows that for elementary predictable processes G,

$$||G_\bullet \widetilde{N}||_{M^2([0,u])} \quad = \quad ||G||_{L^2(P\otimes\lambda_{(0,u)}\otimes\nu)}. \tag{10.4.6}$$

On the other hand, it can be shown that any predictable process $G \in L^2(P \otimes \lambda_{(0,u)} \otimes \nu)$ is a limit w.r.t. the $L^2(P\otimes\lambda_{(0,u)}\otimes\nu)$ norm of a sequence $(G^{(k)})$ consisting of elementary predictable processes. Hence isometric extension of the linear map $G \mapsto G_\bullet \widetilde{N}$ can be used to define $G_\bullet \widetilde{N} \in M^2_d(0,u)$ for any predictable $G \in L^2(P \otimes \lambda_{(0,u)} \otimes \nu)$ in such a way that

$$G^{(k)}_\bullet \widetilde{N} \quad \longrightarrow \quad G_\bullet \widetilde{N} \quad \text{in } M^2 \qquad \text{whenever} \quad G^{(k)} \to G \text{ in } L^2.$$

Theorem 10.14 (Itô isometry and stochastic integrals w.r.t. compensated PPP). *Suppose that $u \in (0,\infty]$. Then there is a unique linear isometry $G \mapsto G_\bullet \widetilde{N}$ from $L^2(\Omega \times (0,u) \times S, \mathcal{P}, P \otimes \lambda \otimes \nu)$ to $M^2_d([0,u])$ such that $G_\bullet \widetilde{N}$ is given by (10.4.4) for any elementary predictable process G of the form (10.4.3).*

Proof. As pointed out above, by (10.4.6), the stochastic integral extends isometrically to the closure $\bar{\mathcal{E}}$ of the subspace of elementary predictable processes in the Hilbert space $L^2(\Omega \times (0,u) \times S, \mathcal{P}, P \otimes \lambda \otimes \nu)$. It only remains to show that *any* square integrable predictable process G is contained in $\bar{\mathcal{E}}$, i.e., G is an L^2 limit of elementary predictable processes. This holds by dominated convergence for bounded left-continuous processes, and by a monotone class argument or a direct approximation for general bounded predictable processes, and hence also for predictable processes in L^2. The details are left to the reader. □

The definition of stochastic integrals w.r.t. compensated Poisson point processes can be extended to locally square integrable predictable processes G by localization – we refer to [5] for details.

Example (Deterministic integrands). If $H_s(y)(\omega) = h(y)$ for some function $h \in \mathcal{L}^2(S, \mathcal{S}, \nu)$ then

$$(H_\bullet \widetilde{N})_t \quad = \quad \int h(y)\, \widetilde{N}_t(dy) \quad = \quad \widetilde{N}_t(h),$$

i.e., $H_\bullet \widetilde{N}$ is a Lévy martingale with jump intensity measure $\nu \circ h^{-1}$.

10.5 Lévy processes with infinite jump intensity

In this section, we are going to construct general Lévy processes from Poisson point processes and Brownian motion. Afterwards, we will consider several important classes of Lévy jump processes with infinite jump intensity.

Construction from Poisson point processes

Let $\nu(dy)$ be a positive measure on $\mathbb{R}^d \setminus \{0\}$ such that $\int (1 \wedge |y|^2)\, \nu(dy) < \infty$, i.e.,

$$\nu(|y| > \varepsilon) \quad < \quad \infty \quad \text{for any } \varepsilon > 0, \text{ and} \tag{10.5.1}$$

$$\int_{|y|\le 1} |y|^2\, \nu(dy) \quad < \quad \infty. \tag{10.5.2}$$

Note that the condition (10.5.1) is necessary for the existence of a Lévy process with jump intensity ν. Indeed, if (10.5.1) would be violated for some $\varepsilon > 0$ then a corresponding Lévy process should have infinitely many jumps of size greater than ε in finite time. This contradicts the càdlàg property of the paths. The square integrability condition (10.5.2) controls the intensity of small jumps. It is crucial for the construction of a Lévy process with jump intensity ν given below, and actually it turns out to be also necessary for the existence of a corresponding Lévy process.

In order to construct the Lévy process, let $N_t(dy)$, $t \ge 0$, be a Poisson point process with intensity measure ν defined on a probability space $(\Omega, \mathcal{A}, P)$, and let $\widetilde{N}_t(dy) :=$

$N_t(dy) - t\,\nu(dy)$ denote the compensated process. For a measure μ and a measurable set A, we denote by

$$\mu^A(B) \quad = \quad \mu(B \cap A)$$

the part of the measure on the set A, i.e., $\mu^A(dy) = I_A(y)\mu(dy)$. The following decomposition property is immediate from the definition of a Poisson point process:

Remark (Decomposition of Poisson point processes). If $A, B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ are disjoint sets then $(N_t^A)_{t\geq 0}$ and $(N_t^B)_{t\geq 0}$ are independent Poisson point processes with intensity measures ν^A, ν^B respectively.

If $A \cap B_e(y) = \emptyset$ for some $\varepsilon > 0$ then the measure ν^A has finite total mass $\nu^A(\mathbb{R}^d) = \nu(A)$ by (10.5.1). Therefore,

$$X_t^A \;:=\; \int_A y\, N_t(dy) \;=\; \int y\, N_t^A(dy)$$

is a compound Poisson process with intensity measure ν^A, and characteristic exponent

$$\psi_{X^A}(p) \;=\; \int_A (1 - \exp(ip \cdot y))\, \nu(dy).$$

On the other hand, if $\int_A |y|^2\, \nu(dy) < \infty$ then

$$M_t^A \;=\; \int_A y\, \widetilde{N}_t(dy) \;=\; \int y\, \widetilde{N}_t^A(dy)$$

is a square integrable martingale. If both conditions are satisfied simultaneously then

$$M_t^A \;=\; X_t^A - t \int_A y\, \nu(dy).$$

In particular, in this case M^A is a Lévy process with characteristic exponent

$$\psi_{M^A}(p) \;=\; \int_A (1 - \exp(ip \cdot y) + ip \cdot y)\, \nu(dy).$$

By (10.5.1) and (10.5.2), we are able to construct a Lévy process with jump intensity measure ν that is given by

$$\widetilde{X}_t^r \;=\; \int_{|y|>r} y\, N_t(dy) \;+\; \int_{|y|\leq r} y\, \widetilde{N}_t(dy). \tag{10.5.3}$$

for any $r \in (0, \infty)$. Indeed, let

$$X_t^r \;:=\; \int_{|y|>r} y\, N_t(dy) \;=\; \int_{(0,t]\times\mathbb{R}^d} y\, I_{\{|y|>r\}}\, N(ds\, dy), \qquad \text{and} \quad (10.5.4)$$

$$M_t^{\varepsilon,r} \;:=\; \int_{\varepsilon<|y|\le r} y\, \widetilde{N}_t(dy). \qquad (10.5.5)$$

for $\varepsilon, r \in [0, \infty)$ with $\varepsilon < r$. As a consequence of the Itô isometry for Poisson point processes, we obtain:

Theorem 10.15 (Existence of Lévy processes with infinite jump intensity). *Let ν be a positive measure on $\mathbb{R}^d \setminus \{0\}$ satisfying $\int (1 \wedge |y|^2)\, \nu(dy) < \infty$.*

1) *For any $r > 0$, (X_t^r) is a compound Poisson process with intensity measure $\nu^r(dy) = I_{\{|y|>r\}}\, \nu(dy)$.*

2) *The process $(M_t^{0,r})$ is a Lévy martingale with characteristic exponent*

$$\psi_r(p) \quad = \quad \int_{|y|\le r} (1 - e^{ip\cdot y} + ip\cdot y)\, \nu(dy) \qquad \forall\, p \in \mathbb{R}^d. \qquad (10.5.6)$$

Moreover, for any $u \in (0, \infty)$,

$$E\left[\sup_{0\le t\le u} |\, M_t^{\varepsilon,r} - M_t^{0,r}|^2\right] \quad \to \quad 0 \qquad \text{as } \varepsilon \downarrow 0. \qquad (10.5.7)$$

3) *The Lévy processes $(M_t^{0,r})$ and (X_t^r) are independent, and $\widetilde{X}_t^r := X_t^r + M_t^{0,r}$ is a Lévy process with characteristic exponent*

$$\widetilde{\psi}_r(p) \quad = \quad \int \left(1 - e^{ip\cdot y} + ip\cdot y I_{\{|y|\le r\}}\right) \nu(dy) \qquad \forall\, p \in \mathbb{R}^d. \qquad (10.5.8)$$

Proof. 1) is a consequence of Theorem 10.10.

2) By (10.5.2), the stochastic integral $(M_t^{0,r})$ is a square integrable martingale on $[0, u]$ for any $u \in (0, \infty)$. Moreover, by the Itô isometry,

$$\|M^{0,r} - M^{\varepsilon,r}\|^2_{M^2([0,u])} \;=\; \|M^{0,\varepsilon}\|^2_{M^2([0,u])} \;=\; \int_0^u \int |y|^2\, I_{\{|y|\le\varepsilon\}}\, \nu(dy)\, dt \;\to\; 0$$

as $\varepsilon \downarrow 0$. By Theorem 10.10, $(M_t^{\varepsilon,r})$ is a compensated compound Poisson process with intensity $I_{\{\varepsilon<|y|\leq r\}}\,\nu(dy)$ and characteristic exponent

$$\psi_{\varepsilon,r}(p) \;=\; \int_{\varepsilon<|y|\leq r} (1-e^{ip\cdot y}+ip\cdot y)\,\nu(dy).$$

As $\varepsilon \downarrow 0$, $\psi_{\varepsilon,r}(p)$ converges to $\psi_r(p)$ since $1-e^{ip\cdot y}+ip\cdot y = \mathcal{O}(|y|^2)$. Hence the limit martingale $M_t^{0,r} \;=\; \lim\limits_{n\to\infty} M_t^{1/n,r}$ also has independent and stationary increments, and characteristic function

$$E[\exp(ip\cdot M_t^{0,r})] \;=\; \lim_{n\to\infty} E[\exp(ip\cdot M_t^{1/n,1})] \;=\; \exp(-t\psi_r(p)).$$

3) Since $I_{\{|y|\leq r\}}\,N_t(dy)$ and $I_{\{|y|>r\}}\,N_t(dy)$ are independent Poisson point processes, the Lévy processes $(M_t^{0,r})$ and (X_t^r) are also independent. Hence $\widetilde{X}_t^r = M_t^{0,r} + X_t^r$ is a Lévy process with characteristic exponent

$$\widetilde{\psi}_r(p) = \psi_r(p) + \int_{|y|>r} (1-e^{ip\cdot y})\,\nu(dy).$$

$\square$

Remark. All the partially compensated processes $(\widetilde{X}_t^r)$, $r\in(0,\infty)$, are Lévy processes with jump intensity ν. Actually, these processes differ only by a finite drift term, since for any $0<\varepsilon<r$,

$$\widetilde{X}_t^\varepsilon \;=\; \widetilde{X}_t^r + bt, \qquad \text{where } \; b = \int_{\varepsilon<|y|\leq r} y\,\nu(dy).$$

A totally uncompensated Lévy process

$$X_t \;=\; \lim_{n\to\infty} \int_{|y|\geq 1/n} y\,N_t(dy)$$

does exist only under additional assumptions on the jump intensity measure:

Corollary 10.16 (Existence of uncompensated Lévy jump processes). *Suppose that* $\int (1\wedge|y|)\,\nu(dy) < \infty$, *or that* ν *is symmetric (i.e.,* $\nu(B) = \nu(-B)$ *for any* $B \in$

$\mathcal{B}(\mathbb{R}^d \setminus \{0\})$) and $\int (1 \wedge |y|^2)\, \nu(dy) < \infty$. Then there exists a Lévy process (X_t) with characteristic exponent

$$\psi(p) \quad = \quad \lim_{\varepsilon \downarrow 0} \int_{|y|>\varepsilon} \left(1 - e^{ip\cdot y}\right)\, \nu(dy) \qquad \forall\, p \in \mathbb{R}^d \tag{10.5.9}$$

such that

$$E\left[\sup_{0\le t\le u} |\, X_t - X_t^\varepsilon|^2\right] \;\to\; 0 \qquad \text{as } \varepsilon \downarrow 0. \tag{10.5.10}$$

Proof. For $0 < \varepsilon < r$, we have

$$X_t^\varepsilon \;=\; X_t^r \,+\, M_t^{\varepsilon,r} \,+\, t\int_{\varepsilon<|y|\le r} y\; \nu(dy).$$

As $\varepsilon \downarrow 0$, $M^{\varepsilon,r}$ converges to $M^{0,r}$ in $M^2([0,u])$ for any finite u. Moreover, under the assumption imposed on ν, the integral on the right hand side converges to bt where

$$b \;=\; \lim_{\varepsilon\downarrow 0} \int_{\varepsilon<|y|\le r} y\; \nu(dy).$$

Therefore, (X_t^ε) converges to a process (X_t) in the sense of (10.5.10) as $\varepsilon \downarrow 0$. The limit process is again a Lévy process, and, by dominated convergence, the characteristic exponent is given by (10.5.9). □

Remark (Lévy processes with finite variation paths). If $\int (1 \wedge |y|)\, \nu(dy) < \infty$ then the process $X_t = \int y\, N_t(dy)$ is defined as a Lebesgue integral. As remarked above, in that case the paths of (X_t) are almost surely of finite variation:

$$V_t^{(1)}(X) \;\le\; \int |y|\, N_t(dy) \;<\; \infty \qquad \text{a.s.}$$

The Lévy-Itô decomposition

We have constructed Lévy processes corresponding to a given jump intensity measure ν under adequate integrability conditions as limits of compound Poisson processes or partially compensated compound Poisson processes, respectively. Remarkably, it turns out that by taking linear combinations of these Lévy jump processes and Gaussian Lévy processes, we obtain all Lévy processes. This is the content of the Lévy-Itô decomposition theorem that we will now state before considering in more detail some important classes of Lévy processes.

Already the classical Lévy-Khinchin formula for infinity divisible random variables (see Corollary 10.18 below) shows that any Lévy process on $\mathbb{R}^d$ can be *characterized by three quantities*: a non-negative definite symmetric matrix $a \in \mathbb{R}^{d\times d}$, a vector $b \in \mathbb{R}^d$, and a σ-finite measure ν on $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$ such that

$$\int (1 \wedge |y|^2)\, \nu(dy) \quad < \quad \infty \qquad . \tag{10.5.11}$$

Note that (10.5.11) holds if and only if ν is finite on complements of balls around 0, and $\int_{|y|\leq 1} |y|^2\, \nu(dy) < \infty$. The Lévy-Itô decomposition gives an explicit representation of a Lévy process with characteristics (a, b, ν).

Let $\sigma \in \mathbb{R}^{d\times d}$ with $a = \sigma\sigma^T$, let (B_t) be a d-dimensional Brownian motion, and let (N_t) be an independent Poisson point process with intensity measure ν. We define a Lévy process (X_t) by setting

$$X_t = \sigma B_t + bt + \int_{|y|>1} y\ N_t(dy) + \int_{|y|\leq 1} y\ \left(N_t(dy) - t\nu(dy)\right). \tag{10.5.12}$$

The first two summands are the diffusion part and the drift of a Gaussian Lévy process, the third summand is a pure jump process with jumps of size greater than 1, and the last summand represents small jumps compensated by drift. As a sum of independent Lévy processes, the process (X_t) is a Lévy process with characteristic exponent

$$\psi(p) \quad = \quad \frac{1}{2}\, p\cdot ap - ib\cdot p + \int_{\mathbb{R}^d\setminus\{0\}} (1 - e^{ip\cdot y} + ip\cdot y\ I_{\{|y|\leq 1\}})\ \nu(dy). \tag{10.5.13}$$

We have thus proved the first part of the following theorem:

Theorem 10.17 (Lévy-Itô decomposition).

1) The expression (10.5.12) defines a Lévy process with characteristic exponent ψ.

2) *Conversely, any Lévy process* (X_t) *can be decomposed as in (10.5.12) with the Poisson point process*

$$N_t \quad = \quad \sum_{\substack{s \le t \\ \Delta X_s \neq 0}} \Delta X_s \qquad , \quad t \geq 0, \tag{10.5.14}$$

an independent Brownian motion (B_t)*, a matrix* $\sigma \in \mathbb{R}^{d \times d}$*, a vector* $b \in \mathbb{R}^d$*, and a* σ*-finite measure* ν *on* $\mathbb{R}^d \setminus \{0\}$ *satisfying (10.5.11).*

We will not prove the second part of the theorem here. The principal way to proceed is to define (N_t) via (10.5.11), and to consider the difference of (X_t) and the integrals w.r.t. (N_t) on the right hand side of (10.5.12). One can show that the difference is a continuous Lévy process which can then be identified as a Gaussian Lévy process by the Lévy characterization, cf. Section 11.1 below. Carrying out the details of this argument, however, is still a lot of work. A detailed proof is given in [5].

As a byproduct of the Lévy-Itô decomposition, one recovers the classical Lévy-Khinchin formula for the characteristic functions of infinitely divisible random variables, which can also be derived directly by an analytic argument.

Corollary 10.18 (Lévy-Khinchin formula). *For a function* $\psi : \mathbb{R}^d \to \mathbb{C}$ *the following statements are all equivalent:*

(i) ψ *is the characteristic exponent of a Lévy process.*

(ii) $\exp(-\psi)$ *is the characteristic function of an infinitely divisible random variable.*

(iii) ψ *satisfies (10.5.13) with a non-negative definite symmetric matrix* $a \in \mathbb{R}^{d \times d}$*, a vector* $b \in \mathbb{R}^d$*, and a measure* ν *on* $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$ *such that* $\int (1 \wedge |y|^2)\, \nu(dy) < \infty$.

Proof. (iii)⇒(i) holds by the first part of Theorem 10.17.
(i)⇒(ii): If (X_t) is a Lévy process with characteristic exponent ψ then $X_1 - X_0$ is an

infinitely divisible random variable with characteristic function $\exp(-\psi)$.
(ii)⇒(iii) is the content of the classical Lévy-Khinchin theorem, see e.g. [17]. □

We are now going to consider several important subclasses of Lévy processes. The class of Gaussian Lévy processes of type

$$X_t \quad = \quad \sigma B_t + bt$$

with $\sigma \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$, and a d-dimensional Brownian motion (B_t) has already been introduced before. The Lévy-Itô decomposition states in particular that these are the only Lévy processes with continuous paths!

Subordinators

A **subordinator** is by definition a non-decreasing real-valued Lévy process. The name comes from the fact that subordinators are used to change the time-parametrization of a Lévy process, cf. below. Of course, the deterministic processes $X_t = bt$ with $b \geq 0$ are subordinators. Furthermore, any compound Poisson process with non-negative jumps is a subordinator. To obtain more interesting examples, we assume that ν is a positive measure on $(0, \infty)$ with

$$\int_{(0,\infty)} (1 \wedge y)\, \nu(dy) \quad < \quad \infty.$$

Then a Poisson point process (N_t) with intensity measure ν satisfies almost surely

$$\mathrm{supp}(N_t) \quad \subset \quad [0, \infty) \qquad \text{for any } t \geq 0.$$

Hence the integrals

$$X_t \quad = \quad \int y\, N_t(dy) \quad , \quad t \geq 0,$$

define a non-negative Lévy process with $X_0 = 0$. By stationarity, all increments of (X_t) are almost surely non-negative, i.e., (X_t) is increasing. In particular, the sample paths are (almost surely) of finite variation.

Example (Gamma process). The Gamma distributions form a convolution semigroup of probability measures on $(0, \infty)$, i.e.,

$$\Gamma(r, \lambda) * \Gamma(s, \lambda) \quad = \quad \Gamma(r + s, \lambda) \qquad \text{for any } r, s, \lambda > 0.$$

Therefore, for any $a, \lambda > 0$ there exists an increasing Lévy process $(\Gamma_t)_{t \geq 0}$ with increment distributions

$$\Gamma_{t+s} - \Gamma_s \quad \sim \quad \Gamma(at, \lambda) \qquad \text{for any } s, t \geq 0.$$

Computation of the Laplace transform yields

$$E[\exp(-u\Gamma_t)] \;=\; \left(1 + \frac{u}{\lambda}\right)^{-at} \;=\; \exp\left(-t\int_0^\infty (1 - e^{-uxy}) a y^{-1} e^{-\lambda y}\, dy\right) \tag{10.5.15}$$

for every $u \geq 0$, cf. e.g. [28, Lemma 1.7]. Since $\Gamma_t \geq 0$, both sides in (10.5.15) have a unique analytic extension to $\{u \in \mathbb{C} \,:\, \Re(u) \geq 0\}$. Therefore, we can replace u by $-ip$ in (10.5.15) to conclude that the characteristic exponent of (Γ_t) is

$$\psi(p) \quad = \quad \int_0^\infty (1 - e^{ipy})\, \nu(dy), \qquad \text{where } \nu(dy) = a y^{-1} e^{-\lambda y}\, dy.$$

Hence the Gamma process is a non-decreasing pure jump process with jump intensity measure ν.

Example (Inverse Gaussian processes). An explicit computation of the characteristic function shows that the Lévy subordinator (T_s) is a pure jump Lévy process with Lévy measure

$$\nu(dy) \quad = \quad (2\pi)^{-1/2}\, y^{-3/2}\, I_{(0,\infty)}(y)\, dx.$$

More generally, if $X_t = \sigma B_t + bt$ is a Gaussian Lévy process with coefficients $\sigma > 0$, $b \in \mathbb{R}$, then the right inverse

$$T_s^X \quad = \quad \inf\{t \geq 0 \,:\, X_t = s\} \quad , \quad s \geq 0,$$

is a Lévy jump process with jump intensity

$$\nu(dy) \quad = \quad (2\pi)^{-1/2} y^{-3/2} \exp(-b^2 y/2) I_{(0,\infty)}(y)\, dy.$$

Remark (Finite variation Lévy jump processes on $\mathbb{R}^1$).
Suppose that (N_t) is a Poisson point process on $\mathbb{R} \setminus \{0\}$ with jump intensity measure ν satisfying $\int (1 \wedge |y|)\, \nu(dy) < \infty$. Then the decomposition $N_t = N_t^{(0,\infty)} + N_t^{(-\infty,0)}$ into the independent restrictions of (N_t) to $\mathbb{R}_+$, $\mathbb{R}_-$ respectively induces a corresponding decomposition

$$X_t = X_t^{\nearrow} + X_t^{\searrow} \quad , \quad X_t^{\nearrow} = \int y \, N_t^{(0,\infty)}(dy) \quad , \quad X_t^{\searrow} = \int y \, N_t^{(-\infty,0)}(dy),$$

of the associated Lévy jump process $X_t = \int y \, N_t(dy)$ into a subordinator $X_t^{\nearrow}$ and a decreasing Lévy process $X_t^{\searrow}$. In particular, we see once more that (X_t) has almost surely paths of finite variation.

An important property of subordinators is that they can be used for random time transformations of Lévy processes:

Exercise (Time change by subordinators). Suppose that (X_t) is a Lévy process with Laplace exponent $\eta_X : \mathbb{R}_+ \to \mathbb{R}$, i.e.,

$$E[\exp(-\alpha X_t)] \quad = \quad \exp(-t\eta_X(\alpha)) \quad \text{for any } \alpha \geq 0.$$

Prove that if (T_s) is an independent subordinator with Laplace exponent η_T then the time-changed process

$$\widetilde{X}_s \quad := \quad X_{T_s} \quad , \quad s \geq 0,$$

is again a Lévy process with Laplace exponent

$$\widetilde{\eta}(p) \quad = \quad \eta_T(\eta_X(p)).$$

The characteristic exponent can be obtained from this identity by analytic continuation.

Example (Subordinated Lévy processes). Let (B_t) be a Brownian motion.

1) If (N_t) is an independent Poisson process with parameter $\lambda > 0$ then (B_{N_t}) is a compensated Poisson process with Lévy measure

$$\nu(dy) \quad = \quad \lambda(2\pi)^{-1/2} \exp(-y^2/2)\, dy.$$

2) If (Γ_t) is an independent Gamma process then for $\sigma, b \in \mathbb{R}$ the process

$$X_t = \sigma B_{\Gamma_t} + b\Gamma_t$$

is called a **Variance Gamma process**. It is a Lévy process with characteristic exponent $\psi(p) = \int (1 - e^{ipy})\, \nu(dy)$, where

$$\nu(dy) = c|y|^{-1} \left(e^{-\lambda y} I_{(0,\infty)}(y) + e^{-\mu|y|} I_{(-\infty,0)}(y) \right) dy$$

with constants $c, \lambda, \mu > 0$. In particular, a Variance Gamma process satisfies $X_t = \Gamma_t^{(1)} - \Gamma_t^{(2)}$ with two independent Gamma processes. Thus the increments of (X_t) have exponential tails. Variance Gamma processes have been introduced and applied to option pricing by Madan and Seneta [31] as an alternative to Brownian motion taking into account longer tails and allowing for a wider modeling of skewness and kurtosis.

3) **Normal Inverse Gaussian (NIG) processes** are time changes of Brownian motions with drift by inverse Gaussian subordinators [6]. Their increments over unit time intervals have a normal inverse Gaussian distribution, which has slower decaying tails than a normal distribution. NIG processes are applied in statistical modelling in finance and turbulence.

Stable processes

We have noted in (10.3.2) that the jump intensity measure of a strictly α-stable process in $\mathbb{R}^1$ is given by

$$\nu(dy) = \left(c_+ I_{(0,\infty)}(y) + c_- I_{(-\infty,0)}(y) \right) |y|^{-1-\alpha}\, dy \tag{10.5.16}$$

with constants $c_+, c_- \in [0, \infty)$. Note that for any $\alpha \in (0, 2)$, the measure ν is finite on $\mathbb{R} \setminus (-1, 1)$, and $\int_{[-1,1]} |y|^2 \nu(dy) < \infty$.

We will prove now that if $\alpha \in (0, 1) \cup (1, 2)$ then for each choice of the constants c_+ and c_-, there is a strictly α-stable process with Lévy measure (10.5.16). For $\alpha = 1$ this is only true if $c_+ = c_-$, whereas a non-symmetric 1-stable process is given by $X_t = bt$ with $b \in \mathbb{R} \setminus \{0\}$. To define the corresponding α-stable processes, let

$$X_t^\varepsilon = \int_{\mathbb{R} \setminus [-\varepsilon, \varepsilon]} y\, N_t(dy)$$

where (N_t) is a Poisson point process with intensity measure ν. Setting $||X||_u = E[\sup_{t\leq a}|X_t|^2]^{1/2}$, an application of Theorem 10.15 yields:

Corollary 10.19 (Construction of α-stable processes). *Let ν be the probability measure on $\mathbb{R}\setminus\{0\}$ defined by (10.5.16) with $c_+, c_- \in [0,\infty)$.*

1) *If $c_+ = c_-$ then there exists a symmetric α-stable process X with characteristic exponent $\psi(p) = \gamma\,|p|^\alpha$, $\gamma = \int(1-\cos y)\,\nu(dy)\ \in\mathbb{R}$, such that $||X^{1/n} - X||_u \to 0$ for any $u \in (0,\infty)$.*

2) *If $\alpha \in (0,1)$ then $\int(1\wedge|y|)\,\nu(dy) < \infty$, and $X_t = \int y\,N_t(dy)$ is an α-stable process with characteristic exponent $\psi(p) = z\,|p|^\alpha$, $z = \int\left(1-e^{iy}\right)\nu(dy)\ \in\mathbb{C}$.*

3) *For $\alpha = 1$ and $b\in\mathbb{R}$, the deterministic process $X_t = bt$ is α-stable with characteristic exponent $\psi(p) = -ibp$.*

4) *Finally, if $\alpha\in(1,2)$ then $\int(|y|\wedge|y|^2)\,\nu(dy) < \infty$, and the compensated process $X_t = \int y\,\widetilde{N}_t(dy)$ is an α-stable martingale with characteristic exponent $\psi(p) = \widetilde{z}\cdot|p|^\alpha$, $\widetilde{z} = \int(1-e^{iy}+iy)\,\nu(dy)$.*

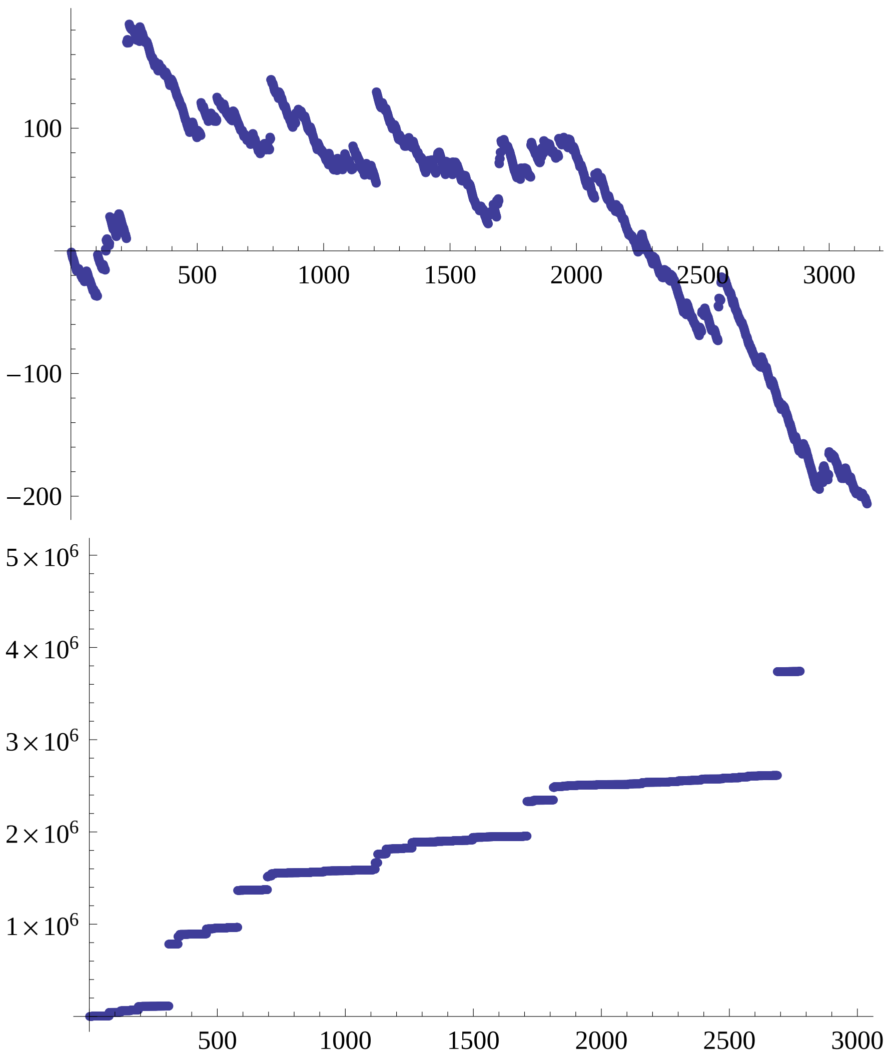
Proof. By Theorem 10.15 it is sufficient to prove convergence of the characteristic exponents

$$\begin{aligned}
\psi_\varepsilon(p) &= \int\limits_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]}\left(1-e^{ipy}\right)\nu(dy) &&= |p|^\alpha\int\limits_{\mathbb{R}\setminus[-\varepsilon p,\varepsilon p]}\left(1-e^{ix}\right)\nu(dx),\\
\widetilde{\psi}_\varepsilon(p) &= \int\limits_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]}\left(1-e^{ipy}+ipy\right)\nu(dy) &&= |p|^\alpha\int\limits_{\mathbb{R}\setminus[-\varepsilon p,\varepsilon p]}\left(1-e^{ix}+ix\right)\nu(dx)
\end{aligned}$$

to $\psi(p)$, $\widetilde{\psi}(p)$ respectively as $\varepsilon\downarrow 0$. This is easily verified in cases 1), 2) and 4) by noting that $1-e^{ix}+1-e^{-ix} = 2(1-\cos x) = \mathcal{O}(x^2)$, $1-e^{ix} = \mathcal{O}(|x|)$, and $1-e^{ix}+ix = \mathcal{O}(|x|^2)$. □

Notice that although the characteristic exponents in the non-symmetric cases 2), 3) and 4) above take a similar form (but with different constants), the processes are actually very different. In particular, for $\alpha > 1$, a strictly α-stable process is always a limit of compensated compound Poisson processes and hence a martingale!

Example (α-stable subordinators vs. α-stable martingales). For $c_- = 0$ and $\alpha \in (0,1)$, the α-stable process with jump intensity ν is increasing, i.e., it is an **α-stable subordinator**. For $c_- = 0$ and $\alpha \in (1,2)$ this is not the case since the jumps are "**compensated by an infinite drift**". The graphics below show simulations of samples from α-stable processes for $c_- = 0$ and $\alpha = 3/2$, $\alpha = 1/2$ respectively. For $\alpha \in (0,2)$, a symmetric α-stable process has the same law as $(\sqrt{2}B_{T_s})$ where (B_t) is a Brownian motion and (T_s) is an independent $\alpha/2$-stable subordinator.



Chapter 11

Transformations of SDE

Let $U \subseteq \mathbb{R}^n$ be an open set. We consider a stochastic differential equation of the form

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \qquad (11.0.1)$$

with a d-dimensional Brownian motion (B_t) and measurable coefficients $b : [0, \infty) \times U \to \mathbb{R}^n$ and $\sigma : [0, \infty) \times U \to \mathbb{R}^{n \times d}$. In applications one is often not interested in the random variables $X_t : \Omega \to \mathbb{R}$ themselves but only in their joint distribution. In that case, it is usually irrelevant w.r.t. which Brownian motion (B_t) the SDE (11.0.1) is satisfied. Therefore, we can "solve" the SDE in a very different way: Instead of constructing the solution from a **given** Brownian motion, we first construct a stochastic process (X_t, P) by different types of transformations or approximations, and then we verify that the process satisfies (11.0.1) w.r.t. **some** Brownian motion (B_t) that is usually **defined through (11.0.1)**.

Definition (Weak and strong solutions). *A **(weak) solution** of the stochastic differential equation (11.0.1) is given by*

*(i) a **"setup"** consisting of a probability space $(\Omega, \mathcal{A}, P)$, a filtration $(\mathcal{F}_t)_{t \geq 0}$ on $(\Omega, \mathcal{A})$ and an $\mathbb{R}^d$-valued $(\mathcal{F}_t)$ Brownian motion $(B_t)_{t \geq 0}$ on $(\Omega, \mathcal{A}, P)$,*

(ii) a continuous $(\mathcal{F}_t)$ adapted stochastic process $(X_t)_{t<S}$ where S is an $(\mathcal{F}_t)$ stopping time such that $b(\cdot, X) \in \mathcal{L}^1_{a,loc}([0,S), \mathbb{R}^n)$, $\sigma(\cdot, X) \in \mathcal{L}^2_{a,loc}([0,S), \mathbb{R}^{n\times d})$, and

$$X_t \quad = \quad X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \qquad \text{for any } t < S \text{ a.s.}$$

*It is called a **strong solution** w.r.t. the given setup if and only if (X_t) is adapted w.r.t. the filtration $\big(\sigma\big(\mathcal{F}_t^{B,P}, X_0\big)\big)_{t\geq 0}$ generated by the Brownian motion and the initial condition.*

Here $\mathcal{L}^q_{a,\mathrm{loc}}([0,S),\mathbb{R}^n)$ consists of all $\mathbb{R}^n$ valued processes $(\omega, t) \mapsto H_t(\omega)$ defined for $t < S(\omega)$ such that there exists an increasing sequence of $(\mathcal{F}_t)$ stopping times $T_n \uparrow S$ and $(\mathcal{F}_t)$ adapted processes $(H_t^{(n)})_{t\geq 0}$ in $\mathcal{L}^q(P \otimes \lambda_{(0,\infty)})$ with $H_t = H_t^{(n)}$ for any $t < T_n$ and $n \in \mathbb{N}$. *Note that the concept of a weak solution of an SDE is not related to the analytic concept of a weak solution of a PDE !*

Remark. A solution $(X_t)_{t\geq 0}$ is a strong solution up to $S = \infty$ w.r.t. a given setup if and only if there exists a measurable map $F : \mathbb{R}_+ \times \mathbb{R}^n \times C\big(\mathbb{R}_+, \mathbb{R}^d\big) \to \mathbb{R}^n$, $(t, x_0, y) \mapsto F_t(x_0, y)$, such that the process $(F_t)_{t\geq 0}$ is adapted w.r.t. the filtration $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{B}_t$, $\mathcal{B}_t = \sigma(y \mapsto y(s) : 0 \leq s \leq t)$, and

$$X_t \quad = \quad F_t(X_0, B) \qquad \text{for any } \ t \geq 0$$

holds almost surely. Hence strong solutions are (almost surely) functions of the **given** Brownian motion and the initial value!

There are SDE that have weak but no strong solutions. An example is given in Section 11.1. The definition of weak and strong solutions can be generalized to other types of SDE including in particular functional equations of the form

$$dX_t \quad = \quad b_t(X)\, dt + \sigma_t(X)\, dB_t$$

where (b_t) and (σ_t) are $(\mathcal{B}_t)$ adapted stochastic processes defined on $C(\mathbb{R}_+, \mathbb{R}^n)$, as well as SDE driven by Poisson point processes, cf. Chapter 13.

Different types of transformations of a stochastic process (X_t, P) are useful for constructing weak solutions. These include:

• **Random time changes:** $(X_t)_{t\geq 0} \to (X_{T_a})_{a\geq 0}$ where $(T_a)_{a\geq 0}$ is an increasing stochastic process on $\mathbb{R}_+$ such that T_a is a stopping time for any $a \geq 0$.

• **Transformations of the paths in space:** These include for example coordinate changes $(X_t) \to (\phi(X_t))$, random translations $(X_t) \to (X_t + H_t)$ where (H_t) is another adapted process, and, more generally, a transformation that maps (X_t) to the strong solution (Y_t) of an SDE driven by (X_t).

• **Change of measure:** Here the random variables X_t are kept fixed but the underlying probability measure P is replaced by a new measure $\widetilde{P}$ such that both measures are mutually absolutely continuous on each of the σ-algebras $\mathcal{F}_t$, $t \in \mathbb{R}_+$ (but usually not on $\mathcal{F}_\infty$).

In Sections 11.2, 11.3 and 11.4, we study these transformations as well as relations between them. For identifying the transformed processes, the Lévy characterizations in Section 11.1 play a crucial rôle. Section 11.5 contains an application to large deviations on Wiener space, and, more generally, random perturbations of dynamical systems. Section 12.2 focusses on Stratonovich differential equations. As the Stratonovich integral satisfies the usual chain rule, these are adequate for studying stochastic processes on Riemannian manifolds. Stratonovich calculus also leads to a tranformation of an SDE in terms of the flow of a corresponding ODE that is useful for example in the one-dimensional case. The concluding Section 12.4 considers numerical approximation schemes for solutions of stochastic differential equations.

11.1 Lévy characterizations and martingale problems

Let $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ be a given filtered probability space. We first note that Lévy processes can be characterized by their exponential martingales:

Lemma 11.1. *Let $\psi : \mathbb{R}^d \to \mathbb{C}$ be a given function. An $(\mathcal{F}_t)$ adapted càdlàg process $X_t : \Omega \to \mathbb{R}^d$ is an $(\mathcal{F}_t)$ Lévy process with characteristic exponent ψ if and only if the complex-valued processes*

$$Z_t^p \quad := \quad \exp\big(ip \cdot X_t + t\psi(p)\big) \qquad , \quad t \geq 0,$$

are $(\mathcal{F}_t)$ martingales, or, equivalently, local $(\mathcal{F}_t)$ martingales for any $p \in \mathbb{R}^d$.

Proof. By Corollary 10.5, the processes Z^p are martingales if X is a Lévy process with characteristic exponent ψ. Conversely, suppose that Z^p is a local martingale for any $p \in R^d$. Then, since these processes are uniformly bounded on finite time intervals, they are martingales. Hence for $0 \leq s \leq t$ and $p \in \mathbb{R}^d$,

$$E\big[\exp\big(ip \cdot (X_t - X_s)\big)\big|\mathcal{F}_s\big] \quad = \quad \exp(-(t-s)\psi(p)),$$

which implies that $X_t - X_s$ is independent of $\mathcal{F}_s$ with characteristic function equal to $\exp(-(t-s)\psi)$. □

Exercise (Characterization of Poisson point processes). Let $(S, \mathcal{S}, \nu)$ be a σ-finite measure space. Suppose that $(N_t)_{t\geq 0}$ on $(\Omega, \mathcal{A}, P)$ is an $(\mathcal{F}_t)$ adapted process taking values in the space $M_c^+(S)$ consisting of all counting measures on S. Prove that the following statements are equivalent:

(i) (N_t) is a Poisson point processes with intensity measure ν.

(ii) For any function $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$, the real valued process

$$N_t(f) \quad = \quad \int f(y)\, N_t\,(dy), \qquad t \geq 0,$$

is a compound Poisson process with jump intensity measure $\mu \circ f^{-1}$.

(iii) For any function $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$, the complex valued process

$$M_t^{[f]} \quad = \quad \exp(iN_t(f) + t\psi(f)), \qquad t \geq 0, \quad \psi(f) \quad = \quad \int \big(1 - e^{if}\big)\, d\nu,$$

is a local $(\mathcal{F}_t)$ martingale.

Show that the statements are also equivalent if only elementary functions $f \in L^1(S, \mathcal{S}, \nu)$ are considered.

Lévy's characterization of Brownian motion

By Lemma 11.1, an $\mathbb{R}^d$-valued process (X_t) is a Brownian motion if and only if the processes $\exp\big(ip \cdot X_t + t|p|^2/2\big)$ are local martingales for all $p \in \mathbb{R}^d$. This can be applied to prove the remarkable fact that any continuous $\mathbb{R}^d$ valued martingale with the right covariations is a Brownian motion:

Theorem 11.2 (Lévy's characterization for multidimensional Brownian motion). *Suppose that $M^1, \dots, M^d$ are continuous local $(\mathcal{F}_t)$ martingales with*

$$[M^k, M^l]_t \quad = \quad \delta_{kl} t \qquad P\text{-a.s. for any } t \ge 0.$$

Then $M = (M^1, \dots, M^d)$ is a d-dimensional Brownian motion.

The following proof is due to Kunita and Watanabe (1967):

Proof. Fix $p \in \mathbb{R}^d$ and let $\Phi_t := \exp(ip \cdot M_t)$. By Itô's formula,

$$\begin{aligned} d\Phi_t \quad &= \quad ip\, \Phi_t \cdot dM_t - \frac{1}{2} \sum_{k,l=1}^d \Phi_t\, p_k p_l\, d[M^k, M^l]_t \\ &= \quad ip\, \Phi_t \cdot dM_t - \frac{1}{2} \Phi_t\, |p|^2\, dt. \end{aligned}$$

Since the first term on the right hand side is a local martingale increment, the product rule shows that the process $\Phi_t \cdot \exp(|p|^2\, t/2)$ is a local martingale. Hence by Lemma 11.1, M is a Brownian motion. □

Lévy's characterization of Brownian motion has a lot of remarkable direct consequences.

Example (Random orthogonal transformations). Suppose that $X_t : \Omega \to \mathbb{R}^n$ is a solution of an SDE

$$dX_t \quad = \quad O_t\, dB_t, \qquad X_0 = x_0, \tag{11.1.1}$$

w.r.t. a d-dimensional Brownian motion (B_t), a product-measurable adapted process $(t, \omega) \mapsto O_t(\omega)$ taking values in $\mathbb{R}^{n \times d}$, and an initial vale $x_0 \in \mathbb{R}^n$. We verify that X is an n-dimensional Brownian motion provided

$$O_t(\omega)\, O_t(\omega)^T \quad = \quad I_n \qquad \text{for any } t \ge 0, \quad \text{almost surely.} \tag{11.1.2}$$

Indeed, by (11.1.1) and (11.1.2), the components

$$X_t^i \quad = \quad x_0^i + \sum_{k=1}^d \int_0^t O_s^{ik}\, dB_s^k$$

are continuous local martingales with covariations

$$[X^i, X^j] \quad = \quad \sum_{k,l} \int O^{ik}\, O^{jl}\, d[B^k, B^l] \quad = \quad \int \sum_k O^{ik}\, O^{jk}\, dt \quad = \quad \delta_{ij}\, dt.$$

Applications include infinitesimal random rotations ($n = d$) and random orthogonal projections ($n < d$). The next example is a special application.

Example (Bessel process). We derive an SDE for the radial component $R_t = |B_t|$ of Brownian motion in $\mathbb{R}^d$. The function $r(x) = |x|$ is smooth on $\mathbb{R}^d \setminus \{0\}$ with $\nabla r(x) = e_r(x)$, and $\Delta r(x) = (d-1) \cdot |x|^{-1}$ where $e_r(x) = x/|x|$. Applying Itô's formula to functions $r_\varepsilon \in C^\infty(\mathbb{R}^d)$, $\varepsilon > 0$, with $r_\varepsilon(x) = r(x)$ for $|x| \geq \varepsilon$ yields

$$dR_t \quad = \quad e_r(B_t) \cdot dB_t \, + \, \frac{d-1}{2R_t}\, dt \qquad \text{for any } \; t < T_0$$

where T_0 is the first hitting time of 0 for (B_t). By the last example, the process

$$W_t \quad := \quad \int_0^t e_r(B_s) \cdot dB_s, \qquad t \geq 0,$$

is a one-dimensional Brownian motion defined for all times (the value of e_r at 0 being irrelevant for the stochastic integral). Hence (B_t) is a weak solution of the SDE

$$dR_t \quad = \quad dW_t \, + \, \frac{d-1}{2R_t}\, dt \tag{11.1.3}$$

up to the first hitting time of 0. The equation (11.1.3) makes sense for any particular $d \in \mathbb{R}$ and is called the **Bessel equation**. Much more on Bessel processes can be found in Revuz and Yor [37] and other works by M. Yor.

Exercise (Exit times and ruin probabilities for Bessel and compound Poisson processes). a) Let (X_t) be a solution of the Bessel equation

$$dX_t \; = \; -\frac{d-1}{2X_t}\, dt \, + \, dB_t, \qquad X_0 = x_0,$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion and d is a real constant.

i) Find a non-constant function $u : \mathbb{R} \to \mathbb{R}$ such that $u(X_t)$ is a local martingale up to the first hitting time of 0.

ii) Compute the ruin probabilities $P[T_a < T_b]$ for $a, b \in \mathbb{R}_+$ with $x_0 \in [a, b]$.

iii) Proceeding similarly, determine the mean exit time $E[T]$, where $T = \min\{T_a, T_b\}$.

b) Now let $(X_t)_{t\geq 0}$ be a compound Poisson process with $X_0 = 0$ and jump intensity measure $\nu = N(m, 1),\ m > 0$.

i) Determine $\lambda \in \mathbb{R}$ such that $\exp(\lambda X_t)$ is a local martingale up to T_0.

ii) Prove that for $a < 0$,

$$P[T_a < \infty] \ = \ \lim_{b\to\infty} P[T_a < T_b] \ \leq \ \exp(ma/2).$$

Why is it not as easy as above to compute the ruin probability $P[T_a < T_b]$ exactly ?

The next application of Lévy's characterization of Brownian motion shows that there are SDE that have weak but no strong solutions.

Example (Tanaka's example. Weak vs. strong solutions). Consider the one dimensional SDE

$$dX_t \quad = \quad \operatorname{sgn}(X_t)\, dB_t \tag{11.1.4}$$

where (B_t) is a Brownian motion and $\operatorname{sgn}(x) := \begin{cases} +1 & \text{for } x \geq 0, \\ -1 & \text{for } x < 0 \end{cases}$. Note the unusual convention $\operatorname{sgn}(0) = 1$ that is used below. We prove the following statements:

1) X is a weak solution of (11.1.4) on $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ if and only if X is an $(\mathcal{F}_t)$ Brownian motion. In particular, *a weak solution exists* and its *law is uniquely determined* by the law of the initial value X_0.

2) If X is a weak solution w.r.t. a setup $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$ then for any $t \geq 0$, $B_t - B_0$ is measurable w.r.t. the σ-algebra $\mathcal{F}_t^{|X|,P} = \sigma(|X_s| : s \leq t)^P$.

3) There is *no strong solution* to (11.1.4) with initial condition $X_0 = 0$.

4) *Pathwise uniqueness does not hold:* If X is a solution to (11.1.4) with $X_0 = 0$ then $-X$ solves the same equation with the same Brownian motion.

The proof of 1) is again a consequence of the first example above: If X is a weak solution then X is a Brownian motion by Lévy's characterization. Conversely, if X is an $(\mathcal{F}_t)$ Brownian motion then the process

$$B_t \quad := \quad \int_0^t \operatorname{sgn}(X_s)\, dX_s$$

is a Brownian motion as well, and

$$\int_0^t \operatorname{sgn}(X_s)\, dB_s \quad = \quad \int_0^t \operatorname{sgn}(X_s)^2\, dX_s \quad = \quad X_t - X_0,$$

i.e., X is a weak solution to (11.1.4).

For proving 2) , we approximate $r(x) = |x|$ by symmetric and concave functions $r_\varepsilon \in C^\infty(\mathbb{R})$ satisfying $r_\varepsilon(x) = |x|$ for $|x| \geq \varepsilon$. Then the associative law, the Itô isometry, and Itô's formula imply

$$\begin{aligned} B_t - B_0 \quad &= \quad \int_0^t \operatorname{sgn}(X_s)\, dX_s \quad = \quad \lim_{\varepsilon \downarrow 0} \int_0^t r_\varepsilon''(X_s)\, dX_s \\ &= \quad \lim_{\varepsilon \downarrow 0} \Big(r_\varepsilon(X_t) - r_\varepsilon(X_0) - \frac{1}{2} \int_0^t r_\varepsilon''(X_s)\, ds \Big) \\ &= \quad \lim_{\varepsilon \downarrow 0} \Big(r_\varepsilon(|X_t|) - r_\varepsilon(|X_0|) - \frac{1}{2} \int_0^t r_\varepsilon''(|X_s|)\, ds \Big) \end{aligned}$$

with almost sure convergence along a subsequence $\varepsilon_n \downarrow 0$.

Finally by 2), if X would be a strong solution w.r.t. a Brownian motion B then X_t would also be measurable w.r.t. the σ-algebra generated by $\mathcal{F}_0$ and $\mathcal{F}_t^{|X|,P}$. This leads to a contradiction as one can verify that the event $\{X_t \geq 0\}$ is not measurable w.r.t. this σ-algebra for a Brownian motion (X_t).

Martingale problem for Itô diffusions

Next we consider a solution of a stochastic differential equation

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \qquad X_0 = x_0, \tag{11.1.5}$$

defined on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$. We assume that (B_t) is an $(\mathcal{F}_t)$ Brownian motion taking values in $\mathbb{R}^d, b, \sigma_1, \ldots, \sigma_d : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^n$ are measurable and locally bounded (i.e., bounded on $[0,t] \times K$ for any $t \geq 0$ and any compact set $K \subset \mathbb{R}^d$) time-dependent vector fields, and $\sigma(t,x) = (\sigma_1(t,x) \cdots \sigma_d(t,x))$ is the $n \times d$ matrix with column vectors $\sigma_i(t,x)$. A solution of (11.1.5) is a continuous $(\mathcal{F}_t^P)$ semimartingale (X_t) satisfying

$$X_t \quad = \quad x_0 + \int_0^t b(s, X_s)\, ds + \sum_{k=1}^d \int_0^t \sigma_k(s, X_s)\, dB_s^k \qquad \forall\, t \geq 0 \;\; a.s. \qquad (11.1.6)$$

If X is a solution then

$$\begin{aligned} [X^i, X^j]_t &= \sum_{k,l} \Big[\int \sigma_k^i(s,X)\, dB^k, \int \sigma_l^j(s,X)\, dB^l \Big]_t \\ &= \sum_{k,l} \int_0^t (\sigma_k^i\, \sigma_l^j)(s,X)\, d[B^k, B^l] \quad = \quad \int_0^t a^{ij}(s, X_s)\, ds \end{aligned}$$

where $a^{ij} = \sum_k \sigma_k^i \sigma_k^j$, i.e.,

$$a(s,x) = \sigma(s,x)\sigma(s,x)^T \in \mathbb{R}^{n \times n}.$$

Therefore, Itô's formula applied to the process (t, X_t) yields

$$\begin{aligned} dF(t,X) &= \frac{\partial F}{\partial t}(t,X)\, dt + \nabla_x F(t,X) \cdot dX + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 F}{\partial x^i \partial x^j}(t,X)\, d[X^i, X^j] \\ &= (\sigma^T \nabla_x F)(t,X) \cdot dB + \Big(\frac{\partial F}{\partial t} + \mathcal{L}F \Big)(t,X)\, dt, \end{aligned}$$

for any $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$, where

$$(\mathcal{L}F)(t,x) \quad = \quad \frac{1}{2} \sum_{i,j=1}^d a^{ij}(t,x) \frac{\partial^2 F}{\partial x^i \partial x^j}(t,x) + \sum_{i=1}^d b^i(t,x) \frac{\partial F}{\partial x^i}(t,x).$$

We have thus derived the **Itô-Doeblin formula**

$$\boxed{F(t, X_t) - F(0, X_0) \quad = \quad \int_0^t (\sigma^T \nabla F)(s, X_s) \cdot dB_s + \int_0^t \big(\frac{\partial F}{\partial t} + \mathcal{L}F \big)(s, X_s)\, ds} \qquad (11.1.7)$$

The formula provides a semimartingale decomposition for $F(t, X_t)$. It establishes a connection between the stochastic differential equation (11.1.5) and partial differential equations involving the operator $\mathcal{L}$.

Example (Exit distributions and boundary value problems). Suppose that $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$ is a classical solution of the p.d.e.

$$\frac{\partial F}{\partial t}(t,x) + (\mathcal{L}F)(t,x) \quad = \quad -g(t,x) \qquad \forall\, t \geq 0,\, x \in U$$

on an open subset $U \subset \mathbb{R}^n$ with boundary values

$$F(t,x) \quad = \quad \phi(t,x) \qquad \forall\, t \geq 0,\, x \in \partial U.$$

Then by (11.1.7), the process

$$M_t \quad = \quad F(t, X_t) + \int_0^t g(s, X_s)\, ds$$

is a local martingale. If F and g are bounded on $[0,t] \times \overline{U}$, then the process M^T stopped at the first exit time $T = \inf\{t \geq 0 \;:\; X_t \notin U\}$ is a martingale. Hence, if T is almost surely finite then

$$E[\phi(T, X_T)] + E\Big[\int_0^T g(s, X_s)\, ds\Big] \quad = \quad F(0, x_0).$$

This can be used, for example, to compute exit distributions (for $g \equiv 0$) and mean exit times (for $\phi \equiv 0$, $g \equiv 1$) analytically or numerically.

Similarly as in the example, the Feynman-Kac-formula and other connections between Brownian motion and the Laplace operator carry over to Itô diffusions and their generator $\mathcal{L}$ in a straightforward way. Of course, the resulting partial differential equation usually can not be solved analytically, but there is a wide range of well-established numerical methods for linear PDE available for explicit computations of expectation values.

Exercise (Feynman-Kac formula for Itô diffusions). Fix $t \in (0,\infty)$, and suppose that $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $V : [0,t] \times \mathbb{R}^n \to [0,\infty)$ are continuous functions. Show that if $u \in C^2((0,t] \times \mathbb{R}^n) \cap C([0,t] \times \mathbb{R}^n)$ is a bounded solution of the heat equation

$$\begin{aligned} \frac{\partial u}{\partial s}(s,x) \quad &= \quad (\mathcal{L}u)(s,x) - V(s,x)u(s,x) \qquad \text{for } s \in (0,t],\ x \in \mathbb{R}^n, \\ u(0,x) \quad &= \quad \varphi(x), \end{aligned}$$

then u has the stochastic representation

$$u(t,x) \;=\; E_x\left[\varphi(X_t)\,\exp\left(-\int_0^t V(t-s,X_s)\,ds\right)\right].$$

Hint: Consider the time reversal $\hat{u}(s,x) := u(t-s,x)$ *of* u *on* $[0,t]$. *Show first that* $M_r := \exp(-A_r)\hat{u}(r,X_r)$ *is a local martingale if* $A_r := \int_0^r \hat{V}(s,X_s)\,ds$.

Often, the solution of an SDE is only defined up to some explosion time ζ where it diverges or exits a given domain. By localization, we can apply the results above in this case as well. Indeed, suppose that $U \subseteq \mathbb{R}^n$ is an open set, and let

$$U_k \quad = \quad \{x \in U \;:\; |x| < k \text{ and } \operatorname{dist}(x,U^c) > 1/k\}\,, \qquad k \in \mathbb{N}.$$

Then $U = \bigcup U_k$. Let T_k denote the first exit time of (X_t) from U_k. A solution (X_t) of the SDE (11.1.5) up to the explosion time $\zeta = \sup T_k$ is a process $(X_t)_{t\in[0,\zeta)\cup\{0\}}$ such that for every $k \in \mathbb{N}$, $T_k < \zeta$ almost surely on $\{\zeta \in (0,\infty)\}$, and the stopped process X^{T_k} is a semimartingale satisfying (11.1.6) for $t \le T_k$. By applying Itô's formula to the stopped processes, we obtain:

Theorem 11.3 (Martingale problem for Itô diffusions). *If* $X_t : \Omega \to U$ *is a solution of (11.1.5) up to the explosion time* ζ*, then for any* $F \in C^2(\mathbb{R}_+ \times U)$ *and* $x_0 \in U$*, the process*

$$M_t \quad := \quad F(t,X_t) - \int_0^t \left(\frac{\partial F}{\partial t} + \mathcal{L}F\right)(s,X_s)\,ds, \qquad t < \zeta,$$

is a local martingale up to the explosion time ζ*, and the stopped processes* M^{T_k}*,* $k \in \mathbb{N}$*, are localizing martingales.*

Proof. We can choose functions $F_k \in \mathcal{C}_b^2([0,a]\times U)$, $k \in \mathbb{N}$, $a \ge 0$, such that $F_k(t,x) = F(t,x)$ for $t \in [0,a]$ and x in a neighbourhood of $\overline{U}_k$. Then for $t \le a$,

$$M_t^{T_k} \quad = \quad M_{t\wedge T_k} \quad = \quad F_k(t,X_{t\wedge T_k}) - \int_0^t \left(\frac{\partial F_k}{\partial t} + \mathcal{L}F_k\right)(s,X_{s\wedge T_k})\,ds.$$

By (11.1.7), the right hand side is a bounded martingale. □

Lévy characterization of weak solutions

Lévy's characterization of Brownian motion can be extended to solutions of stochastic differential equations of type

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{11.1.8}$$

driven by a d-dimensional Brownian motion (B_t). As a consequence, one can show that a process is a weak solution of (11.1.8) if and only if it solves the corresponding martingale problem. As above, we assume that the coefficients $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are measurable and locally bounded, and we set

$$\mathcal{L} \quad = \quad \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(t,x)\frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^i(t,x)\frac{\partial}{\partial x^i} \tag{11.1.9}$$

where $a(t, x) = \sigma(t, x)\sigma(t, x)^T$.

Theorem 11.4 (Weak solutions and the martingale problem). *If the matrix $\sigma(t, x)$ is invertible for any t and x, and $(t, x) \mapsto \sigma(t, x)^{-1}$ is a locally bounded function on $\mathbb{R}_+ \times \mathbb{R}^d$, then the following statements are equivalent:*

(i) (X_t) *is a weak solution of (11.1.8) on the setup* $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$.

(ii) The processes $M^i_t := X^i_t - X^i_0 - \int_0^t b^i(s, X_s)\, ds$, $1 \le i \le d$, *are continuous local* $(\mathcal{F}^P_t)$ *martingales with covariations*

$$[M^i, M^j]_t \quad = \quad \int_0^t a^{ij}(s, X_s)\, ds \qquad P\text{-a.s. for any } t \ge 0. \tag{11.1.10}$$

(iii) The processes $M^{[f]}_t := f(X_t) - f(X_0) - \int_0^t (\mathcal{L}f)(s, X_s)\, ds$, $f \in C^2(\mathbb{R}^d)$, *are continuous local* $(\mathcal{F}^P_t)$ *martingales.*

(iv) The processes $\hat{M}^{[f]}_t := f(t, X_t) - f(0, X_0) - \int_0^t \left(\frac{\partial f}{\partial t} + \mathcal{L}f\right)(s, X_s)\, ds$, $f \in C^2(\mathbb{R}_+ \times \mathbb{R}^d)$, *are continuous local* $(\mathcal{F}^P_t)$ *martingales.*

Proof. (i)⇒(iv) is a consequence of the Itô-Doeblin formula, cf. Theorem 11.3 above.
(iv)⇒(iii) trivially holds.
(iii)⇒(ii) follows by choosing for f polynomials of degree ≥ 2. Indeed, for $f(x) = x^i$, we obtain $\mathcal{L}f = b^i$, hence

$$M_t^i \quad = \quad X_t^i - X_t^0 - \int_0^t b^i(s, X_s)\, ds \quad = \quad M_t^{[f]} \tag{11.1.11}$$

is a local martingale by (iii). Moreover, if $f(x) = x^i x^j$ then $\mathcal{L}f = a^{ij} + x^i b^j + x^j b^i$ by the symmetry of a, and hence

$$X_t^i X_t^j - X_0^i X_0^j \;=\; M_t^{[f]} + \int_0^t \left(a^{ij}(s, X_s) + X_s^i\, b^j(s, X_s) + X_s^j\, b^i(s, X_s)\right) ds. \tag{11.1.12}$$

On the other hand, by the product rule and (11.1.11),

$$\begin{aligned} X_t^i X_t^j - X_0^i X_0^j \quad &= \quad \int_0^t X_s^i\, dX_s^j + \int_0^t X_s^j\, dX_s^i + [X^i, X^j]_t \qquad (11.1.13) \\ &= \quad N_t + \int_0^t \left(X_s^i\, b^j(s, X_s) + X_s^j\, b^i(s, X_s)\right) ds + [X^i, X^j]_t \end{aligned}$$

with a continuous local martingale N. Comparing (11.1.12) and (11.1.13) we obtain

$$[M^i, M^j]_t \quad = \quad [X^i, X^j]_t \quad = \quad \int_0^t a^{ij}(s, X_s)\, ds$$

since a continuous local martingale of finite variation is constant.

(ii)⇒(i) is a consequence of Lévy's characterization of Brownian motion: If (ii) holds then

$$dX_t \quad = \quad dM_t + b(t, X_t)\, dt \quad = \quad \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt$$

where $M_t = \left(M_t^1, \ldots, M_t^d\right)$ and $B_t := \int_0^t \sigma(s, X_s)^{-1}\, dM_s$ are continuous local martingales with values in $\mathbb{R}^d$ because σ^{-1} is locally bounded. To identify B as a Brownian motion it suffices to note that

$$\begin{aligned} [B^k, B^l]_t \quad &= \quad \int_0^t \sum_{i,j} \left(\sigma_{ki}^{-1}\, \sigma_{lj}^{-1}\right)(s, X_s)\, d[M^i, M^j] \\ &= \quad \int_0^t \left(\sigma^{-1} a (\sigma^{-1})^T\right)_{kl} (s, X_s)\, ds \quad = \quad \delta_{kl}\, t \end{aligned}$$

for any $k, l = 1, \ldots, d$ by (11.1.10). □

Remark (Degenerate case). If $\sigma(t,x)$ is degenerate then a corresponding assertion still holds. However, in this case the Brownian motion (B_t) only exists on an extension of the probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$. The reason is that in the degenerate case, the Brownian motion can not be recovered directly from the solution (X_t) as in the proof above, see [38] for details.

The martingale problem formulation of weak solutions is powerful in many respects: It is stable under weak convergence and therefore well suited for approximation arguments, it carries over to more general state spaces (including for example Riemannian manifolds, Banach spaces, spaces of measures), and, of course, it provides a direct link to the theory of Markov processes. Do not miss to have a look at the classics by Stroock and Varadhan [40] and by Ethier and Kurtz [16] for much more on the martingale problem and its applications to Markov processes.

11.2 Random time change

Random time change is already central to the work of Doeblin from 1940 that has been discovered only recently [3]. Independently, Dambis and Dubins-Schwarz have developed a theory of random time changes for semimartingales in the 1960s [25], [37]. In this section we study random time changes with a focus on applications to SDE, in particular, but not exclusively, in dimension one.

Throughout this section we fix a **right-continuous filtration** $(\mathcal{F}_t)$ such that $\mathcal{F}_t = \mathcal{F}^P$ for any $t \geq 0$. Right-continuity is required to ensure that the time transformation is given by $(\mathcal{F}_t)$ stopping times.

Continuous local martingales as time-changed Brownian motions

Let $(M_t)_{t\geq 0}$ be a continuous local $(\mathcal{F}_t)$ martingale w.r.t. the underlying probability measure P such that $M_0 = 0$. Our aim is to show that M_t can be represented as $B_{[M]_t}$ with a one-dimensional Brownian motion (B_a). For this purpose, we consider the random time substitution $a \mapsto T_a$ where $T_a = \inf\{u : [M]_u > a\}$is the first passage time to the level u. Note that $a \mapsto T_a$ is the **right inverse** of the quadratic variation $t \mapsto [M]_t$, i.e.,

$$[M]_{T_A} \quad = \quad a \qquad \text{on} \quad \{T_a \; < \; \infty\}\,, \quad \text{and,}$$

$$T_{[M]_t} \quad = \quad \inf\{u : [M]_u > [M]_t\} \quad = \quad \sup\{u : [M]_u \; = \; [M]_t\}$$

by continuity of $[M]$. If $[M]$ is strictly increasing then $T = [M]^{-1}$. By right-continuity of $(\mathcal{F}_t)$, T_a is an $(\mathcal{F}_t)$ stopping time for any $a \geq 0$.

Theorem 11.5 (Dambis, Dubins-Schwarz). *If M is a continuous local $(\mathcal{F}_t)$ martingale with $[M]_\infty = \infty$ almost surely then the time-changed process $B_a := M_{T_a}$, $a \geq 0$, is an $(\mathcal{F}_{T_a})$ Brownian motion, and*

$$M_t \quad = \quad B_{[M]_t} \qquad \textit{for any } t \geq 0, \quad \textit{almost surely.} \tag{11.2.1}$$

The proof is again based on Lévy's characterization.

Proof. 1) We first note that $B_{[M]_t} = M_t$ almost surely. Indeed, by definition, $B_{[M]_t} = M_{T_{[M]_t}}$. It remains to verify that M is almost surely constant on the interval $[t, T_{[M]_t}]$. This holds true since the quadratic variation $[M]$ is constant on this interval, cf. the exercise below.

2) Next, we verify that $B_a = M_{T_a}$ is almost surely continuous. Right-continuity holds since M and T are both right-continuous. To prove left-continuity note that for $a > 0$,

$$\lim_{\varepsilon \downarrow 0} M_{T_{a-\varepsilon}} \quad = \quad M_{T_{a-}} \qquad \text{for any } a \geq 0$$

by continuity of M. It remains to show $M_{T_{a-}} = M_{T_a}$ almost surely. This again holds true by the exercise below, because T_{a-} and T_a are stopping times, and

$$[M]_{T_{a-}} \quad = \quad \lim_{\varepsilon \downarrow 0} [M]_{T_{a-\varepsilon}} \quad = \quad \lim_{\varepsilon \downarrow 0} (a - \varepsilon) \quad = \quad a \quad = \quad [M]_{T_a}$$

by continuity of $[M]$.

3) We now show that (B_a) is a square-integrable $(\mathcal{F}_{T_a})$ martingale. Since the random variables T_a are $(\mathcal{F}_t)$ stopping times, (B_a) is $(\mathcal{F}_{T_a})$ adapted. Moreover, for any a, the stopped process $M_t^{T_a} = M_{t\wedge T_a}$ is a continuous local martingale with

$$E\big[[M^{T_a}]_\infty\big] \quad = \quad E\big[[M]_{T_a}\big] \quad = \quad a \quad < \quad \infty.$$

Hence M^{T_a} is in $M_c^2\big([0,\infty]\big)$, and

$$E[B_a^2] \quad = \quad E[M_{T_a}^2] \quad = \quad E[(M_\infty^{T_a})^2] \quad = a \qquad \text{for any } \ a \geq 0.$$

This shows that (B_a) is square-integrable, and, moreover,

$$E[B_a|\mathcal{F}_{T_r}] \quad = \quad E[M_{T_a}|\mathcal{F}_{T_r}] \quad = \quad M_{T_r} \quad = \quad B_r \qquad \text{for any } \ 0 \leq r \leq a$$

by the Optional Sampling Theorem applied to M^{T_a}.

Finally, we note that $[B]_a = \langle B\rangle_a = a$ almost surely. Indeed, by the Optional Sampling Theorem applied to the martingale $(M^{T_a})^2 - [M^{T_a}]$, we have

$$\begin{aligned} E\big[B_a^2 - B_r^2|\mathcal{F}_{T_r}\big] \quad &= \quad E\big[M_{T_a}^2 - M_{T_r}^2|\mathcal{F}_{T_r}\big] \\ &= \quad E\big[[M]_{T_a} - [M]_{T_r}|\mathcal{F}_{T_r}\big] \quad = \quad a - r \quad \text{for } \ 0 \leq r \leq a. \end{aligned}$$

Hence $B_a^2 - a$ is a martingale, and thus by continuity, $[B]_a = \langle B\rangle_a = a$ almost surely.

We have shown that (B_a) is a continuous square-integrable $(\mathcal{F}_{T_a})$ martingale with $[B]_a = a$ almost surely. Hence B is a Brownian motion by Lévy's characterization. □

Remark. The assumption $[M]_\infty = \infty$ in Theorem 11.5 ensures $T_a < \infty$ almost surely. If the assumption is violated then M can still be represented in the form (11.2.1) with a Brownian motion B. However, in this case, B is only defined on an extended probability space and can not be obtained as a time-change of M for all times, cf. e.g. [37].

Exercise. Let M be a continuous local $(\mathcal{F}_t)$ martingale, and let S and T be $(\mathcal{F}_t)$ stopping times such that $S \leq T$. Prove that if $[M]_S = [M]_T < \infty$ almost surely, then M is almost surely constant on the stochastic interval $[S,T]$. Use this fact to complete the missing step in the proof above.

We now consider several applications of Theorem 11.5. Let $(W_t)_{t\geq 0}$ be a Brownian motion with values in $\mathbb{R}^d$ w.r.t. the underlying probability measure P.

Time-change representations of stochastic integrals

By Theorem 11.5 and the remark below the theorem, stochastic integrals w.r.t. Brownian motions are time-changed Brownian motions. For any integrand $G \in \mathcal{L}^2_{a,loc}(\mathbb{R}_+, \mathbb{R}^d)$, there exists a one-dimensional Brownian motion B, possibly defined on an enlarged probability space, such that almost surely,

$$\int_0^t G_s \cdot dW_s \quad = \quad B_{\int_0^t |G_s|^2\, ds} \qquad \text{for any } \ t \geq 0.$$

Example (Gaussian martingales). If G is a deterministic function then the stochastic integral is a Gaussian process that is obtained from the Brownian motion B by a deterministic time substitution. This case has already been studied in Section 8.3 in [14].

Doeblin [3] has developed a stochastic calculus based on time substitutions instead of Itô integrals. For example, an SDE in $\mathbb{R}^1$ of type

$$X_t - X_0 \quad = \quad \int_0^t \sigma(s, X_s)\, dW_s + \int_0^t b(s, X_s)\, ds$$

can be rephrased in the form

$$X_t - X_0 \quad = \quad B_{\int_0^t \sigma(s,X_s)^2\, ds} + \int_0^t b(s, X_s)\, ds$$

with a Brownian motion B. The one-dimensional Itô-Doeblin formula then takes the form

$$f(t, X_t) - f(0, X_0) \quad = \quad B_{\int_0^t \sigma(s,X_s)^2\, f'(s,X_s)^2\, ds} + \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{L}f \right)(s, X_s)\, ds$$

with $\mathcal{L}f = \frac{1}{2}\, \sigma^2 f'' + bf'$.

Time substitution in stochastic differential equations

To see how time substitution can be used to construct weak solutions, we consider at first an SDE of type

$$dY_t \quad = \quad \sigma(Y_t)\, dB_t \tag{11.2.2}$$

in $\mathbb{R}^1$ where $\sigma : \mathbb{R} \to (0, \infty)$ is a strictly positive continuous function. If Y is a weak solution then by Theorem 11.5 and the remark below,

$$Y_t \quad = \quad X_{A_t} \qquad \text{with} \quad A_t = [Y]_t = \int_0^t \sigma(Y_r)^2 \, dr \tag{11.2.3}$$

and a Brownian motion X. Note that A depends on Y, so at first glace (11.2.3) seems not to be useful for solving the SDE (11.2.2). However, the inverse time substitution $T = A^{-1}$ satisfies

$$T' \quad = \quad \frac{1}{A' \circ T} \quad = \quad \frac{1}{\sigma(Y \circ T)^2} \quad = \quad \frac{1}{\sigma(X)^2},$$

and hence

$$T_a \quad = \quad \int_0^a \frac{1}{\sigma(X_u)^2} \, du.$$

Therefore, we can construct a weak solution Y of (11.2.2) from a given Brownian motion X by first computing T, then the inverse function $A = T^{-1}$, and finally setting $Y = X \circ A$. More generally, the following result holds:

Theorem 11.6. *Suppose that (X_a) on $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ is a weak solution of an SDE of the form*

$$dX_a \quad = \quad \sigma(X_a) \, dB_a + b(X_a) \, da \tag{11.2.4}$$

with locally bounded measurable coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ such that $\sigma(x)$ is invertible for almost all x, and σ^{-1} is again locally bounded. Let $\varrho : \mathbb{R}^d \to (0, \infty)$ be a measurable function such that almost surely,

$$T_a \quad := \quad \int_0^a \varrho(X_u) \, du \quad < \quad \infty \qquad \forall a \in (0, \infty), \quad \text{and} \quad T_\infty = \infty. \tag{11.2.5}$$

Then the time-changed process defined by

$$Y_t \quad := \quad X_{A_t}, \qquad A \quad := \quad T^{-1},$$

is a weak solution of the SDE

$$dY_t \quad = \quad \left(\frac{\sigma}{\sqrt{\varrho}}\right)(Y_t) \, dB_t + \left(\frac{b}{\varrho}\right)(Y_t) \, dt. \tag{11.2.6}$$

We only give a sketch of the proof of the theorem:

Proof of 11.6. (Sketch). The process X is a solution of the martingale problem for the operator $\mathcal{L} = \frac{1}{2}\sum a_{ij}(x)\,\frac{\partial^2}{\partial x^i \partial x^j} + b(x)\cdot\nabla$ where $a = \sigma\sigma^T$, i.e.,

$$M_a^{[f]} \quad = \quad f(X_a) - F(X_0) - \int_0^a (\mathcal{L}f)(X_u)\,du$$

is a local $(\mathcal{F}_a)$ martingale for any $f \in C^2$. Therefore, the time-changed process

$$\begin{aligned} M_{A_t}^{[f]} \quad &= \quad f(X_{A_t} - f(X_{A_0}) - \int_0^{A_t} (\mathcal{L}f)(X_u)\,du \\ &= \quad f(Y_t) - f(Y_0) - \int_0^t (\mathcal{L}f)(Y_r)A'_r\,dr \end{aligned}$$

is a local $(\mathcal{F}_{A_t})$ martingale. Noting that

$$A'_r \quad = \quad \frac{1}{T'(A_r)} \quad = \quad \frac{1}{\varrho(X_{A_r})} \quad = \quad \frac{1}{\varrho(Y_r)},$$

we see that w.r.t. the filtration $(\mathcal{F}_{A_t})$, the process Y is a solution of the martingale problem for the operator

$$\widetilde{\mathcal{L}} \quad = \quad \frac{1}{\varrho}\mathcal{L} \quad = \quad \frac{1}{2}\sum_{i,j}\frac{a_{ij}}{\varrho}\,\frac{\partial^2}{\partial x^i \partial x^j} + \frac{b}{\varrho}\cdot\nabla.$$

Since $\frac{a}{\varrho} = \frac{\sigma}{\varrho}\,\frac{\sigma^T}{\varrho}$, this implies that Y is a weak solution of (11.2.6). $\square$

In particular, the theorem shows that if X is a Brownian motion and condition (11.2.5) holds then the time-changed process Y solves the SDE $dY = \varrho(Y)^{-1/2}\,dB$.

Example (Non-uniqueness of weak solutions). Consider the one-dimensional SDE

$$dY_t \quad = \quad |Y_t|^\alpha\,dB_t, \qquad Y_0 \;= 0, \tag{11.2.7}$$

with a one-dimensional Brownian motion (B_t) and $\alpha > 0$. If $\alpha < 1/2$ and x is a Brownian motion with $X_0 = 0$ then the time-change $T_a = \int_0^a \varrho(X_u)\,du$ with $\varrho(x) = |x|^{-2\alpha}$ satisfies

$$\begin{aligned} E[T_a] \quad &= \quad E\Big[\int_0^a \varrho(X_u)\,du\Big] \quad = \quad \int_0^a E[|X_u|^{-2\alpha}]\,du \\ &= \quad E[|X_1|^{-2\alpha}]\cdot\int_0^a u^{-\alpha}\,du \quad < \quad \infty \end{aligned}$$

for any $a \in (0, \infty)$. Hence (11.2.5) holds, and therefore the process $Y_t = X_{A_t}$, $A = T^{-1}$, is a non-trivial weak solution of (11.2.7). On the other hand, $Y_t \equiv 0$ is also a weak solution. Hence for $\alpha < 1/2$, uniqueness in distribution of weak solutions fails. For $\alpha \geq 1/2$, the theorem is not applicable since Assumption (11.2.5) is violated. One can prove that in this case indeed, the trivial solution $Y_t \equiv 0$ is the unique weak solution.

Exercise (Brownian motion on the unit sphere). Let $Y_t = B_t/|B_t|$ where $(B_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^n$, $n > 2$. Prove that the time-changed process

$$Z_a = Y_{T_a}, \qquad T = A^{-1} \text{ with } A_t = \int_0^t |B_s|^{-2}\, ds \, ,$$

is a diffusion taking values in the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ with generator

$$\mathcal{L}f(x) \;=\; \frac{1}{2}\left(\Delta f(x) - \sum_{i,j} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right) - \frac{n-1}{2}\sum_i x_i \frac{\partial f}{\partial x_i}(x), \qquad x \in S^{n-1}.$$

One-dimensional SDE

By combining scale and time transformations, one can carry out a rather complete study of weak solutions for non-degenerate SDE of the form

$$dX_t \;=\; \sigma(X_t)\, dB_t + b(X_t)\, dt, \qquad X_0 = x_0, \tag{11.2.8}$$

on a real interval (α, β). We assume that the initial value X_0 is contained in (α, β), and b, $\sigma : (\alpha, \beta) \to \mathbb{R}$ are continuous functions such that $\sigma(x) > 0$ for any $x \in (\alpha, \beta)$. We first simplify (11.2.8) by a coordinate transformation $Y_t = s(X_t)$ where

$$s \;:\; (\alpha, \beta) \quad \to \quad \big(s(\alpha), s(\beta)\big)$$

is C^2 and satisfies $s'(x) > 0$ for all x. The scale function

$$s(z) \quad := \quad \int_{x_0}^{z} \exp\Big(- \int_{x_0}^{y} \frac{2b(x)}{\sigma(x)^2}\, dx\Big)\, dy$$

has these properties and satisfies $\frac{1}{2}\sigma^2 s'' + b s' = 0$. Hence by the Itô-Doeblin formula, the transformed process $Y_t = s(X_t)$ is a local martingale satisfying

$$dY_t \quad = \quad (\sigma s')(X_t)\, dB_t,$$

i.e., Y is a solution of the equation

$$dY_t \quad = \quad \widetilde{\sigma}(Y_t)\, dB_t, \qquad Y_0 \quad = \quad s(x_0), \tag{11.2.9}$$

where $\widetilde{\sigma} := (\sigma s') \circ s^{-1}$. The SDE (11.2.9) is the original SDE in "natural scale". It can be solved explicitly by a time change. By combining scale transformations and time change one obtains:

Theorem 11.7. *The following statements are equivalent:*

(i) The process $(X_t)_{t<\zeta}$ on the setup $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$ is a weak solution of (11.2.8) defined up to a stopping time ζ.

(ii) The process $Y_t = s(X_t)$, $t < \zeta$, on the same setup is a weak solution of (11.2.9) up to ζ.

(iii) The process $(Y_t)_{s<\zeta}$ has a representation of the form $Y_t = \widetilde{B}_{A_t}$, where $\widetilde{B}_t$ is a one-dimensional Brownian motion satisfying $\widetilde{B}_0 = s(x_0)$ and $A = T^{-1}$ with

$$T_r \quad = \quad \int_0^r \varrho\big(\widetilde{B}_u\big)\, du, \qquad \varrho(y) \quad = \quad 1/\widetilde{\sigma}(y)^2.$$

Carrying out the details of the proof is left as an exercise. The measure $m(dy) := \varrho(y)\, dy$ is called the **"speed measure"** of the process Y although Y is moving faster if m is small. The generator of Y can be written in the form $\mathcal{L} = \frac{1}{2}\frac{d}{dm}\frac{d}{dy}$, and the generator of X is obtained from $\mathcal{L}$ by coordinate transformation. For a much more detailed discussion of one dimensional diffusions we refer to Section V.7 in [38]. Here we only note that 11.7 immediately implies existence and uniqueness of a maximal weak solution of (11.2.8):

Corollary 11.8. *Under the regularity and non-degeneracy conditions on σ and b imposed above there exists a weak solution X of (11.2.8) defined up to the first exit time*

$$\zeta = \inf \left\{ t \geq 0 : \lim_{s \uparrow t} X_t \in \{a, b\} \right\}$$

from the interval (α, β). Moreover, the distribution of any two weak solutions $(X_t)_{t<\zeta}$ and $(\bar{X}_t)_{t<\bar{\zeta}}$ on $\bigcup_{u>0} C([0,u),\mathbb{R})$ coincide.

Remark. We have already seen above that uniqueness may fail if σ is degenerate. For example, the solution of the equation$dY_t = |Y_t|^\alpha \, dB_t$, $Y_0 = 0$, is not unique in distribution for $\alpha \in (0, 1/2)$.

Example (Bessel SDE). Suppose that $(R_t)_{t<\zeta}$ is a maximal weak solution of the Bessel equation

$$dR_t \quad = \quad dW_t + \frac{d-1}{2R_t}\, dt, \qquad W \sim \mathrm{BM}(\mathbb{R}^1),$$

on the interval $(\alpha, \beta) = (0, \infty)$ with initial condition $R_0 = r_0 \in (0, \infty)$ and the parameter $d \in \mathbb{R}$. The ODE $\mathcal{L}s = \frac{1}{2}s'' + \frac{d-1}{2r}s' = 0$ for the scale function has a strictly increasing solution

$$s(r) \quad = \quad \begin{cases} \frac{1}{2-d}\, r^{2-d} & \text{for } \ d \neq 2, \\ \log r & \text{for } \ d = 2 \end{cases}$$

(More generally, $cs + d$ is a strictly increasing solution for any $c > 0$ and $d \in \mathbb{R}$).
Note that s is one-to-one from the interval $(0, \infty)$ onto

$$(s(0), s(\infty)) \quad = \quad \begin{cases} (0, \infty) & \text{for } \ d < 2, \\ (-\infty, \infty) & \text{for } \ d = 2, \\ (-\infty, 0) & \text{for } \ d > 2. \end{cases}$$

By applying the scale transformation, we see that

$$P\big[T_b^R < T_a^R\big] \quad = \quad P\big[T_{s(b)}^{s(R)} < T_{s(a)}^{s(R)}\big] \quad = \quad \frac{s(r_0) - s(a)}{s(b) - s(a)}$$

for any $a < r_0 < b$, where T_c^X denoted the first passage time to c for the process X. As a consequence,

$$P\big[\liminf_{t \uparrow \zeta} R_t = 0\big] \quad = \quad P\Big[\bigcap_{a \in (0, r_0)} \ \bigcup_{b \in (r_0, \infty)} \{T_a^R < T_b^R\}\Big] \quad = \quad \begin{cases} 1 & \text{for } \ d \leq 2, \\ & \\ 0 & \text{for } \ d > 2, \end{cases}$$

$$P\big[\limsup_{t\uparrow\zeta} R_t = \infty\big] \quad = \quad P\Big[\bigcap_{b\in(r_0,\infty)} \bigcup_{a\in(0,r_0)} \{T_b^R < T_a^R\}\Big] \quad = \quad \begin{cases} 1 & \text{for } d \geq 2, \\ 0 & \text{for } d < 2. \end{cases}$$

Note that $d = 2$ is the critical dimension in both cases. Rewriting the SDE in natural scale yields

$$d\,s(R) \quad = \quad \widetilde{\sigma}\big(s(R)\big)\,dW \qquad \text{with} \quad \widetilde{\sigma}(y) \quad = \quad s'\big(s^{-1}(y)\big).$$

In the **critical case** $d = 2$, $s(r) = \log r$, $\widetilde{\sigma}(y) = e^{-y}$, and hence $\varrho(y) = \widetilde{\sigma}(y)^{-2} = e^{2y}$. Thus the speed measure is $m(dy) = e^{2y}\,dy$, and $\log R_t = \widetilde{B}_{T^{-1}(t)}$, i.e.,

$$R_t \quad = \quad \exp\left(\widetilde{B}_{T^{-1}(t)}\right) \qquad \text{with} \quad T_a \quad = \quad \int_0^a \exp\left(2\widetilde{B}_u\right)\,du$$

and a one-dimensional Brownian motion $\widetilde{B}$.

11.3 Change of measure

In Section 11.3, 11.4 and 11.5 we study connections between two different ways of transforming a stochastic process (Y, P):

1) **Random transformations of the paths:** For instance, mapping a Brownian motion (Y_t) to the solution (X_t) of s stochastic differential equation of type

$$dX_t \quad = \quad b(t, X_t)\,dt + dY_t \tag{11.3.1}$$

corresponds to a random translation of the paths of (Y_t):

$$X_t(\omega) \quad = \quad Y_t(\omega) + H_t(\omega) \quad \text{where} \quad H_t \quad = \quad \int_0^t b(X_s)\,ds.$$

2) **Change of measure:** Replace the underlying probability measure P by a modified probability measure Q such that P and Q are mutually absolutely continuous on $\mathcal{F}_t$ for any $t \in [0, \infty)$.

In this section we focus mainly on random transformations of Brownian motions and the corresponding changes of measure. To understand which kind of results we can expect in this case, we first look briefly at a simplified situation:

Example (Translated Gaussian random variables in $\mathbb{R}^d$). We consider the equation

$$X \quad = \quad b(X) + Y, \qquad Y \;\sim\; N(0, I_d) \text{ w.r.t. } P, \tag{11.3.2}$$

for random variables $X, Y : \Omega \to \mathbb{R}^d$ where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a "predictable" map in the sense that the i-th component $b^i(x)$ depends only on the first $i-1$ components $X^i, \dots, X^{i-1}$ of X. The predictability ensures in particular that the transformation defined by (11.3.2) is invertible, with $X^1 = Y^1 + b^1$, $X^2 = Y^2 + b^2(X^1)$, $X^3 = Y^3 + b^3(X^1, X^2)$, ... ,$X^n = Y^n + b^n(X^1, \dots, X^{n-1})$.

A random variable (X, P) is a "weak" solution of the equation (11.3.2) if and only if $Y := X - b(X)$ is standard normally distributed w.r.t. P, i.e., if and only if the distribution $P \circ X^{-1}$ is absolutely continuous with density

$$\begin{aligned} f_X^P(x) &= f_Y^P\big(x - b(x)\big) \left| \det \frac{\partial(x - b(x))}{\partial x} \right| \\ &= (2\pi)^{-d/2} e^{-|x-b(x)|^2/2} \\ &= e^{x \cdot b(x) - |b(x)|^2/2} \, \phi^d(x), \end{aligned}$$

where $\phi^d(x)$ denotes the standard normal density in $\mathbb{R}^d$. Therefore we can conclude:

(X, P) **is a weak solution of (11.3.2) if and only if** $X \sim N(0, I_d)$ **w.r.t. the unique probability measure** Q **on** $\mathbb{R}^d$ **satisfying** $P \ll Q$ **with**

$$\frac{dP}{dQ} \quad = \quad \exp\big(X \cdot b(X) - |b(X)|^2/2\big). \tag{11.3.3}$$

In particular, we see that the law μ^b of a weak solution of (11.3.2) is uniquely determined, and μ^b satisfies

$$\mu^b \quad = \quad P \circ X^{-1} \quad \ll \quad Q \circ X^{-1} \quad = \quad N(0, I_d) \quad = \quad \mu^0$$

with relative density

$$\boxed{\frac{d\mu^b}{d\mu^0}(x) \quad = \quad e^{x \cdot b(x) - |b(x)|^2/2}}$$

The example can be extended to Gaussian measures on Hilbert spaces and to more general transformations, leading to Ramer's generalization of the Cameron-Martin Theorem [1]. Here, we study the more concrete situation where Y and X are replaced by a Brownian motion and a solution of the SDE (11.3.1) respectively. We start by recalling Girsanov's Theorem 9.9.

Change of measure for Brownian motion

Let $(\mathcal{F}_t)$ be a filtration on a measurable space $(\Omega, \mathcal{A})$, and fix $t_0 \in (0, \infty)$. We consider two probability measures P and Q on $(\Omega, \mathcal{A})$ that are mutually absolutely continuous on the σ-algebra $\mathcal{F}_{t_0}$ with relative density

$$Z_{t_0} \quad = \quad \frac{dP}{dQ}\Big|_{\mathcal{F}_{t_0}} \quad > \quad 0 \qquad Q\text{-almost surely.}$$

Then P and Q are also mutually absolutely continuous on each of the σ-algebras $\mathcal{F}_t$, $t \le t_0$, with Q- and P-almost surely strictly positive relative densities

$$Z_t \quad = \quad \frac{dP}{dQ}\Big|_{\mathcal{F}_t} \quad = \quad E_Q\big[Z_{t_0}\big|\mathcal{F}_t\big] \qquad \text{and} \qquad \frac{dQ}{dP}\Big|_{\mathcal{F}_t} \quad = \quad \frac{1}{Z_t}.$$

The process $(Z_t)_{t \le t_0}$ is a martingale w.r.t. Q, and, correspondingly, $(1/Z_t)_{t \le t_0}$ is a martingale w.r.t. P. From now on, we always choose a càdlàg version of these martingales. If the probability measures P and Q are mutually absolutely continuous on the σ-algebra $\mathcal{F}_t$, then the Q-martingale $Z_t = \frac{dP}{dQ}\Big|_{\mathcal{F}_t}$ of relative densities is actually an exponential martingale. Indeed, to obtain a corresponding representation let

$$L_t \quad := \quad \int_0^t \frac{1}{Z_{s-}}\, dZ_s$$

denote the **stochastic "logarithm"** of Z. Here we are using stochastic calculus for càdlàg semimartingales, cf. Chapter 14 below. This can be avoided if one assumes that Q-almost surely, $t \mapsto Z_t$ is continuous, i.e., $Z_{t-} = Z_t$ for $t \ge 0$. In any case, the process $(L_t)_{t \le t_0}$ is a well-defined local martingale w.r.t. Q since Q-a.s., (Z_t) is càdlàg and strictly positive. Moreover, by the associative law,

$$dZ_t \quad = \quad Z_{t-}\, dL_t, \qquad Z_0 \;=\; 1,$$

so Z_t is the stochastic exponential of the local Q-martingale (L_t):

$$Z_t \quad = \quad \mathcal{E}_t^L.$$

In particular, if (Z_t) is continuous then

$$Z_t \;=\; e^{L_t - [L]_t/2}\,.$$

Suppose that (X_t) is a Brownian motion in $\mathbb{R}^d$ with $X_0 = 0$ w.r.t. the probability measure Q and the filtration $(\mathcal{F}_t)$, and fix $t_0 \in [0, \infty)$. Let

$$L_t \quad = \quad \int_0^t G_s \cdot dX_s, \quad t \geq 0,$$

with $G \in \mathcal{L}^2_{\text{a,loc}}\big(\mathbb{R}_+, \mathbb{R}^d\big)$. Notice that if $(\mathcal{F}_t)$ is the filtration generated by the Brownian motion X then by Itô's Representation Theorem 9.11, every local martingale L with $L_0 = 0$ can be represented in this form. Hence in this case, we are considering the most general mutually absolutely continuous measure transformation. Since $[L]_t = \int_0^t |G_s|^2\, ds$,

$$Z_t \quad = \quad \exp\Big(\int_0^t G_s \cdot dX_s - \frac{1}{2}\int_0^t |G_s|^2\, ds\Big) \tag{11.3.4}$$

is the exponential of the local Q-martingale L. Recall from Section 9.3 that $(Z_t)_{t \leq t_0}$ is a martingale under Q if and only if $E_Q[Z_{t_0}] = 1$. By Novikov's criterion, a sufficient condition for the global martingale property of $(Z_t)_{t \leq t_0}$ is

$$E\left[\exp\left(\frac{1}{2}\int_0^{t_0} |G_s|^2\, ds\right)\right] \;=\; E\left[\exp\left([L]_{t_0}/2\right)\right] \;<\; \infty,$$

cf. Theorem 9.10. Assuming the global martingale property, there exists a probability measure P on $\mathcal{A}$ satisfying

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_{t_0}} \quad = \quad Z_{t_0} \qquad Q\text{-a.s.} \tag{11.3.5}$$

Girsanov's Theorem 9.9 states that in this case,

$$B_t \quad := \quad X_t - \int_0^t G_s\, ds, \qquad t \leq t_0,$$

is a Brownian motion w.r.t. any such probability measure P.

Applications to SDE

The Girsanov transformation can be used to construct weak solutions of stochastic differential equations. For example, consider an SDE

$$dX_t \quad = \quad b(t, X_t)\,dt + dB_t, \qquad X_0 \;=\; o, \qquad B \;\sim\; \mathrm{BM}(\mathbb{R}^d), \tag{11.3.6}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is continuous, and $o \in \mathbb{R}^d$ is a fixed initial value. If the drift coefficient is not growing too strongly as $|x| \to \infty$, then we can construct a weak solution of (11.3.6) from Brownian motion by a change of measure. To this end let (X, Q) be an $(\mathcal{F}_t)$ Brownian motion with $X_0 = o$ Q-almost surely, and suppose that the following assumption is satisfied:

Assumption (A). The process

$$Z_t \quad = \quad \exp\left(\int_0^t b(s, X_s) \cdot dX_s - \frac{1}{2}\int_0^t |b(s, X_s)|^2\,ds\right), \qquad t \geq 0,$$

is a martingale w.r.t. Q.

By Novikov's criterion, the assumption is always satisfied if b is bounded. More generally, it can be shown that (A) holds if $b(x)$ is growing at most linearly in x. If (A) holds then $E_Q[Z_t] = 1$ for any $t \geq 0$, and, by Kolmogorov's extension theorem, there exists a probability measure P on $(\Omega, \mathcal{A})$ such that

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} \quad = \quad Z_t \qquad Q\text{-almost surely for any } t \geq 0.$$

By Girsanov's Theorem, the process

$$B_t \quad = \quad X_t - \int_0^t b(s, X_s)\,ds, \qquad t \geq 0,$$

is a Brownian motion w.r.t. P, i.e. (X, P) is a weak solution of the SDE (11.3.6).

More generally, instead of starting from a Brownian motion, we may start from a solution (X, Q) of an SDE of the form

$$dX_t \quad = \quad \beta(t, X_t)\,dt \,+\, \sigma(t, X_t)\,dW_t \tag{11.3.7}$$

where W is an $\mathbb{R}^d$-valued Brownian motion w.r.t. the underlying probability measure Q. We change measure via an exponential martingale of type

$$Z_t = \exp\left(\int_0^t b(s, X_s) \cdot dW_s - \frac{1}{2}\int_0^t |b(s, X_s)|^2 \, ds\right)$$

where $b, \beta : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are continuous functions.

Corollary 11.9 (Drift transformations for SDE). *Suppose that (X, Q) is a weak solution of (11.3.7). If $(Z_t)_{t\geq 0}$ is a Q-martingale and $\mathcal{P} \ll Q$ on $\mathcal{F}_t$ with relative density Z_t for any $t \geq 0$, then (X, P) is a weak solution of*

$$dX_t = (\beta + \sigma b)(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \qquad B \sim \mathrm{BM}(\mathbb{R}^d). \tag{11.3.8}$$

Proof. By (11.3.7), the equation (11.3.8) holds with

$$B_t = W_t - \int_0^t b(s, X_s)\, ds.$$

Girsanov's Theorem implies that B is a Brownian motion w.r.t. P. □

Note that the Girsanov transformation induces a corresponding transformation for the **martingale problem**: If (X, Q) solves the martingale problem for the operator

$$\mathcal{L} = \frac{1}{2}\sum_{i,j} a^{ij}\frac{\partial^2}{\partial x^i \partial x^j} + \beta\cdot\nabla, \qquad a = \sigma\sigma^T, \tag{11.3.9}$$

then (X, P) is a solution of the martingale problem for

$$\widetilde{\mathcal{L}} = \mathcal{L} + (\sigma b)\cdot\nabla = \mathcal{L} + b\cdot\sigma^T\nabla.$$

This **"Girsanov transformation for martingale problems"** carries over to diffusion processes with more general state spaces than $\mathbb{R}^n$.

Doob's h-transform

The h-transform is a change of measure involving a space-time harmonic function that applies to general Markov processes. In the case of Itô diffusions, it turns out to be a special case of the drift transform studied above. Indeed, suppose that $h \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$ is a strictly positive space-time harmonic function for the generator (11.3.9) of the Itô diffusion (X, Q), normalized such that $h(0, o) = 1$:

$$\frac{\partial h}{\partial t} + \mathcal{L}h \quad = \quad 0, \qquad h(0,o) \; = \; 1. \tag{11.3.10}$$

Then, by Itô's formula, the process

$$Z_t \quad = \quad h(t, X_t), \qquad t \geq 0,$$

is a positive local Q-martingale satisfying $Z_0 = 1$ Q-almost surely. We can therefore try to change the measure via (Z_t). To understand the effect of such a transformation, we write Z_t in exponential form. By the Itô-Doeblin formula and (11.3.10),

$$dZ_t \quad = \quad (\sigma^T \nabla h)(t, X_t) \cdot dW_t.$$

Hence $Z_t = \exp(L_t - \frac{1}{2}[L]_t)$ where

$$L_t \quad = \quad \int_0^t \frac{1}{Z_s}\, dZ_s \quad = \quad \int_0^t (\sigma^T \nabla \log h)(s, X_s) \cdot dW_s$$

is the stochastic logarithm of Z. Thus if (Z, Q) is a martingale, and $P \ll Q$ with local densities $\frac{dP}{dQ}\big|_{\mathcal{F}_t} = Z_t$ then (X, P) solves the SDE (11.3.7) with $b = \sigma^T \nabla \log h$, i.e.,

$$dX_t \;=\; (\beta + \sigma\sigma^T \nabla \log h)(t, X_t)\, dt \;+\; \sigma(t, X_t)\, dB_t, \quad B \;\sim\; \mathrm{BM}(\mathbb{R}^d) \;\text{ w.r.t. } P. \tag{11.3.11}$$

The proces (X, P) is called the **h-transform** of (X, Q).

Example. If $X_t = W_t$ is a Brownian motion w.r.t. Q then

$$dX_t \quad = \quad \nabla \log h(t, X_t)\, dt + dB_t, \qquad B \;\sim\; \mathrm{BM}(\mathbb{R}^d) \;\text{ w.r.t. } P.$$

For example, choosing $h(t,x) = \exp(\alpha \cdot x - \frac{1}{2}|\alpha|^2 t), \alpha \in \mathbb{R}^d$, (X, P) is a Brownian motion with constant drift α, i.e., $dX_t = \alpha\, dt \;+\; dB_t$.

11.4 Path integrals and bridges

One way of thinking about a stochastic process is to interpret it as a probability measure on path space. This useful point of view will be developed further in this and the following section. We consider an SDE

$$dW_t \;=\; b(W_t)\,dt + dB_t, \qquad W_0 \;=\; o, \quad B \;\sim\; \mathrm{BM}(\mathbb{R}^d) \tag{11.4.1}$$

with initial condition $o \in \mathbb{R}^d$ and $b \in C(\mathbb{R}^d, \mathbb{R}^d)$. We will show that the solution constructed by Girsanov transformation is a Markov process, and we will study its transition function, as well as the bridge process obtained by conditioning on a given value at a fixed time.

Let μ_o denote the law of Brownian motion starting at o on $(\Omega, \mathcal{F}^W_\infty)$ where $\Omega = C(\mathbb{R}_+, \mathbb{R}^d)$ and $W_t(x) = x_t$ is the canonical Brownian motion on (Ω, μ_o). Let

$$Z_t \;=\; \exp\left(\int_0^t b(W_s)\cdot dW_s - \frac{1}{2}\int_0^t |b(W_s)|^2\,ds\right). \tag{11.4.2}$$

Note that if $b(x) = -\nabla H(x)$ for a function $H \in C^2(\mathbb{R}^d)$ then by Itô's formula,

$$Z_t \;=\; \exp\left(H(W_0) - H(W_t) + \frac{1}{2}\int_0^t \left(\Delta H - |\nabla H|^2\right)(W_s)\,ds\right). \tag{11.4.3}$$

This shows that Z is more robust w.r.t. variations of (W_t) if b is a gradient vector field, because (11.4.3) does not involve a stochastic integral. This robustness is crucial for certain applications, see the example below. Similarly as above, we assume:

Assumption (A). The exponential $(Z_t)_{t\geq 0}$ is a martingale w.r.t. μ_o.

We note that by Novikov's criterion, the assumption always holds if

$$|b(x)| \quad\leq\quad c\cdot(1+|x|) \qquad \text{for some finite constant} \quad c>0: \tag{11.4.4}$$

Exercise (Martingale property for exponentials). Prove that (Z_t) is a martingale if (11.4.4) holds. *Hint: Prove first that $E[\exp\int_0^\varepsilon |b(W_s)|^2\,ds] < \infty$ for $\varepsilon > 0$ sufficiently small, and conclude that $E[Z_\varepsilon] = 1$. Then show by induction that $E[Z_{k\varepsilon}] = 1$ for any $k \in \mathbb{N}$.*

If (A) holds then by the Kolmogorov extension theorem, there exists a probability measure μ_o^b on $\mathcal{F}_\infty^W$ such that μ_o^b and μ_o are mutually absolutely continuous on each of the σ-algebras $\mathcal{F}_t^W$, $t \in [0,\infty)$, with relative densities

$$\left.\frac{d\mu_o^b}{d\mu_o}\right|_{\mathcal{F}_t^W} \quad = \quad Z_t \qquad \mu_o\text{-a.s.}$$

Girsanov's Theorem implies:

Corollary 11.10. *Suppose that (A) holds. Then:*

1) *The process (W, μ_o^b) is a weak solution of (11.3.6).*
2) *For any $t \in [0,\infty)$, the law of (W, μ_o^b) is absolutely continuous w.r.t. Wiener measure μ_o on $\mathcal{F}_t^W$ with relative density Z_t.*

The first assertion follows since $B_t = W_t - \int_0^t b(W_s)\,ds$ is a Brownian motion w.r.t. μ_o^b, and the second assertion holds since $\mu_o^b \circ W^{-1} = \mu_o^b$.

Path integral representation

Corollary 11.10 yields a rigorous **path integral representation** for the solution (W, μ_o^b) of the SDE (11.3.6): If $\mu_o^{b,t}$ denotes the law of $(W_s)_{s\le t}$ on $C\big([0,t],\mathbb{R}^d\big)$ w.r.t. μ_o^b then

$$\mu_o^{b,t}(dx) \quad = \quad \exp\left(\int_0^t b(x_s)\cdot dx_s - \frac{1}{2}\int_0^t |b(x_s)|^2\,ds\right)\,\mu_o^{0,t}(dx). \tag{11.4.5}$$

By combining (11.4.5) with the **heuristic** path integral representation

$$\text{“}\ \mu_o^{0,t}(dx) \quad = \quad \frac{1}{\infty}\exp\left(-\frac{1}{2}\int_0^t |x_s'|^2\,ds\right)\,\delta_0(dx_0)\prod_{0<s\le t} dx_s\ \text{”}$$

of Wiener measure, we obtain the non-rigorous but very intuitive representation

$$\text{“}\ \mu_o^{b,t}(dx) \quad = \quad \frac{1}{\infty}\exp\left(-\frac{1}{2}\int_0^t |x_s' - b(x_s)|^2\,ds\right)\,\delta_0(dx_0)\prod_{0<s\le t} dx_s\ \text{”} \tag{11.4.6}$$

of $\mu_o^{b,t}$. Hence intuitively, the "likely" paths w.r.t. $\mu_o^{b,t}$ should be those for which the **action functional**

$$I(x) \quad = \quad \frac{1}{2}\int_0^t \left|x_s' - b(x_s)\right|^2 ds$$

takes small values, and the "most likely trajectory" should be the solution of the deterministic ODE

$$x_s' \quad = \quad b(s, x_s)$$

obtained by setting the noise term in the SDE (11.3.6) equal to zero. Of course, these arguments do not hold rigorously, because $I(x) = \infty$ for $\mu_o^{0,t}$- and $\mu_o^{b,t}$- almost every x. Nevertheless, they provide an extremely valuable guideline to conclusions that can then be verified rigorously, for instance via (11.4.5).

Example (Likelihood ratio test for non-linear filtering). Suppose that we are observing a noisy signal (x_t) taking values in $\mathbb{R}^d$ with $x_0 = o$. We interpret (x_t) as a realization of a stochastic process (X_t). We would like to decide if there is only noise, or if the signal is coming from an object moving with law of motion $dx/dt = -\nabla H(x)$ where $H \in C^2(\mathbb{R}^d)$. The noise is modelled by the increments of a Brownian motion (white noise). This is a simplified form of models that are used frequently in nonlinear filtering (in realistic models often the velocity or the acceleration is assumed to satisfy a similar equation). In a hypothesis test, the null hypothesis and the alternative would be

$$\begin{aligned} H_0 : \qquad X_t \quad &= \quad B_t, \\ H_1 : \qquad dX_t \quad &= \quad b(X_t)\, dt + dB_t, \end{aligned}$$

where (B_t) is a d-dimensional Brownian motion, and $b = -\nabla H$. In a likelihood ratio test based on observations up to time t, the test statistic would be the likelihood ratio $d\mu_o^{b,t}/d\mu_o^{0,t}$ which by (11.4.3) can be represented in the robust form

$$\frac{d\mu_o^{b,t}}{d\mu_o^{0,t}}(x) \quad = \quad \exp\left(H(x_0) - H(x_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(x_s)\, ds\right) \tag{11.4.7}$$

The null hypothesis H_0 would then be rejected if this quantity exceeds some given value c for the observed signal x, i.e. , if

$$H(x_0) - H(x_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(x_s)\, ds \quad > \quad \log c. \tag{11.4.8}$$

Note that the robust representation of the density ensures that the estimation procedure is quite stable, because the $\log$ likelihood ratio in (11.4.8) is continuous w.r.t. the supremum norm on $C([0,t],\mathbb{R}^d)$.

The Markov property

Recall that if (A) holds then there exists a (unique) probability measure μ_o^b on $(\Omega, \mathcal{F}_\infty^W)$ such that

$$\mu_o^b[A] \quad = \quad E_o[Z_t\,;\ A] \qquad \text{for any } \ t \geq 0 \ \text{ and } \ A \in \mathcal{F}_t^W.$$

Here E_x denotes expectation w.r.t. Wiener measure μ_x with start in x. By Girsanov's Theorem, the process (W, μ_o^b) is a weak solution of (11.4.1). Moreover, we can easily verify that (W, μ_o^b) is a Markov process:

Theorem 11.11 (Markov property). *If (A) holds then (W, μ_o^b) is a time-homogeneous Markov process with transition function*

$$p_t^b(x, C) \quad = \quad \mu_x^b[W_t \in C] \quad = \quad E_x[Z_t\,;\ W_t \in C] \qquad \forall\, C \in \mathcal{B}(\mathbb{R}^d).$$

Proof. Let $0 \leq s \leq t$, and let $f : \mathbb{R}^d \to \mathbb{R}_+$ be a non-negative measurable function. Then, by the Markov property for Brownian motion,

$$\begin{aligned} E_o^b[f(W_t)|\mathcal{F}_s^W] \quad &= \quad E_o[f(W_t)Z_t|\mathcal{F}_s^W]/Z_s \\ &= \quad E_o\Big[f(W_t)\exp\left(\int_s^t b(W_r)\cdot dW_r - \frac{1}{2}\int_s^t |b(W_r)|^2\,dr\right)\Big|\mathcal{F}_s^W\Big] \\ &= \quad E_{W_s}[f(W_{t-s})Z_{t-s}] \quad = \quad (p_{t-s}^b f)(W_s) \end{aligned}$$

μ_o- and μ_o^b-almost surely where E_x^b denotes the expectation w.r.t. μ_x^b. □

Remark. 1) If b is time-dependent then one verifies in the same way that (W, μ_o^b) is a time-inhomogeneous Markov process.
2) It is not always easy to prove that solutions of SDE are Markov processes. If the solution is not unique then usually, there are solutions that are not Markov processes.

Bridges and heat kernels

We now restrict ourselves to the time-interval $[0, 1]$, i.e., we consider a similar setup as before with $\Omega = C([0, 1], \mathbb{R}^d)$. Note that $\mathcal{F}_1^W$ is the Borel σ-algebra on the Banach space Ω. Our goal is to condition the diffusion process (W, μ_o^b) on a given terminal value $W_1 = y$, $y \in \mathbb{R}^d$. More precisely, we will construct a **regular version $y \mapsto \mu_{o,y}^b$ of the conditional distribution $\mu_o^b[\cdot|W_1 = y]$** in the following sense:

(i) For any $y \in \mathbb{R}^d$, $\mu_{o,y}^b$ is a probability measure on $\mathcal{B}(\Omega)$,and $\mu_{o,y}^b[W_1 = y] = 1$.

(ii) Disintegration: For any $A \in \mathcal{B}(\Omega)$, the function $y \mapsto \mu_{o,y}^b[A]$ is measurable, and

$$\mu_o^b[A] \quad = \quad \int_{\mathbb{R}^d} \mu_{o,y}^b[A]\, p_1^b(o, dy).$$

(iii) The map $y \mapsto \mu_{o,y}^b$ is continuous w.r.t. weak convergence of probability measures.

Example (Brownian bridge). For $b = 0$, a regular version $y \mapsto \mu_{o,y}$ of the conditional distribution $\mu_o[\,\cdot\,|W_1 = y]$ w.r.t. Wiener measure μ_o can be obtained by linearly transforming the paths of Brownian motion, cf. Theorem 8.11 in [14]: Under μ_o, the process

$$X_t^y \;:=\; W_t - tW_1 + ty, \qquad 0 \le t \le 1,$$

is independent of W_1 with terminal value y, and the law $\mu_{o,y}$ of $(X_t^y)_{t\in[0,1]}$ w.r.t. μ_o is a regular version of $\mu_o[\,\cdot\,|W_1 = y]$. The measure $\mu_{o,y}$ is called **"pinned Wiener measure"**.

The construction of a bridge process described in the example only applies for Brownian motion and other Gaussian processes. For more general diffusions, the bridge can not be constructed from the original process by a linear transformation of the paths. For perturbations of a Brownian motion by a drift, however, we can apply Girsanov's Theorem to construct a bridge measure.

We assume for simplicity again that b is the gradient of a C^2 function:

$$b(x) \quad = \quad -\nabla H(x) \qquad \text{with} \quad H \in C^2(\mathbb{R}^d).$$

Then the exponential martingale (Z_t) takes the form

$$Z_t = \exp\left(H(W_0) - H(W_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(W_s)\,ds\right),$$

cf. (11.4.3). Note that the expression on the right-hand side is defined $\mu_{o,y}$-almost surely for any y. Therefore, (Z_t) can be used for changing the measure w.r.t. the Brownian bridge.

Theorem 11.12 (Heat kernel and Bridge measure). *Suppose that (A) holds. Then:*

1) *The measure $p_1^b(o, dy)$ is absolutely continuous w.r.t. d-dimensional Lebesgue measure with density*

$$p_1^b(o, y) = p_1(o, y) \cdot E_{o,y}[Z_1].$$

2) *A regular version of $\mu_o^b[\,\cdot\,|W_1 = y]$ is given by*

$$\mu_{o,y}^b(dx) = \frac{p_1(o,y)}{p_1^b(o,y)}\frac{\exp H(o)}{\exp H(y)}\exp\left(\frac{1}{2}\int_0^1 (\Delta H - |\nabla H|^2)(x_s)\,ds\right)\mu_{o,y}(dx).$$

The theorem yields the existence and a formula for the heat kernel $p_1^b(o, y)$, as well as a path integral representation for the bridge measure $\mu_{o,y}^b$:

$$\mu_{o,y}^b(dx) \propto \exp\left(\frac{1}{2}\int_0^1 (\Delta H - |\nabla H|^2)(x_s)\,ds\right)\mu_{o,y}(dx). \tag{11.4.9}$$

Proof of 11.12. Let $F : \Omega \to \mathbb{R}_+$ and $g : \mathbb{R}^d \to \mathbb{R}_+$ be measurable functions. By the disintegration of Wiener measure into pinned Wiener measures,

$$E_o^b[F \cdot g(W_1)] = E_o[Fg(W_1)Z_1] = \int E_{o,y}[FZ_1]\, g(y)\, p_1(o, y)\, dy.$$

Choosing $F \equiv 1$, we obtain

$$\int g(y)\, p_1^b(o, dy) = \int g(y)\, E_{o,y}[Z_1]\, p_1(o, y)\, dy$$

for any non-negative measurable function g, which implies 1).
Now, choosing $g \equiv 1$, we obtain by 1) that

$$E_o^b[F] = \int E_{o,y}[FZ_1]\, p_1(o,y)\, dy = \int \frac{E_{o,y}[FZ_1]}{E_{o,y}[Z_1]}\, p_1^b(o,dy) \qquad (11.4.10)$$

$$= \int E_{o,y}^b[F]\, p_1^b(o,dy) \qquad (11.4.11)$$

This proves 2), because $W_1 = y$ $\mu_{o,y}^b$-a.s., and $y \mapsto \mu_{o,y}^b$ is weakly continuous. □

Remark (Non-gradient case). If b is not a gradient then things are more involved because the expressions for the relative densities Z_t involve a stochastic integral. In principle, one can proceed similarly as above after making sense of this stochastic integral for $\mu_{o,y}$-almost every path x.

Example (Reversibility in the gradient case). The representation (11.4.9) immediately implies the following reversibility property of the diffusion bridge when b is a gradient: If $R : C([0,1],\mathbb{R}^d) \to C([0,1],\mathbb{R}^d)$ denotes the time-reversal defined by $(Rx)_t = x_{1-t}$, then the image $\mu_{o,y}^b \circ R^{-1}$ of the bridge measure from o to y coincides with the bridge measure $\mu_{y,o}^b$ from y to o. Indeed, this property holds for the Brownian bridge, and the relative density in (11.4.9) is invariant under time reversal.

SDE for diffusion bridges

An important application of the h-transform is the interpretation of diffusion bridges by a change of measure w.r.t. the law of the unconditioned diffusion process (W, μ_o^b) on $C([0,1],\mathbb{R}^d)$ satisfying

$$dW_t = dB_t + b(W_t)\, dt, \qquad W_0 = o,$$

with an $\mathbb{R}^d$-valued Brownian motion B. We assume that the transition density $(t,x,y) \mapsto p_t^b(x,y)$ is smooth for $t > 0$ and bounded for $t \geq \varepsilon$ for any $\varepsilon > 0$. Then for $y \in \mathbb{R}$, $p_t^b(\cdot,y)$ satisfies the Kolmogorov backward equation

$$\frac{\partial}{\partial t} p_t^b(\cdot,y) = \mathcal{L}^b p_t^b(\cdot,y) \qquad \text{for any } t > 0,$$

where $\mathcal{L}^b = \frac{1}{2}\Delta + b \cdot \nabla$ is the corresponding generator. Hence

$$h(t, z) \quad = \quad p_{1-t}^b(z, y)/p_1^b(o, y), \qquad t \; < \; 1,$$

is a space-time harmonic function with $h(0, o) = 1$. Since h is bounded for $t \le 1 - \varepsilon$ for any $\varepsilon > 0$, the process $h(t, W_t)$ is a martingale under μ_o^b for $t < 1$. Now let $\mu_{o,y}^b$ be the measure on $C([0, 1], \mathbb{R}^d)$ that is absolutely continuous w.r.t. μ_o^b on $\mathcal{F}_t^W$ with relative density $h(t, W_t)$ for any $t < 1$. Then the marginal distributions of the process $(W_t)_{t<1}$ under μ_o^b, $\mu_{o,y}^b$ respectively are

$$\begin{aligned}(W_{t_1}, \ldots, W_{t_k}) \;&\sim\; p_{t_1}^b(o, x_1)p_{t_2-t_1}^b(x_1, x_2) \cdots p_{t_k-t_{k-1}}^b(x_{k-1}, x_k)\; \lambda^k(dx) \qquad \text{w.r.t. } \mu_o^b, \\ &\sim\; \frac{p_{t_1}^b(o, x_1)p_{t_2-t_1}^b(x_1, x_2) \cdots p_{t_k-t_{k-1}}^b(x_{k-1}, x_k)p_{1-t_k}^b(x_k, y)}{p_1^b(o, y)}\; \lambda^k(dx) \qquad \text{w.r.t. } \mu_{o,y}^b.\end{aligned}$$

This shows that $y \to \mu_{o,y}^b$ coincides with the regular version of the conditional distribution of μ_o^b given W_1, i.e., $\mu_{o,y}^b$ is the bridge measure from o to y. Hence, by Corollary 11.9, we have shown:

Theorem 11.13 (SDE for diffusion bridges). *The diffusion bridge $(W, \mu_{o,y}^b)$ is a weak solution of the SDE*

$$dW_t \quad = \quad dB_t \;+\; b(W_t)\, dt \;+\; (\nabla \log p_{1-t}^b(\cdot, y))(W_t)\;\, dt, \qquad t \; < \; 1. \qquad (11.4.12)$$

Note that the additional drift $\beta(t, x) = \nabla \log p_{1-t}^b(\cdot, y)(x)$ is singular as $t \uparrow 1$. Indeed, if at a time close to 1 the process is still far away from y, then a strong drift is required to force it towards y. On the σ-algebra $\mathcal{F}_1^W$, the measures μ_o^b and $\mu_{o,y}^b$ are singular.

Remark (Generalized diffusion bridges). Theorem 11.13 carries over to bridges of diffusion processes with non-constant diffusion coefficients σ. In this case, the SDE (11.4.12) is replaced by

$$dW_t \quad = \quad \sigma(W_t)\, dB_t + b(W_t)\, dt + \left(\sigma\sigma^T \nabla \log p_{1-t}(\cdot, y)\right) (W_t)\, dt. \qquad (11.4.13)$$

The last term can be interpreted as a gradient of the logarithmic heat kernel w.r.t. the Riemannian metric $g = (\sigma\sigma^T)^{-1}$ induced by the diffusion process.

11.5 Large deviations on path spaces

In this section, we apply Girsanov's Theorem to study random perturbations of a dynamical system of type

$$dX_t^\varepsilon \quad = \quad b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\,dB_t, \qquad X_0^\varepsilon \;=\; 0, \tag{11.5.1}$$

asymptotically as $\varepsilon \downarrow 0$. We show that on the exponential scale, statements about the probabilities of rare events suggested by path integral heuristics can be put in a rigorous form as a large deviation principle on path space.

Let $\Omega = C_0([0,1],\mathbb{R}^d)$ endowed with the supremum norm $||\omega|| = \sup\{|\omega(t)| \;:\; t \in [0,1]\}$, let μ denote Wiener measure on $\mathcal{B}(\Omega)$, and let $W_t(\omega) = \omega(t)$.

Support of Wiener measure

For $h \in \Omega$, we consider the translation operator $\tau_h : \Omega \to \Omega$,

$$\tau_h(\omega) \quad = \quad \omega + h,$$

and the translated Wiener measure $\mu_h \;:=\; \mu \circ \tau_h^{-1}$. We recall the Cameron-Martin Theorem from Section 9.2. For the convenience of the reader, we also give a second proof that is based on Girsanov's Theorem:

Theorem 11.14 (Cameron, Martin 1944). *Let $h \in \Omega$. Then $\mu_t \ll \mu$ if and only if h is contained in the* ***Cameron-Martin space***

$$H_{CM} \quad = \quad \left\{h \in \Omega \;:\; h \text{ is absolutely contin. with } h' \in L^2([0,1],\mathbb{R}^d)\right\}.$$

In this case, the relative density of μ_h w.r.t. μ is

$$\frac{d\mu_h}{d\mu} \quad = \quad \exp\Big(\int_0^t h_s' \cdot dW_s - \frac{1}{2}\int_0^t |h_s'|^2\,ds\Big). \tag{11.5.2}$$

Proof. "⇒" is a consequence of Girsanov's Theorem: For $h \in H_{CM}$, the stochastic integral $\int h' \cdot dW$ has finite deterministic quadratic variation $[\int h' \cdot dW]_1 = \int_0^1 |h'|^2\, ds$. Hence by Novikov's criterion,

$$Z_t \quad = \quad \exp\Big(\int_0^t h' \cdot dW - \frac{1}{2}\int_0^t |h'|^2\, ds\Big)$$

is a martingale w.r.t. Wiener measure μ. Girsanov's Theorem implies that w.r.t. the measure $\nu = Z_1 \cdot \mu$, the process (W_t) is a Brownian motion translated by (h_t). Hence

$$\mu_h \quad = \quad \mu \circ (W+h)^{-1} \quad = \quad \nu \circ W^{-1} \quad = \quad \nu.$$

"⇐" To prove the converse implication let $h \in \Omega$, and suppose that $\mu_h \ll \mu$. Since W is a Brownian motion w.r.t. μ, $W - h$ is a Brownian motion w.r.t. μ_h. In particular, it is a semimartingale. Moreover, W is a semimartingale w.r.t. μ and hence also w.r.t. μ_h. Thus $h = W - (W - h)$ is also a semimartingale w.r.t. μ_h. Since h is deterministic, this implies that h has **finite variation**. We now show:

Claim. The map $g \mapsto \int_0^1 g \cdot dh$ is a continuous linear functional on $L^2([0,1],\mathbb{R}^d)$.

The claim implies $h \in H_{CM}$. Indeed, by the claim and the Riesz Representation Theorem, there exists a function $f \in L^2([0,1],\mathbb{R}^d)$ such that

$$\int_0^1 g \cdot dh \quad = \quad \int_0^1 g \cdot f ds \qquad \text{for any } \; g \in L^2([0,1],\mathbb{R}^d).$$

Hence h is absolutely continuous with $h' = f \ \in L^2([0,1],\mathbb{R}^d)$. To prove the claim let (g_n) be a sequence in $L^2([0,1],\mathbb{R}^d)$ with $||g_n||_{L^2} \to 0$. Then by Itô's isometry, $\int g_n\, dW \to 0$ in $L^2(\mu)$, and hence μ- and μ_h-almost surely along a subsequence. Thus also

$$\int g_n \cdot dh \quad = \quad \int g_n \cdot d(W+h) - \int g_n \cdot dW \quad \longrightarrow \quad 0$$

μ-almost surely along a subsequence. Applying the same argument to a subsequence of (g_n), we see that every subsequence $(\widetilde{g}_n)$ has a subsequence $(\hat{g}_n)$ such that $\int \hat{g}_n \cdot dh \to 0$. This shows that $\int g_n \cdot dh$ converges to 0 as well. The claim follows, since (g_n) was an arbitrary null sequence in $L^2([0,1],\mathbb{R}^d)$. □

A first consequence of the Cameron-Martin Theorem is that the support of Wiener measure is the whole space $\Omega = C_0([0,1], \mathbb{R}^d)$:

Corollary 11.15 (Support Theorem). *For any $h \in \Omega$ and $\delta > 0$,*

$$\mu\big[\{\omega \in \Omega \;:\; ||\omega - h|| < \delta\}\big] \quad > \quad 0.$$

Proof. Since the Cameron-Martin space is dense in Ω w.r.t. the supremum norm, it is enough to prove the assertion for $h \in H_{CM}$. In this case, the Cameron-Martin Theorem implies

$$\mu\big[||W - h|| < \delta\big] \quad = \quad \mu_{-h}\big[||W|| < \delta\big] \quad > \quad 0.$$

as $\mu[||W|| < \delta] > 0$ and $\mu_{-h} \ll \mu$. □

Remark (Quantitative Support Theorem). More explicitly,

$$\begin{aligned} \mu\big[||W - h|| < \delta\big] \quad &= \quad \mu_{-h}\big[||W|| < \delta\big] \\ &= \quad E\Big[\exp\Big(-\int_0^1 h' \cdot dW - \frac{1}{2}\int_0^1 |h'|^2\, ds\Big) \;;\; \max_{s\leq 1} |W_s| < \delta\Big] \end{aligned}$$

where the expectation is w.r.t. Wiener measure. This can be used to derive quantitative estimates.

Schilder's Theorem

We now study the solution of (11.5.1) for $b \equiv 0$, i.e.,

$$X_t^\varepsilon \quad = \quad \sqrt{\varepsilon}\, B_t, \qquad t \in [0,1],$$

with $\varepsilon > 0$ and a d-dimensional Brownian motion (B_t). Path integral heuristics suggests that for $h \in H_{CM}$,

$$\text{“}\; P[X^\varepsilon \approx h] \quad = \quad \mu\Big[W \approx \frac{h}{\sqrt{\varepsilon}}\Big] \quad \sim \quad e^{-I(h/\sqrt{\varepsilon})} \quad = \quad e^{-I(h)/\varepsilon} \;\text{”}$$

where $I : \Omega \to [0, \infty]$ is the **action functional** defined by

$$I(\omega) \quad = \quad \begin{cases} \frac{1}{2}\int_0^1 |\omega'(s)|^2\, ds & \text{if } \omega \in H_{CM}, \\ +\infty & \text{otherwise.} \end{cases}$$

The heuristics can be turned into a rigorous statement asymptotically as $\varepsilon \to 0$ on the exponential scale. This is the content of the next two results that together are know as Schilder's Theorem:

Theorem 11.16 (Schilder's large derivation principle, lower bound).

1) For any $h \in H_{CM}$ and $\delta > 0$,

$$\liminf_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu\big[\sqrt{\varepsilon} W \in B(h, \delta)\big] \quad \geq \quad -I(h).$$

2) For any open subset $U \subseteq \Omega$,

$$\liminf_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu\big[\sqrt{\varepsilon} W \in U\big] \quad \geq \quad -\inf_{\omega \in U} I(\omega).$$

Here $B(h, \delta) = \{\omega \in \Omega\ :\ ||\omega - h|| < \delta\}$ denotes the ball w.r.t. the supremum norm.

Proof. 1) Let $c = \sqrt{8I(h)}$. Then for $\varepsilon > 0$ sufficiently small,

$$\begin{aligned} \mu\big[\sqrt{\varepsilon} W \in B(h, \delta)\big] \quad &= \quad \mu\big[W \in B(h/\sqrt{\varepsilon}, \delta/\sqrt{\varepsilon})\big] \\ &= \quad \mu_{-h/\sqrt{\varepsilon}}\big[B(0, \delta/\sqrt{\varepsilon})\big] \\ &= \quad E\Big[\exp\Big(-\frac{1}{\sqrt{\varepsilon}}\int_0^1 h' \cdot dW - \frac{1}{2\varepsilon}\int_0^1 |h'|^2\, ds\Big)\ ;\ B\Big(0, \frac{\delta}{\sqrt{\varepsilon}}\Big)\Big] \\ &\geq \quad \exp\Big(-\frac{1}{\varepsilon} I(h) - \frac{c}{\sqrt{\varepsilon}}\Big)\ \mu\Big[\Big\{\int_0^1 h' \cdot dW \leq c\Big\} \cap B(0, \frac{\delta}{\sqrt{\varepsilon}})\Big] \\ &\geq \quad \frac{1}{2}\exp\left(-\frac{1}{\varepsilon} I(h) - \sqrt{\frac{8I(h)}{\varepsilon}}\right) \end{aligned}$$

where E stands for expectation w.r.t. Wiener measure. Here we have used that

$$\mu\Big[\int_0^1 h' \cdot dW > c\Big] \quad \leq \quad c^{-2} E\Big[\Big(\int_0^1 h' \cdot dW\Big)^2\Big] \quad = \quad 2I(h)/c^2 \quad \leq \quad 1/4$$

by Itô's isometry and the choice of c.

2) Let U be an open subset of Ω. For $h \in U \cap H_{CM}$, there exists $\delta > 0$ such that $B(h,\delta) \subset U$. Hence by 1),

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon} W \in U] \quad \geq \quad -I(h).$$

Since this lower bound holds for any $h \in U \cap H_{CM}$, and since $I = \infty$ on $U \setminus H_{CM}$, we can conclude that

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon} W \in U] \quad \geq \quad -\inf_{h \in U \cap H_{CM}} I(h) \quad = \quad -\inf_{\omega \in U} I(\omega).$$

□

To prove a corresponding upper bound, we consider linear approximations of the Brownian paths. For $n \in \mathbb{N}$ let

$$W_t^{(n)} := (1-s)W_{k/n} + sW_{k+1/n}$$

whenever $t = (k+s)/n$ for $k \in \{0, 1, \dots, n-1\}$ and $s \in [0,1]$.

Theorem 11.17 (Schilder's large deviations principle, upper bound).

1) For any $n \in \mathbb{N}$ and $\lambda \geq 0$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[I(\sqrt{\varepsilon} W^{(n)}) \geq \lambda] \quad \leq \quad -\lambda.$$

2) For any closed subset $A \subseteq \Omega$,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon} W \in A] \quad \leq \quad -\inf_{\omega \in A} I(\omega).$$

Proof. 1) Let $\varepsilon > 0$ and $n \in \mathbb{N}$. Then

$$I(\sqrt{\varepsilon} W^{(n)}) \quad = \quad \frac{1}{2}\varepsilon \sum_{k=1}^{n} n\,(W_{k/n} - W_{(k-1)/n})^2.$$

Since the random variables $\eta_k := \sqrt{n} \cdot (W_{k/n} - W_{(k-1)/n})$ are independent and standard normally distributed, we obtain

$$\begin{aligned} \mu[I(\sqrt{\varepsilon} W^{(n)}) \geq \lambda] &= \mu\Big[\sum |\eta_k|^2 \geq 2\lambda/\varepsilon\Big] \\ &\leq \exp(-2\lambda c/\varepsilon)\, E\Big[\exp\Big(c \sum |\eta_k|^2\Big)\Big], \end{aligned}$$

where the expectation on the right hand side is finite for $c < 1/2$. Hence for any $c < 1/2$,

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[I(\sqrt{\varepsilon} W^{(n)}) \geq \lambda] \quad \leq \quad -2c\lambda.$$

The assertion now follows as $c \uparrow 1/2$.

2) Now fix a closed set $A \subseteq \Omega$ and $\lambda < \inf\{I(\omega) : \omega \in A\}$. To prove the second assertion it suffices to show

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[\sqrt{\varepsilon} W \in A] \quad \leq \quad -\lambda. \tag{11.5.3}$$

By the Theorem of Arzéla-Ascoli, the set $\{I \leq \lambda\}$ is a **compact** subset of the Banach space Ω. Indeed, by the Cauchy-Schwarz inequality,

$$|\omega(t) - \omega(s)| \quad = \quad \Big|\int_s^t \omega'(u)\, du\Big| \quad \leq \quad \sqrt{2\lambda}\,\sqrt{t-s} \qquad \forall\, s,t \in [0,1]$$

holds for any $\omega \in \Omega$ satisfying $I(\omega) \leq \lambda$. Hence the paths in $\{I \leq \lambda\}$ are equicontinuous, and the Arzéla-Ascoli Theorem applies.

Let δ denote the distance between the sets A and $\{I \leq \lambda\}$ w.r.t. the supremum norm. Note that $\delta > 0$, because A is closed, $\{I \leq \lambda\}$ is compact, and both sets are disjoint by the choice of λ. Hence for $\varepsilon > 0$, we can estimate

$$\mu[\sqrt{\varepsilon} W \in A] \quad \leq \quad \mu[I(\sqrt{\varepsilon} W^{(n)}) > \lambda] + \mu[||\sqrt{\varepsilon} W - \sqrt{\varepsilon} W^{(n)}||_{\sup} > \delta].$$

The assertion (11.5.3) now follows from

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[I(\sqrt{\varepsilon} W^{(n)}) > \lambda] \quad \leq \quad -\lambda, \quad \text{and} \tag{11.5.4}$$

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[||W - W^{(n)}||_{\sup} > \delta/\sqrt{\varepsilon}] \quad \leq \quad -\lambda. \tag{11.5.5}$$

The bound (11.5.4) holds by 1) for any $n \in \mathbb{N}$. The proof of (11.5.5) reduces to an estimate of the supremum of a Brownian bridge on an interval of length $1/n$. We leave it as an exercise to verify that (11.5.5) holds if n is large enough. $\square$

Remark (Large deviation principle for Wiener measure). Theorems 11.16 and 11.17 show that

$$\mu[\sqrt{\varepsilon}W \in A] \quad \simeq \quad \exp\Big(-\frac{1}{\varepsilon}\inf_{\omega \in A} I(\omega)\Big)$$

holds on the exponential scale in the sense that a lower bound holds for open sets and an upper bound holds for closed sets. This is typical for large deviation principles, see e.g. [10] or [11]. The proofs above based on "exponential tilting" of the underlying Wiener measure (Girsanov transformation) for the lower bound, and an exponential estimate combined with exponential tightness for the upper bound are typical for the proofs of many large deviation principles.

Random perturbations of dynamical systems

We now return to our original problem of studying small random perturbations of a dynamical system

$$dX_t^\varepsilon \quad = \quad b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\,dB_t, \qquad X_0^\varepsilon \;=\; 0. \tag{11.5.6}$$

This SDE can be solved pathwise:

Lemma 11.18 (Control map). *Suppose that $b : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous. Then:*

1) *For any function $\omega \in C([0,1],\mathbb{R}^d)$ there exists a unique function $x \in C([0,1],\mathbb{R}^d)$ such that*

$$x(t) \quad = \quad \int_0^t b(x(s))\,ds + \omega(t) \qquad \forall\, t \in [0,1]. \tag{11.5.7}$$

The function x is absolutely continuous if and only if ω is absolutely continuous, and in this case,

$$x'(t) \quad = \quad b(x(t)) + \omega'(t) \qquad \text{for a.e. } t \in [0,1]. \tag{11.5.8}$$

2) *The **control map** $\mathcal{J} : C([0,1],\mathbb{R}^d) \to C([0,1],\mathbb{R}^d)$ that maps ω to the solution $\mathcal{J}(\omega) = x$ of (11.5.7) is continuous.*

Proof. 1) Existence and uniqueness holds by the classical Picard-Lindelöf Theorem.
2) Suppose that $x = \mathcal{J}(\omega)$ and $\widetilde{x} = \mathcal{J}(\widetilde{\omega})$ are solutions of (11.5.7) w.r.t. driving paths $\omega, \widetilde{\omega} \in C[0,1], \mathbb{R}^d)$. Then for $t \in [0,1]$,

$$\begin{aligned} |x(t) - \widetilde{x}(t)| &= \Big| \int_0^t (b(\omega(s)) - b(\widetilde{\omega}(s)))\, ds + \sqrt{\varepsilon}(\omega(t) - \widetilde{\omega}(t)) \Big| \\ &\leq L \int_0^t |\omega(s) - \widetilde{\omega}(s)|\, ds + \sqrt{\varepsilon}|(t)m - \widetilde{\omega}(t)|. \end{aligned}$$

where $L \in \mathbb{R}_+$ is a Lipschitz constant for b. Gronwall's Lemma now implies

$$|x(t) - \widetilde{x}(t)| \quad \leq \quad \exp(tL)\, \sqrt{\varepsilon}\, ||\omega - \widetilde{\omega}||_{\sup} \qquad \forall\, t \in [0,1],$$

and hence

$$||x - \widetilde{x}||_{\sup} \quad \leq \quad \exp(L)\, \sqrt{\varepsilon}\, ||\omega - \widetilde{\omega}||_{\sup}.$$

This shows that the control map $\mathcal{J}$ is even Lipschitz continuous. □

For $\varepsilon > 0$, the unique solution of the SDE (11.5.6) on $[0,1]$ is given by

$$X^\varepsilon \quad = \quad \mathcal{J}(\sqrt{\varepsilon}B).$$

Since the control map $\mathcal{J}$ is continuous, we can apply Schilder's Theorem to study the large deviations of X^ε as $\varepsilon \downarrow 0$:

Theorem 11.19 (Fredlin & Wentzel 1970, 1984). *If b is Lipschitz continuous then the large deviations principle*

$$\liminf_{\varepsilon \downarrow 0}\ \varepsilon \log\ P[X^\varepsilon \in U] \quad \geq \quad - \inf_{x \in U} I_b(x) \qquad \textit{for any open set } U \subseteq \Omega,$$

$$\liminf_{\varepsilon \downarrow 0}\ \varepsilon \log\ P[X^\varepsilon \in A] \quad \geq \quad - \inf_{x \in A} I_b(x) \qquad \textit{for any closed set } A \subseteq \Omega,$$

holds, where the rate function $I_b : \Omega \to [0, \infty]$ is given by

$$I_b(x) \quad = \quad \begin{cases} \frac{1}{2} \int_0^1 |x'(s) - b(x(s))|^2\, ds & \textit{for } x \in H_{CM}, \\ +\infty & \textit{for } x \in \Omega \setminus H_{CM}. \end{cases}$$

Proof. For any set $A \subseteq \Omega$, we have

$$P[X^\varepsilon \in A] \quad = \quad P[\sqrt{\varepsilon} B \in \mathcal{J}^{-1}(A)] \quad = \quad \mu[\sqrt{\varepsilon} W \in \mathcal{J}^{-1}(A)].$$

If A is open then $\mathcal{J}^{-1}(A)$ is open by continuity of $\mathcal{J}$, and hence

$$\liminf_{\varepsilon \downarrow 0} \, \varepsilon \log \, P[X^\varepsilon \in A] \quad \geq \quad - \inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega))$$

by Theorem 11.16. Similarly, if A is closed then $\mathcal{J}^{-1}(A)$ is closed, and hence the corresponding upper bound holds by Theorem 11.17. Thus it only remains to show that

$$\inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega) \quad = \quad \inf_{x \in A} I_b(x).$$

To this end we note that $\omega \in \mathcal{J}^{-1}(A)$ if and only if $x = \mathcal{J}(\omega) \in A$, and in this case $\omega' = x' - b(x)$. Therefore,

$$\begin{aligned} \inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega) \quad &= \quad \inf_{\omega \in \mathcal{J}^{-1}(A) \cap H_{CM}} \frac{1}{2} \int_0^1 |\omega'(s)|^2 \, ds \\ &= \quad \inf_{x \in A \cap H_{CM}} \frac{1}{2} |x'(s) - b(x(s))|^2 \, ds \quad = \quad \inf_{x \in A} I_b(x). \end{aligned}$$

□

Remark. The large deviation principle in Theorem 11.19 generalizes to non-Lipschitz continuous vector fields b and to SDEs with multiplicative noise. However, in this case, there is no continuous control map that can be used to reduce the statement to Schilder's Theorem. Therefore, a different proof is required, cf. e.g. [10].

Chapter 12

Extensions of Itô calculus

This chapter contains an introduction to some important extensions of Itô calculus and the type of SDE considered so far. We will consider SDE for jump processes driven by white and Poisson noise, Stratonovich calculus and Brownian motion on curved surfaces, stochastic Taylor expansions and numerical methods for SDE, local times and a singular SDE for reflected Brownian motion, as well as stochastic flows.

We start by recalling a crucial martingale inequality that we will apply frequently to derive L^p estimates for semimartingales. For real-valued càdlàg functions $x = (x_t)_{t\geq 0}$ we set

$$x_t^\star \quad := \quad \sup_{s<t} |x_s| \qquad \text{for } \ t > 0, \qquad \text{and} \qquad x_0^\star \ := \ |x_0|.$$

Then the **Burkholder-Davis-Gundy inequality** states that for any $p \in (0, \infty)$ there exist universal constants c_p, $C_p \in (0, \infty)$ such that the estimates

$$c_p \cdot E[[M]_\infty^{p/2}] \quad \leq \quad E[(M_\infty^\star)^p] \quad \leq \quad C_p \cdot E[[M]_\infty^{p/2}] \tag{12.0.1}$$

hold for any continuous local martingale M satisfying $M_0 = 0$, cf. [37]. The inequality shows in particular that for continuous martingales, the $\mathcal{H}^p$ norm, i.e., the L^p norm of $M_\infty^\star$, is equivalent to $E[[M]_\infty^{p/2}]^{1/p}$. Note that for $p = 2$, by Itô's isometry and Doob's L^2 maximal inequality, Equation (12.0.1) holds with $c_p = 1$ and $C_p = 4$. The Burkholder-Davis-Gundy inequality can thus be used to generalize arguments based on Itô's isometry from an L^2 to an L^p setting. This is, for example, important for proving the existence of a continuous stochastic flow corresponding to an SDE, see Section 12.6 below.

In these notes, we only prove an easy special case of the Burkholder-Davis-Gundy inequality that will be sufficient for our purposes: For any $p \in [2, \infty)$,

$$E[(M_T^\star)^p]^{1/p} \quad \leq \quad \sqrt{e/2}\, p\, E[[M]_T^{p/2}]^{1/p} \tag{12.0.2}$$

This estimate also holds for càdlàg local martingales and is proven in Theorem 14.24.

12.1 SDE with jumps

Let $(S, \mathcal{S}, \nu)$ be a σ-finite measure space, and let $d, n \in \mathbb{N}$. Suppose that on a probability space $(\Omega, \mathcal{A}, P)$, we are given an $\mathbb{R}^d$-valued Brownian motion (B_t) and a Poisson random measure $N(dt\ dy)$ over $\mathbb{R}_+ \times S$ with intensity measure $\lambda_{(0,\infty)} \otimes \nu$. Let $(\mathcal{F}_t)$ denote a complete filtration such that (B_t) is an $(\mathcal{F}_t)$ Brownian motion and $N_t(B) = N((0,t] \times B)$ is an $(\mathcal{F}_t)$ Poisson point process, and let

$$\widetilde{N}(dt\ dy) \;=\; N(dt\ dy) - \lambda_{(0,\infty)}(dt)\, \nu(dy).$$

If T is an $(\mathcal{F}_t)$ stopping time then we call a predictable process $(\omega, t) \mapsto G_t(\omega)$ or $(\omega, t, y) \mapsto G_t(y)(\omega)$ defined for finite $t \leq T(\omega)$ and $y \in S$ **locally square integrable** iff there exists an increasing sequence (T_n) of $(\mathcal{F}_t)$ stopping times with $T = \sup T_n$ such that for any n, the trivially extended process $G_t I_{\{t \leq T_n\}}$ is contained in $\mathcal{L}^2(P \otimes \lambda)$, $\mathcal{L}^2(P \otimes \lambda \otimes \nu)$ respectively. For locally square integrable predictable integrands, the stochastic integrals $\int_0^t G_s\, dB_s$ and $\int_{(0,t]\times S} G_s(y)\, \widetilde{N}(ds\ dy)$ respectively are local martingales defined for $t \in [0, T)$.

In this section, we are going to study existence and pathwise uniqueness for solutions of stochastic differential equations of type

$$dX_t \quad = \quad b_t(X)\, dt + \sigma_t(X)\, dB_t + \int_{y \in S} c_{t-}(X, y)\, \widetilde{N}(dt\ dy). \tag{12.1.1}$$

Here $b : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \to \mathbb{R}^n$, $\sigma : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \to \mathbb{R}^{n \times d}$, and $c : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \times S \to \mathbb{R}^n$ are càdlàg functions in the first variable such that b_t, σ_t and c_t

are measurable w.r.t. the σ-algebras $\mathcal{B}_t := \sigma(x \mapsto x_s : s \le t)$, $\mathcal{B}_t \otimes \mathcal{S}$ respectively for any $t \ge 0$. We also assume **local boundedness** of the coefficients, i.e.,

$$\sup_{s<t} \sup_{x:x_t^\star<r} \sup_{y\in S} \left(|b_s(x)| + \|\sigma_s(x)\| + |c_s(x,y)|\right) \; < \; \infty \tag{12.1.2}$$

for any $t, r \in (0,\infty)$.

Note that the assumptions imply that b is progressively measurable, and hence $b_t(x)$ is a measurable function of the path $(x_s)_{s\le t}$ up to time t. Therefore, $b_t(x)$ is also well-defined for càdlàg paths $(x_s)_{s<\zeta}$ with finite life-time ζ provided $\zeta > t$. Corresponding statements hold for σ_t and c_t. Condition (12.1.2) implies in particular that the jump sizes are locally bounded. Locally unbounded jumps could be taken into account by extending the SDE (12.1.1) by an additional term consisting of an integral w.r.t. an uncompensated Poisson point process.

Definition. *Suppose that T is an $(\mathcal{F}_t)$ stopping time.*

1) *A **solution** of the stochastic differential equation (12.1.1) for $t < T$ is a càdlàg $(\mathcal{F}_t)$ adapted stochastic process $(X_t)_{t<T}$ taking values in $\mathbb{R}^n$ such that almost surely,*

$$X_t \;=\; X_0 + \int_0^t b_s(X)\,ds + \int_0^t \sigma_s(X)\,dB_s + \int_{(0,t]\times S} c_{s-}(X,y)\,\widetilde{N}(ds\,dy) \tag{12.1.3}$$

holds for any $t < T$.

2) *A solution $(X_t)_{t<T}$ is called **strong** iff it is adapted w.r.t. the completed filtration $\mathcal{F}_t^0 = \sigma(X_0, \mathcal{F}_t^{B,N})^P$ generated by the initial value, the Brownian motion and the Poisson point process.*

For a *strong solution*, X_t is almost surely a measurable function of the initial value X_0 and the processes $(B_s)_{s\le t}$ and $(N_s)_{s\le t}$ driving the SDE up to time t. In Section 11.1, we saw an example of a solution to an SDE that does not possess this property.

Remark. The stochastic integrals in (12.1.3) are well-defined strict local martingales. Indeed, the local boundedness of the coefficients guarantees local square integrability of the integrands as well as local boundedness of the jumps for the integral w.r.t. $\widetilde{N}$. The process $\sigma_s(X)$ is not necessarily predictable, but observing that $\sigma_s(X(\omega)) = \sigma_{s-}(X(\omega))$ for $P \otimes \lambda$ almost every (ω, s), we may define

$$\int \sigma_s(X)\, dB_s \;:=\; \int \sigma_{s-}(X)\, dB_s.$$

L^p Stability

In addition to the assumptions above, we assume from now on that the coefficients in the SDE (12.1.1) satisfy a **local Lipschitz condition**:

Assumption (A1). For any $t_0 \in \mathbb{R}$, and for any open bounded set $U \subset \mathbb{R}^n$, there exists a constant $L \in \mathbb{R}_+$ such that the following Lipschitz condition $\mathrm{Lip}(t_0, U)$ holds:

$$|b_t(x) - b_t(\widetilde{x})| + ||\sigma_t(x) - \sigma_t(\widetilde{x})|| + \|c_t(x, \bullet) - c_t(\widetilde{x}, \bullet)\|^2_{L^2(\nu)} \;\leq\; L \cdot \sup_{s \leq t} |x_s - \widetilde{x}_s|$$

for any $t \in [0, t_0]$ and $x, \widetilde{x} \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n)$ with $x_s, \widetilde{x}_s \in U$ for $s \leq t_0$.

We now derive an a priori estimate for solutions of (12.1.1) that is crucial for studying existence, uniqueness, and dependence on the initial condition:

Theorem 12.1 (A priori estimate). *Fix $p \in [2, \infty)$ and an open set $U \subseteq \mathbb{R}^n$, and let T be an $(\mathcal{F}_t)$ stopping time. Suppose that (X_t) and $(\widetilde{X}_t)$ are solutions of (12.1.1) taking values in U for $t < T$, and let*

$$\varepsilon_t \quad := \quad E\left[\sup_{s < t \wedge T} |X_s - \widetilde{X}_s|^p\right].$$

If the Lipschitz condition $\mathrm{Lip}(t_0, U)$ holds then there exists a finite constant $C \in \mathbb{R}_+$ depending only on p and on the Lipschitz constant L such that for any $t \leq t_0$,

$$\varepsilon_t \quad \leq \quad C \cdot \left(\varepsilon_0 + \int_0^t \varepsilon_s\, ds\right), \quad \textit{and} \tag{12.1.4}$$

$$\varepsilon_t \quad \leq \quad C \cdot e^{Ct}\, \varepsilon_0. \tag{12.1.5}$$

Proof. We only prove the assertion for $p = 2$. For $p > 2$, the proof can be carried out in a similar way by relying on Burkholder's inequality instead of Itô's isometry.

Clearly, (12.1.5) follows from (12.1.4) by Gronwell's lemma. To prove (12.1.4), note that

$$X_t \;=\; X_0 + \int_0^t b_s(X)\,ds + \int_0^t \sigma_s(X)\,dB_s + \int_{(0,t]\times S} c_{s-}(X,y)\,\widetilde{N}(ds\,dy) \qquad \forall\, t < T,$$

and an analogue equation holds for $\widetilde{X}$. Hence for $t \le t_0$,

$$(X - \widetilde{X})^\star_{t\wedge T} \quad\le\quad \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}, \qquad \text{where} \tag{12.1.6}$$

$$\begin{aligned}
\mathrm{I} \quad &= \quad |X_0 - \widetilde{X}_0|,\\
\mathrm{II} \quad &= \quad \int_0^{t\wedge T} |b_s(X) - b_s(\widetilde{X})|\,ds,\\
\mathrm{III} \quad &= \quad \sup_{u<t\wedge T} \Big|\int_0^u (\sigma_s(X) - \sigma_s(\widetilde{X}))\,dB_s\Big|, \qquad \text{and}\\
\mathrm{IV} \quad &= \quad \sup_{u<t\wedge T} \Big|\int_{(0,u]\times S} (c_{s-}(X,y) - c_{s-}(\widetilde{X},y))\,\widetilde{N}(ds\,dy)\Big|.
\end{aligned}$$

The squared L^2-norms of the first two expressions are bounded by

$$\begin{aligned}
E[I^2] \quad &= \quad \varepsilon_0, \qquad \text{and}\\
E[\mathrm{II}^2] \quad \le \quad L^2 t\, E\Big[\int_0^{t\wedge T} (X - \widetilde{X})^{\star\,2}_s\,ds\Big] \quad &\le \quad L^2 t \int_0^t \varepsilon_s\,ds.
\end{aligned}$$

Denoting by M_u and K_u the stochastic integrals in III and IV respectively, Doob's inequality and Itô's isometry imply

$$\begin{aligned}
E[\mathrm{III}^2] \quad &= \quad E[M^{\star\,2}_{t\wedge T}] \quad \le \quad 4E[M^2_{t\wedge T}]\\
&= \quad 4E\Big[\int_0^{t\wedge T} ||\sigma_s(X) - \sigma_s(\widetilde{X})||^2\,ds\Big] \quad \le \quad 4L^2 \int_0^t \varepsilon_s\,ds,
\end{aligned}$$

$$\begin{aligned}
E[\mathrm{IV}^2] \quad &= \quad E[K^{\star\,2}_{t\wedge T}] \quad \le \quad 4E[K^2_{t\wedge T}]\\
&= \quad 4E\Big[\int_0^{t\wedge T}\int |c_{s-}(X,y) - c_{s-}(\widetilde{X},y)|^2\,\nu(dy)\,ds\Big] \quad \le \quad 4L^2 \int_0^t \varepsilon_s\,ds.
\end{aligned}$$

The assertion now follows since by (12.1.6),

$$\varepsilon_t \quad = \quad E\big[(X-\widetilde{X})^{\star\,2}_{t\wedge T}\big] \quad \leq \quad 4\cdot E[\mathrm{I}^2+\mathrm{II}^2+\mathrm{III}^2+\mathrm{IV}^2\big].$$

□

The a priori estimate shows in particular that under a global Lipschitz condition, solutions depend continuously on the initial condition in mean square. Moreover, it implies pathwise uniqueness under a local Lipschitz condition:

Corollary 12.2 (Pathwise uniqueness). *Suppose that Assumption (A1) holds. If (X_t) and $(\widetilde{X}_t)$ are strong solutions of (12.0.1) with $X_0=\widetilde{X}_0$ almost surely then*

$$P\big[X_t \;=\; \widetilde{X}_t \quad \textit{for any } t\big] \quad = \quad 1.$$

Proof. For any open bounded set $U\subset\mathbb{R}^n$ and $t_0\in\mathbb{R}_+$, the a priori estimate in Theorem 12.1 implies that X and $\widetilde{X}$ coincide almost surely on $[0,t_0\wedge T_{U^c})$ where T_{U^c} denotes the first exit time from U. □

Existence of strong solutions

To prove existence of strong solutions, we need an additional assumption:

Assumption (A2). For any $t_0\in\mathbb{R}_+$**,**

$$\sup_{t<t_0}\int |c_t(0,y)|^2\,\nu(dy) \quad < \quad \infty.$$

Here 0 **denotes the constant path** $x\equiv 0$ **in** $\mathcal{D}(\mathbb{R}_+,\mathbb{R}^n)$**.**

Note that the assumption is always satisfied if $c\equiv 0$.

Remark (Linear growth condition). If both (A2) and a global Lipschitz condition $\mathrm{Lip}(t_0,\mathbb{R}^n)$ hold then there exists a finite constant $C(t_0)$ such that for any $x\in\mathcal{D}(\mathbb{R}_+,\mathbb{R}^n)$,

$$\sup_{t<t_0}\Big(|b_t(x)|^2+||\sigma_t(x)||^2+\int |c_t(x,y)|^2\,\nu(dy)\Big) \quad \leq \quad C(t_0)\cdot(1+x^\star_{t_0})^2. \tag{12.1.7}$$

Theorem 12.3 (Itô). *Let $\xi : \Omega \to \mathbb{R}^n$ be a random variable that is independent of the Brownian motion B and the Poisson random measure N.*

1) *Suppose that the local Lipschitz condition (A1) and (A2) hold. Then (12.0.1) has a strong solution $(X_t)_{t<\zeta}$ with initial condition $X_0 = \xi$ that is defined up to the explosion time*

$$\zeta \quad = \quad \sup T_k, \qquad \textit{where } \ T_k \quad = \quad \inf\{t \geq 0 : |X_t| \geq k\}.$$

2) *If, moreover, the global Lipschitz condition $\mathrm{Lip}(t_0, \mathbb{R}^n)$ holds for any $t_0 \in \mathbb{R}_+$, then $\zeta = \infty$ almost surely.*

Proof of 12.3. We first prove existence of a global strong solution $(X_t)_{t\in[0,\infty)}$ assuming (A2) and a global Lipschitz condition $\mathrm{Lip}(t_0, \mathbb{R}^n)$ for any $t_0 \in \mathbb{R}_+$. The first assertion will then follow by localization.

For proving global existence we may assume w.l.o.g. that ξ is bounded and thus square integrable. We then construct a sequence (X^n) of approximate solutions to (12.0.1) by a **Picard-Lindelöf iteration**, i.e., for $t \geq 0$ and $n \in \mathbb{Z}_+$ we define inductively

$$\begin{aligned} X_t^0 &:= \xi, \qquad\qquad (12.1.8) \\ X_t^{n+1} &:= \xi + \int_0^t b_s(X^n)\, ds + \int_0^t \sigma_s(X^n)\, dB_s + \int_{(0,t]\times S} c_{s-}(X^n, y)\, \widetilde{N}(ds\, dy). \end{aligned}$$

Fix $t_0 \in [0, \infty)$. We will show below that Assumption (A2) and the global Lipschitz condition imply that

(i) for any $n \in \mathbb{N}$, X^n is a square integrable $(\mathcal{F}_t^0)$ semimartingale on $[0, t_0]$ (i.e., the sum of a square integrable martingale and an adapted process with square integrable total variation), and

(ii) there exists a finite constant $C(t_0)$ such that the mean square deviations

$$\Delta_t^n \quad := \quad E[(X^{n+1} - X^n)_t^{\star\, 2}].$$

of the approximations X^n and X^{n+1} satisfy

$$\Delta_t^{n+1} \quad \leq \quad C(t_0) \int_0^t \Delta_s^n \, ds \qquad \text{for any } n \geq 0 \text{ and } t \leq t_0.$$

Then, by induction,

$$\Delta_t^n \quad \leq \quad C(t_0)^n \, \frac{t^n}{n!} \, \Delta_t^0 \qquad \text{for any } n \in \mathbb{N} \text{ and } t \leq t_0.$$

In particular, $\sum_{n=1}^{\infty} \Delta_{t_0}^n < \infty$. An application of the Borel-Cantelli Lemma now shows that the limit $X_s = \lim_{n\to\infty} X_s^n$ exists uniformly for $s \in [0, t_0]$ with probability one. Moreover, X is a fixed point of the Picard-Lindelöf iteration, and hence a solution of the SDE (12.0.1). Since t_0 has been chosen arbitrarily, the solution is defined almost surely on $[0, \infty)$, and by construction it is adapted w.r.t. the filtration $(\mathcal{F}_t^0)$.

We now show by induction that Assertion (i) holds. If X^n is a square integrable $(\mathcal{F}_t^0)$ semimartingale on $[0, t_0]$ then, by the linear growth condition (12.1.7), the process $|b_s(X^n)|^2 + ||\sigma_s(X^n)||^2 + \int |c_s(X^n, y)|^2 \, \nu(dy)$ is integrable w.r.t. the product measure $P \otimes \lambda_{(0,t_0)}$. Therefore, by Itô's isometry, the integrals on the right hand side of (12.1.8) all define square integrable $(\mathcal{F}_t^0)$ semimartingales, and thus X^{n+1} is a square integrable $(\mathcal{F}_t^0)$ semimartingale, too.

Assertion (ii) is a consequence of the global Lipschitz condition. Indeed, by the Cauchy-Schwarz inequality, Itô's isometry and $\mathrm{Lip}(t_0, \mathbb{R}^n)$, there exists a finite constant $C(t_0)$ such that

$$\begin{aligned}
\Delta_t^{n+1} &= E\left[\left(X^{n+2} - X^{n+1}\right)_t^{\star\, 2}\right] \\
&\leq 3t\, E\left[\int_0^t \left|b_s(X^{n+1}) - b_s(X^n)\right|^2 \, ds\right] + 3\, E\left[\int_0^t \left\|\sigma_s(X^{n+1}) - \sigma_s(X^n)\right\|^2 \, ds\right] \\
&\quad + 3\, E\left[\int_0^t \int \left|c_s(X^{n+1}, y) - c_s(X^n, y)\right|^2 \, \nu(dy)\, ds\right] \\
&\leq C(t_0) \int_0^t \Delta_s^n \, ds \qquad\qquad \text{for any } n \geq 0 \text{ and } t \leq t_0.
\end{aligned}$$

This completes the proof of global existence under a global Lipschitz condition.

Finally, suppose that the coefficients b, σ and c only satisfy the local Lipschitz condition (A1). Then for $k \in \mathbb{N}$ and $t_0 \in \mathbb{R}_+$, we can find functions b^k, σ^k and c^k that are globally

Lipschitz continuous and that agree with b, σ and c on paths (x_t) taking values in the ball $B(0,k)$ for $t \leq t_0$. The solution $X^{(k)}$ of the SDE with coefficients b^k, σ^k, c^k is then a solution of (12.0.1) up to $t \wedge T_k$ where T_k denotes the first exit time of $X^{(k)}$ from $B(0,k)$. By pathwise uniqueness, the local solutions obtained in this way are consistent. Hence they can be combined to construct a solution of (12.0.1) that is defined up to the explosion time $\zeta = \sup T_k$. □

Non-explosion criteria

Theorem 12.3 shows that under a global Lipschitz and linear growth condition on the coefficients, the solution to (12.0.1) is defined for all times with probability one. However, this condition is rather restrictive, and there are much better criteria to prove that the explosion time ζ is almost surely infinite. Arguably the most generally applicable non-explosion criteria are those based on **stochastic Lyapunov functions**. Consider for example an SDE of type

$$dX_t \quad = \quad b(X_t)\,dt + \sigma(X_t)\,dB_t \tag{12.1.9}$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ are locally Lipschitz continuous, and let

$$\mathcal{L} \quad = \quad \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x)\cdot\nabla, \qquad a(x) \;=\; \sigma(x)\sigma(x)^T,$$

denote the corresponding generator.

Theorem 12.4 (Lyapunov condition for non-explosion). *Suppose that there exists a function $\phi \in C^2(\mathbb{R}^n)$ such that*

(i) $\phi(x) \geq 0$ *for any $x \in \mathbb{R}^n$,*

(ii) $\phi(x) \to \infty$ *as $|x| \to \infty$, and*

(iii) $\mathcal{L}\phi \leq \lambda\phi$ *for some $\lambda \in \mathbb{R}_+$.*

Then the strong solution of (12.0.1) with initial value $x_0 \in \mathbb{R}^n$ exists up to $\zeta = \infty$ almost surely.

Proof. We first remark that by (iii), $Z_t := \exp(-\lambda t)\phi(X_t)$ is a supermartingale up to the first exit time T_k of the local solution X from a ball $B(0,k) \subset \mathbb{R}^n$. Indeed, by the product rule and the Itô-Doeblin formula,

$$dZ \;=\; -\lambda e^{-\lambda t}\phi(X)\,dt \;+\; e^{-\lambda_t}\,d\phi(X) \;=\; dM \;+\; e^{-\lambda t}(\mathcal{L}\phi - \lambda\phi)(X)\,dt$$

holds on $[0, T_k]$ with a martingale M up to T_k.

Now we fix $t \geq 0$. Then, by the Optional Stopping Theorem and by Condition (i),

$$\begin{aligned}
\varphi(x_0) \;&=\; E\left[\varphi(X_0)\right] \;\geq\; E\left[\exp(-\lambda(t\wedge T_k))\,\varphi(X_{t\wedge T_k})\right] \\
&\geq\; E\left[\exp(-\lambda t)\,\varphi(X_{T_k})\,;\,T_k \leq t\right] \\
&\geq\; \exp(-\lambda t)\,\inf_{|y|=k}\phi(y)\;P\left[T_k \leq t\right]
\end{aligned}$$

for any $k \in \mathbb{N}$. As $k \to \infty$, $\inf_{|y|=k}\phi(y) \to \infty$ by (ii). Therefore,

$$P[\sup T_k \leq t] \;=\; \lim_{k\to\infty} P[T_k \leq t] \;=\; 0$$

for any $t \geq 0$, i.e., $\zeta = \sup T_k = \infty$ almost surely. □

By applying the theorem with the function $\varphi(x) = 1 + |x|^2$ we obtain:

Corollary 12.5. *If there exists $\lambda \in \mathbb{R}_+$ such that*

$$2x\cdot b(x) + \mathrm{tr}(a(x)) \quad\leq\quad \lambda\cdot(1+|x|^2) \qquad \textit{for any } x \in \mathbb{R}^n$$

then $\zeta = \infty$ almost surely.

Note that the condition in the corollary is satisfied if

$$\frac{x}{|x|}\cdot b(x) \quad\leq\quad \text{const.}\cdot|x| \qquad \text{and} \qquad \mathrm{tr}\,a(x) \leq \text{const.}\cdot|x|^2$$

for sufficiently large $x \in \mathbb{R}^n$, i.e., if the outward component of the drift is growing at most linearly, and the trace of the diffusion matrix is growing at most quadratically.

12.2 Stratonovich differential equations

Replacing Itô by Statonovich integrals has the advantage that the calculus rules (product rule, chain rule) take the same form as in classical differential calculus. This is useful for explicit computations (Doss-Sussman method), for approximating solutions of SDE by solutions of ordinary differential equations, and in stochastic differential geometry. For simplicity, we only consider Stratonovich calculus for continuous semimartingales, cf. [36] for the discontinuous case.

Let X and Y be continuous semimartingales on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$.

Definition (Fisk-Stratonovich integral). *The Stratonovich integral $\int X \circ dY$ is the continuous semimartingale defined by*

$$\int_0^t X_s \circ dY_s \quad := \quad \int_0^t X_s \, dY_s + \frac{1}{2}[X, Y]_t \qquad \text{for any } t \geq 0.$$

Note that a Stratonovich integral w.r.t. a martingale is not a local martingale in general. The Stratonovich integral is a limit of trapezoidal Riemann sum approximations:

Lemma 12.6. *If (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\operatorname{mesh}(\pi_n) \to 0$ then*

$$\int_0^t X_s \circ dY_s \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \frac{X_s + X_{s'\wedge t}}{2} \left(Y_{s'\wedge t} - Y_s\right) \qquad \text{in the ucp sense.}$$

Proof. This follows since $\int_0^t X \, dY = \text{ucp-}\lim \sum_{s<t} X_s \, (Y_{s'\wedge t} - Y_s)$ and $[X, Y]_t = \text{ucp-}\lim \sum_{s<t} (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s)$ by the results above. □

Itô-Stratonovich formula

For Stratonovich integrals w.r.t. continuous semimartingales, the classical chain rule holds:

Theorem 12.7. *Let $X = (X^1, \ldots, X^d)$ with continuous semimartingales X^i. Then for any function $F \in C^2(\mathbb{R}^d)$,*

$$F(X_t) - F(X_0) \quad = \quad \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x^i}(X_s) \circ dX_s^i \qquad \forall\, t \geq 0. \tag{12.2.1}$$

Proof. To simplify the proof we assume $F \in C^3$. Under this condition, (12.2.1) is just a reformulation of the Itô rule

$$F(X_t) - F(X_0) \quad = \quad \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x^i}(X_s)\, dX_s^i + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s)\, d[X^i, X^j]_s. \tag{12.2.2}$$

Indeed, applying Itô's rule to the C^2 function $\frac{\partial F}{\partial x^i}$ shows that

$$\frac{\partial F}{\partial x^i}(X_t) \quad = \quad A_t + \sum_j \int \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s)\, dX_s^j$$

for some continuous finite variation process A. Hence the difference between the Statonovich integral in (12.2.1) and the Itô integral in (12.2.2) is

$$\frac{1}{2}\left[\frac{\partial F}{\partial x^i}(X), X^i\right]_t \quad = \quad \frac{1}{2} \sum_j \int \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s)\, d[X^j, X^i]_s.$$

□

Remark. For the extension of the proof to C^2 functions F see e.g. [36], where also a generalization to càdlàg semimartingales is considered.

The product rule for Stratonovich integrals is a special case of the chain rule:

Corollary 12.8. *For continuous semimartingales X, Y,*

$$X_t Y_t - X_0 Y_0 \quad = \quad \int_0^t X_s \circ dY_s + \int_0^t Y_s \circ dX_s \qquad \forall\, t \geq 0.$$

Exercise (Associative law). Prove an associative law for Stratonovich integrals.

Stratonovich SDE

Since Stratonovich integrals differ from the corresponding Itô integrals only by the covariance term, equations involving Stratonovich integrals can be rewritten as Itô equations and vice versa, provided the coefficients are sufficiently regular. We consider a Stratonovich SDE in $\mathbb{R}^d$ of the form

$$\circ\, dX_t \quad = \quad b(X_t)\, dt + \sum_{k=1}^{d} \sigma_k(X_t) \circ dB_t^k, \qquad X_0 \;=\; x_0, \tag{12.2.3}$$

with $x_0 \in \mathbb{R}^n$, continuous vector fields $b, \sigma_1, \ldots, \sigma_d \in C(\mathbb{R}^n, \mathbb{R}^n)$, and an $\mathbb{R}^d$-valued Brownian motion (B_t).

Exercise (Stratonovich to Itô conversion). 1) Prove that for $\sigma_1, \ldots, \sigma_d \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, the Stratonovich SDE (12.2.3) is equivalent to the Itô SDE

$$dX_t \quad = \quad \widetilde{b}(X_t)\, dt + \sum_{k=1}^{d} \sigma_k(X_t)\, dB_t^k, \qquad X_0 \;=\; x_0, \tag{12.2.4}$$

where

$$\widetilde{b} \quad := \quad b + \frac{1}{2} \sum_{k=1}^{d} \sigma_k \cdot \nabla \sigma_k.$$

2) Conclude that if $\widetilde{b}$ and $\sigma_1, \ldots, \sigma_d$ are Lipschitz continuous, then there is a unique strong solution of (12.2.3).

Theorem 12.9 (Martingale problem for Stratonovich SDE). *Let $b \in C(\mathbb{R}^n, \mathbb{R}^n)$ and $\sigma_1, \ldots, \sigma_d \in C^2(\mathbb{R}^n, \mathbb{R}^n)$, and suppose that $(X_t)_{t\geq 0}$ is a solution of (12.2.3) on a given setup $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$. Then for any function $F \in C^3(\mathbb{R}^n)$, the process*

$$\begin{aligned} M_t^F \quad &= \quad F(X_t) - F(X_0) - \int_0^t (\mathcal{L}F)(X_s)\, ds, \\ \mathcal{L}F \quad &= \quad \frac{1}{2} \sum_{k=1}^{d} \sigma_k \cdot \nabla(\sigma_k \cdot \nabla F) + b \cdot \nabla F, \end{aligned}$$

is a local $(\mathcal{F}_t^P)$ martingale.

Proof. By the Stratonovich chain rule and by (12.2.3),

$$\begin{aligned} F(X_t) - F(X_0) &= \int_0^t \nabla F(X) \cdot \circ dX \\ &= \int_0^t (b \cdot \nabla F)(X)\, dt + \sum_k \int_0^t (\sigma_k \cdot \nabla F)(X) \circ dB^k. \end{aligned} \tag{12.2.5}$$

By applying this formula to $\sigma_k \cdot \nabla F$, we see that

$$(\sigma_k \cdot \nabla F)(X_t) = A_t + \sum_l \int \sigma_l \cdot \nabla(\sigma_k \cdot \nabla F)(X)\, dB^l$$

with a continuous finite variation process (A_t). Hence

$$\begin{aligned} \int_0^t (\sigma_k \cdot \nabla F)(X) \circ dB^k &= \int_0^t (\sigma_k \cdot \nabla F)(X)\, dB^k + [(\sigma_k \cdot \nabla F)(X), B^k]_t \\ &= \text{local martingale } + \int_0^t \sigma_k \cdot \nabla(\sigma_k \cdot \nabla F)(X) dt. \end{aligned} \tag{12.2.6}$$

The assertion now follows by (12.2.5) and (12.2.6). □

The theorem shows that the generator of a diffusion process solving a Stratonovich SDE is in sum of squares form. In geometric notation, one briefly writes b for the derivative $b \cdot \nabla$ in the direction of the vector field b. The generator then takes the form

$$\mathcal{L} = \frac{1}{2} \sum_k \sigma_k^2 + b$$

Brownian motion on hypersurfaces

One important application of Stratonovich calculus is stochastic differential geometry. Itô calculus can not be used directly for studying stochastic differential equations on manifolds, because the classical chain rule is essential for ensuring that solutions stay on the manifold if the driving vector fields are tangent vectors. Instead, one considers Stratonovich equations. These are converted to Itô form when computing expectation values. To avoid differential geometric terminology, we only consider Brownian motion on a hypersurface in $\mathbb{R}^{n+1}$, cf. [38], [20] and [22] for stochastic calculus on more general Riemannian manifolds.

Let $f \in C^\infty(\mathbb{R}^{n+1})$ and suppose that $c \in \mathbb{R}$ is a regular value of f, i.e., $\nabla f(x) \neq 0$ for any $x \in f^{-1}(c)$. Then by the implicit function theorem, the level set

$$M_c \quad = \quad f^{-1}(c) \quad = \quad \left\{x \in \mathbb{R}^{n+1} \; : \; f(x) = c\right\}$$

is a smooth n-dimensional submanifold of $\mathbb{R}^{n+1}$. For example, if $f(x) = |x|^2$ and $c = 1$ then M_c is the n-dimensional unit sphere S^n.

For $x \in M_c$, the vector

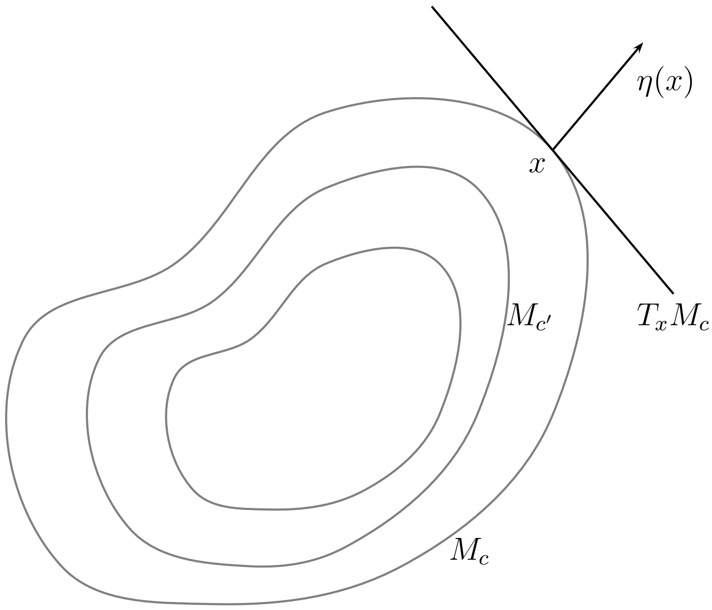
$$\mathfrak{n}(x) \quad = \quad \frac{\nabla f(x)}{|\nabla f(x)|} \quad \in S^n$$

is the **unit normal** to M_c at x. The **tangent space** to M_c at x is the orthogonal complement

$$T_x M_c \quad = \quad \operatorname{span}\left\{\mathfrak{n}(x)\right\}^\perp .$$

Let $P(x) : \mathbb{R}^{n+1} \to T_x M_c$ denote the orthogonal projection onto the tangent space w.r.t. the Euclidean metric, i.e.,

$$P(x)v \quad = \quad v - v \cdot \mathfrak{n}(x)\, \mathfrak{n}(x), \quad v \in \mathbb{R}^{n+1}.$$



For $k \in \{1, \dots, n+1\}$, we set $P_k(x) = P(x)e_k$.

Definition. *A* ***Brownian motion on the hypersurface*** $\boldsymbol{M_c}$ *with initial value* $x_0 \in M_c$ *is a solution* (X_t) *of the Stratonovich SDE*

$$\circ\, dX_t \;=\; P(X_t) \circ dB_t \;=\; \sum_{k=1}^{n+1} P_k(X_t) \circ dB_t^k, \qquad X_0 \;=\; x_0, \tag{12.2.7}$$

with respect to a Brownian motion (B_t) *on* $\mathbb{R}^{n+1}$.

We now assume for simplicity that M_c is compact. Then, since c is a regular value of f, the vector fields P_k are smooth with bounded derivatives of all orders in a neighbourhood U of M_c in $\mathbb{R}^{n+1}$. Therefore, there exists a unique strong solution of the SDE (12.2.7) in $\mathbb{R}^{n+1}$ that is defined up to the first exit time from U. Indeed, this solution stays on the submanifold M_c for all times:

Theorem 12.10. *If X is a solution of (12.2.7) with $x_0 \in M_c$ then almost surely, $X_t \in M_c$ for any $t \geq 0$.*

The proof is very simple, but it relies on the classical chain rule in an essential way:

Proof. We have to show that $f(X_t)$ is constant. This is an immediate consequence of the Stratonovich formula:

$$f(X_t) - f(X_0) \;=\; \int_0^t \nabla f(X_s) \cdot \circ dX_s \;=\; \sum_{k=1}^{n+1} \int_0^t \nabla f(X_s) \cdot P_k(X_s) \circ dB_s^k \;=\; 0$$

since $P_k(x)$ is orthogonal to $\nabla f(x)$ for any x. □

Although we have defined Brownian motion on the Riemannian manifold M_c in a non-intrinsic way, one can verify that it actually is an intrinsic object and does not depend on the embedding of M_c into $\mathbb{R}^{n+1}$ that we have used. We only convince ourselves that the corresponding generator is an intrinsic object. By Theorem 12.9, the Brownian motion (X_t) constructed above is a solution of the martingale problem for the operator

$$\mathcal{L} \quad = \quad \frac{1}{2}\sum_{k=1}^{n+1}(P_k \cdot \nabla)P_k \cdot \nabla \quad = \quad \frac{1}{2}\sum_{k=1}^{n+1} P_k^2.$$

From differential geometry it is well-known that this operator is $\frac{1}{2}\Delta_{M_c}$ where Δ_{M_c} denotes the (intrinsic) **Laplace-Beltrami operator** on M_c.

Exercise (Itô SDE for Brownian motion on M_c). Prove that the SDE (12.2.7) can be written in Itô form as

$$dX_t \quad = \quad P(X_t)\, dB_t - \frac{1}{2}\kappa(X_t)\mathfrak{n}(X_t)\, dt$$

where $\kappa(x) = \frac{1}{n} \operatorname{div} \mathfrak{n}(x)$ is the mean curvature of M_c at x.

Doss-Sussmann method

Stratonovich calculus can also be used to obtain explicit solutions for stochastic differential equations in $\mathbb{R}^n$ that are driven by a **one-dimensional** Brownian motion (B_t). We consider the SDE

$$\circ\, dX_t \quad = \quad b(X_t)\, dt \;+\; \sigma(X_t) \circ dB_t, \qquad X_0 \;=\; a, \tag{12.2.8}$$

where $a \in \mathbb{R}^n$, $b : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous and $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is C^2 with bounded derivatives. Recall that (12.2.8) is equivalent to the Itô SDE

$$dX_t \quad = \quad \big(b + \frac{1}{2}\sigma \cdot \nabla\sigma\big)(X_t)\, dt + \sigma(X_t)\, dB_t, \qquad X_0 = a. \tag{12.2.9}$$

We first determine an explicit solution in the case $b \equiv 0$ by the ansatz $X_t = F(B_t)$ where $F \in C^2(\mathbb{R}, \mathbb{R}^n)$. By the Stratonovich rule,

$$\circ dX_t \quad = \quad F'(B_t) \circ dB_t \quad = \quad \sigma(F(B_t)) \circ dB_t$$

provided F is a solution of the ordinary differential equation

$$F'(s) \quad = \quad \sigma(F(s)). \tag{12.2.10}$$

Hence a solution of (12.2.8) with initial condition $X_0 = a$ is given by

$$X_t \quad = \quad F(B_t, a)$$

where $(s,x) \mapsto F(s,x)$ is the **flow** of the vector field σ, i.e., $F(\cdot, a)$ is the unique solution of (12.2.10) with initial condition a.

Recall from the theory of ordinary differential equations that the flow of a vector field σ as above defines a diffeomorphism $a \mapsto F(s,a)$ for any $s \in \mathbb{R}$. To obtain a solution of (12.2.8) in the general case, we try the "variation of constants" ansatz

$$X_t \quad = \quad F(B_t, C_t) \tag{12.2.11}$$

with a continuous semimartingale (C_t) satisfying $C_0 = a$. In other words: we make a time-dependent coordinate transformation in the SDE that is determined by the flow F and the driving Brownian path (B_t). By applying the chain rule to (12.2.11), we obtain

$$\begin{aligned} \circ dX_t \quad &= \quad \frac{\partial F}{\partial s}(B_t, C_t) \circ dB_t \; + \; \frac{\partial F}{\partial x}(B_t, C_t) \circ dC_t \\ &= \quad \sigma(X_t) \circ dB_t \; + \; \frac{\partial F}{\partial x}(B_t, C_t) \circ dC_t \end{aligned}$$

where $\frac{\partial F}{\partial x}(s, \cdot)$ denotes the Jacobi matrix of the diffeomorphism $F(s, \cdot)$. Hence (X_t) is a solution of the SDE (12.2.8) provided (C_t) is almost surely absolutely continuous with derivative

$$\frac{d}{dt} C_t \quad = \quad \frac{\partial F}{\partial x}(B_t, C_t)^{-1} \, b(F(B_t, C_t)). \tag{12.2.12}$$

For every given ω, the equation (12.2.12) is an ordinary differential equation for $C_t(\omega)$ which has a unique solution. Working out these arguments in detail yields the following result:

Theorem 12.11 (Doss 1977, Sussmann 1978). *Suppose that $b : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous and $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is C^2 with bounded derivatives. Then the flow F of the vector field σ is well-defined, $F(s, \cdot)$ is a C^2 diffeomorphism for any $s \in \mathbb{R}$, and the equation (12.2.12) has a unique pathwise solution $(C_t)_{t \geq 0}$ satisfying $C_0 = a$. Moreover, the process $X_t = F(B_t, C_t)$ is the unique strong solution of the equation (12.2.8), (12.2.9) respectively.*

We refer to [25] for a detailed proof.

Exercise (Computing explicit solutions). Solve the following Itô stochastic differential equations explicitly:

$$dX_t = \frac{1}{2}X_t\,dt + \sqrt{1+X_t^2}\,dB_t, \qquad X_0 = 0, \tag{12.2.13}$$

$$dX_t = X_t(1+X_t^2)\,dt + (1+X_t^2)\,dB_t, \qquad X_0 = 1. \tag{12.2.14}$$

Do the solutions explode in finite time?

Exercise (Variation of constants). We consider nonlinear stochastic differential equations of the form

$$dX_t = f(t, X_t)\,dt + c(t)X_t\,dB_t, \qquad X_0 = x,$$

where $f : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ and $c : \mathbb{R}^+ \to \mathbb{R}$ are continuous (deterministic) functions. Proceed as follows :

a) Find an explicit solution Z_t of the equation with $f \equiv 0$.

b) To solve the equation in the general case, use the Ansatz

$$X_t = C_t \cdot Z_t\,.$$

Show that the SDE gets the form

$$\frac{dC_t(\omega)}{dt} = f(t, Z_t(\omega) \cdot C_t(\omega))/Z_t(\omega)\,; \qquad C_0 = x. \tag{12.2.15}$$

Note that for each $\omega \in \Omega$, this is a *deterministic* differential equation for the function $t \mapsto C_t(\omega)$. We can therefore solve (12.2.15) with ω as a parameter to find $C_t(\omega)$.

c) Apply this method to solve the stochastic differential equation

$$dX_t = \frac{1}{X_t}\,dt + \alpha X_t\,dB_t\,; \qquad X_0 = x > 0\,,$$

where α is constant.

d) Apply the method to study the solution of the stochastic differential equation

$$dX_t = X_t^\gamma\,dt + \alpha X_t\,dB_t\,; \qquad X_0 = x > 0\,,$$

where α and γ are constants. For which values of γ do we get explosion?

Wong Zakai approximations of SDE

A natural way to approximate the solution of an SDE driven by a Brownian motion is to replace the Brownian motion by a smooth approximation. The resulting equation can then be solved pathwise as an ordinary differential equation. It turns out that the limit of this type of approximations as the driving smoothed processes converge to Brownian motion will usually solve the corresponding Stratonovich equation.

Suppose that $(B_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^d$ with $B_0 = 0$. For notational convenience we define $B_t := 0$ for $t < 0$. We approximate B by the smooth processes

$$B^{(k)} \quad := \quad B \star \phi_{1/k}, \qquad \phi_\varepsilon(t) \quad = \quad (2\pi\varepsilon)^{-1/2} \exp\big(-\frac{t^2}{2\varepsilon}\big).$$

Other smooth approximations could be used as well, cf. [25] and [23]. Let $X^{(k)}$ denote the unique solution to the ordinary differential equation

$$\frac{d}{dt} X_t^{(k)} \quad = \quad b(X_t^{(k)}) + \sigma(X_t^{(k)})\, \frac{d}{dt} B_t^{(k)}, \qquad X_0^{(k)} \;=\; a \tag{12.2.16}$$

with coefficients $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$.

Theorem 12.12 (Wong, Zakai 1965). *Suppose that b is C^1 with bounded derivatives and σ is C^2 with bounded derivatives. Then almost surely as $k \to \infty$,*

$$X_t^{(k)} \quad \longrightarrow \quad X_t \qquad \textit{uniformly on compact intervals},$$

where (X_t) is the unique solution of the Stratonovich equation

$$\circ dX_t \quad = \quad b(X_t)\, dt + \sigma(X_t) \circ dB_t, \qquad X_0 \;=\; a.$$

If the driving Brownian motion is one-dimensional, there is a simple proof based on the Doss-Sussman representation of solutions. This shows that $X^{(k)}$ and X can be represented in the form $X_t^{(k)} = F(B_t^{(k)}, C_t^{(k)})$ and $X_t = F(B_t, C_t)$ with the flow F of the same vector field σ, and the processes $C^{(k)}$ and C solving (12.2.12) w.r.t. $B^{(k)}$, B respectively. Therefore, it is not difficult to verify that almost surely, $X^{(k)} \to X$ uniformly on compact time intervals, cf. [25]. The proof in the more interesting general case is much more involved, cf. e.g. Ikeda & Watanabe [23, Ch. VI, Thm. 7.2].

12.3 Stochastic Taylor expansions

In the next section we will study numerical schemes for Itô stochastic differential equations of type

$$dX_t \;=\; b(X_t)\,dt \;+\; \sum_{k=1}^{d} \sigma_k(X_t)\,dB_t^k \tag{12.3.1}$$

in $\mathbb{R}^N$, $N \in \mathbb{N}$. A key tool for deriving and analyzing such schemes are stochastic Taylor expansions that are introduced in this section.

We will assume throughout the next two sections that the coefficients $b, \sigma_1, \ldots, \sigma_d$ are C^∞ vector fields on $\mathbb{R}^N$, and $B = (B^1, \ldots, B^d)$ is a d-dimensional Brownian motion. Below, it will be convenient to set

$$B_t^0 \;:=\; t.$$

A solution of (12.3.1) satisfies

$$X_{t+h} \;=\; X_t \;+\; \int_t^{t+h} b(X_s)\,ds \;+\; \sum_{k=1}^{d} \int_t^{t+h} \sigma_k(X_s)\,dB_s^k \tag{12.3.2}$$

for any $t, h \geq 0$. By approximating $b(X_s)$ and $\sigma_k(X_s)$ in (12.3.2) by $b(X_t)$ and $\sigma_k(X_t)$ respectively, we obtain an Euler approximation of the solution with step size h. Similarly, higher order numerical schemes can be obtained by approximating $b(X_s)$ and $\sigma_k(X_s)$ by stochastic Taylor approximations.

Itô-Taylor expansions

Suppose that X is a solution of (12.3.1), and let $f \in C^\infty(\mathbb{R}^N)$. Then the Itô-Doeblin formula for $f(X)$ on the interval $[t, t+h]$ can be written in the compact form

$$f(X_{t+h}) \;=\; f(X_t) \;+\; \sum_{k=0}^{d} \int_t^{t+h} (\mathcal{L}_k f)(X_s)\,dB_s^k \tag{12.3.3}$$

for any $t, h \geq 0$,where $B_t^0 = t$, $a = \sigma\sigma^T$,

$$\mathcal{L}_0 f \;=\; \frac{1}{2}\sum_{i,j=1}^{N} a^{ij}\frac{\partial^2 f}{\partial x^i \partial x^j} \;+\; b\cdot\nabla f, \qquad \text{and} \tag{12.3.4}$$

$$\mathcal{L}_k f \;=\; \sigma_k \cdot \nabla f, \qquad\qquad \text{for } k = 1, \ldots, d. \tag{12.3.5}$$

By iterating this formula, we obtain Itô-Taylor expansions for $f(X)$. For example, a first iteration yields

$$f(X_{t+h}) \;=\; f(X_t) + \sum_{k=0}^{d} (\mathcal{L}_k f)(X_t) \int_t^{t+h} dB_s^k + \sum_{k,l=0}^{d} \int_t^{t+h} \int_t^s (\mathcal{L}_l \mathcal{L}_k f)(X_r)\, dB_r^l\, dB_s^k.$$

The first two terms on the right hand side constitute a first order Taylor expansion for $f(X)$ in terms of the processes B^k, $k = 0, 1, \ldots, d$, and the iterated Itô integral in the third term is the corresponding remainder. Similarly, we obtain higher order expansions in terms of iterated Itô integrals where the remainders are given by higher order iterated integrals, cf. Theorem 12.14 below. The next lemma yields L^2 bounds on the remainder terms:

Lemma 12.13. *Suppose that $G : \Omega \times (t, t+h) \to \mathbb{R}$ is an adapted process in $\mathcal{L}^2(P \otimes \lambda_{(t,t+h)})$. Then*

$$E\left[\left(\int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{n-1}} G_{s_n}\, dB_{s_n}^{k_n} \cdots dB_{s_2}^{k_2}\, dB_{s_1}^{k_1}\right)^2\right] \;\le\; \frac{h^{n+m(k)}}{n!} \sup_{s \in (t,t+h)} E\left[G_s^2\right]$$

for any $n \in \mathbb{N}$ and $k = (k_1, \ldots, k_n) \in \{0, 1, \ldots, d\}^n$, where

$$m(k) \;:=\; |\{1 \le i \le n : k_i = 0\}|$$

denotes the number of integrations w.r.t. dt.

Proof. By Itô's isometry and the Cauchy-Schwarz inequality,

$$E\left[\left(\int_t^{t+h} G_s\, dB_s^k\right)^2\right] \;\le\; \int_t^{t+h} E\left[G_s^2\right]\, ds \qquad \text{for any } k \neq 0, \text{ and}$$

$$E\left[\left(\int_t^{t+h} G_s\, ds\right)^2\right] \;\le\; h \int_t^{t+h} E\left[G_s^2\right]\, ds.$$

By iteratively applying these estimates we see that the second moment of the iterated integral in the assertion is bounded from above by

$$h^{m(k)} \int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{n-1}} E[G_{s_n}^2]\, ds_n \cdots ds_2\, ds_1.$$

$\square$

The lemma can be applied to control the strong convergence order of stochastic Taylor expansions. For $k \in \mathbb{N}$ we denote by $C_b^k(\mathbb{R})$ the space of all C^k functions with bounded derivatives up to order k. Notice that we do not assume that the functions in C_b^k are bounded.

Definition (Stochastic convergence order). *Suppose that A_h, $h > 0$, and A are random variables, and let $\alpha > 0$.*

*1) A_h converges to A with **strong $\mathbf{L^2}$ order** α iff*

$$E\left[|A_h - A|^2\right]^{1/2} \;=\; O(h^\alpha).$$

*2) A_h converges to A with **weak order** α iff*

$$E\left[f(A_h)\right] - E\left[f(A)\right] \;=\; O(h^\alpha) \qquad \textit{for any } f \in C_b^{\lceil 2(\alpha+1)\rceil}(\mathbb{R}).$$

Notice that convergence with strong order α requires that the random variables are defined on a common probability space. For convergence with weak order α this is not necessary. If A_h converges to A with strong order α then we also write

$$A_h \;=\; A + \mathcal{O}(h^\alpha).$$

Examples. 1) If B is a *Brownian motion* then B_{t+h} converges to B_t almost surely as $h \downarrow 0$. By the law of the iterated logarithm, the pathwise convergence order is

$$B_{t+h} - B_t \;=\; O(h^{1/2} \log\log h^{-1}) \qquad \text{almost surely.}$$

On the other hand, the strong L^2 order is $1/2$, and the weak order is 1 since by Kolmogorov's forward equation,

$$E[f(B_{t+h})] - E[f(B_t)] \;=\; \int_t^{t+h} E[\frac{1}{2}\Delta f(B_s)]\,ds \;\leq\; \frac{h}{2}\sup \Delta f$$

for any $f \in C_b^2$. The exercise below shows that similar statements hold for more general Itô diffusions.

2) The n-fold iterated Itô integrals w.r.t. Brownian motion considered in Lemma 12.13 have strong order $(n+m)/2$ where m is the number of time integrals.

Exercise (Order of Convergence for Itô diffusions). Let $(X_t)_{t\geq 0}$ be an N-dimensional stochastic process satisfying the SDE (12.3.1) where $b, \sigma_k : \mathbb{R}^N \to \mathbb{R}^N$, $k = 1, \ldots, d$, are bounded continuous functions, and B is a d-dimensional Brownian motion. Prove that as $h \downarrow 0$,

1) X_{t+h} converges to X_t with strong L^2 order $1/2$.

2) X_{t+h} converges to X_t with weak order 1.

Theorem 12.14 (Itô-Taylor expansion with remainder of order α). *Suppose that $\alpha = j/2$ for some $j \in \mathbb{N}$. If X is a solution of (12.3.1) with coefficients $b, \sigma_1, \ldots, \sigma_d \in C_b^{\lfloor 2\alpha \rfloor}(\mathbb{R}^N, \mathbb{R}^N)$ then the following expansions hold for any $f \in C_b^{\lfloor 2\alpha+1 \rfloor}(\mathbb{R}^N)$:*

$$f(X_{t+h}) = \sum_{n<2\alpha} \sum_{k:n+m(k)<2\alpha} \left(\mathcal{L}_{k_n}\mathcal{L}_{k_{n-1}}\cdots\mathcal{L}_{k_1} f\right)(X_t) \times \tag{12.3.6}$$

$$\times \int_t^{t+h}\int_t^{s_1}\cdots\int_t^{s_{n-1}} G_{s_n}\, dB_{s_n}^{k_n}\cdots dB_{s_2}^{k_2}\, dB_{s_1}^{k_1} + \mathcal{O}(h^\alpha),$$

$$E\left[f(X_{t+h})\right] = \sum_{n<\alpha} E\left[\left(\mathcal{L}_0^n f\right)(X_t)\right] \frac{h^n}{n!} + O(h^\alpha). \tag{12.3.7}$$

Proof. Iteration of the Itô-Doeblin formula (12.3.3) shows that (12.3.6) holds with a remainder term that is a sum of iterated integrals of the form

$$\int_t^{t+h}\int_t^{s_1}\cdots\int_t^{s_{n-1}} \left(\mathcal{L}_{k_n}\mathcal{L}_{k_{n-1}}\cdots\mathcal{L}_{k_1} f\right)(X_{s_n})\, dB_{s_n}^{k_n}\cdots dB_{s_2}^{k_2}\, dB_{s_1}^{k_1}$$

with $k = (k_1, \ldots, k_n)$ satisfying $n + m(k) > 2\alpha$ and $n - 1 + m(k_1, \ldots, k_{n-1}) < 2\alpha$. By Lemma 12.13, these iterated integrals are of strong L^2 order $(n+m(k))/2$. Hence the full remainder term is of the order $\mathcal{O}(h^\alpha)$.

Equation (12.3.7) follows easily by iterating the Kolmogorov forward equation

$$E\left[f(X_{t+h})\right] = E\left[f(X_t)\right] + \int_t^{t+h} E\left[(\mathcal{L}_0 f)(X_s)\right] ds.$$

Alternatively, it can be derived from (12.3.6) by noting that all iterated integrals involving at least one integration w.r.t. a Brownian motion have mean zero. □

Remark (Computation of iterated Itô integrals). Iterated Itô integrals involving only a single one dimensional Brownian motion B can be computed explicitly from the Brownian increments. Indeed,

$$\int_t^{t+h}\int_t^{s_1}\cdots\int_t^{s_{n-1}} dB_{s_n}\cdots dB_{s_2}\,dB_{s_1} \;=\; h_n(h, B_{t+h}-B_t)/n!,$$

where h_n denotes the n-th Hermite polynomial, cf. (14.5.1). In the multi-dimensional case, however, the iterated Itô integrals can not be represented in closed form as functions of Brownian increments. Therefore, in higher order numerical schemes, these integrals have to be approximated separately. For example, the second iterated Itô integral

$$I_h^{kl} \;=\; \int_0^h\int_0^s dB_r^k\,dB_s^l \;=\; \int_0^h B_s^k\,dB_s^l$$

of two components of a d dimensional Brownian motion satisfies $I_h^{kl}+I_h^{lk} = B_h^k B_h^l$. Hence the symmetric part can be computed easily. However, the antisymmetric part $I_h^{kl}-I_h^{lk}$ is the *Lévy area process* of the two dimensional Brownian motion (B^k, B^l). The Lévy area can not be computed explicitly from the increments if $k \neq l$. Controlling the Lévy area is crucial for a pathwise stochastic integration theory, cf. [18, 19, 29].

Exercise (Lévy Area). If $c(t) = (x(t), y(t))$ is a smooth curve in $\mathbb{R}^2$ with $c(0) = 0$, then

$$A(t) \;=\; \int_0^t (x(s)y'(s)-y(s)x'(s))\,ds \;=\; \int_0^t x\,dy - \int_0^t y\,dx$$

describes the area that is covered by the secant from the origin to $c(s)$ in the interval $[0,t]$. Analogously, for a two-dimensional Brownian motion $B_t = (X_t, Y_t)$ with $B_0 = 0$, one defines the *Lévy Area*

$$A_t \;:=\; \int_0^t X_s\,dY_s \;-\; \int_0^t Y_s\,dX_s\,.$$

1) Let $\alpha(t)$, $\beta(t)$ be C^1-functions, $p\in\mathbb{R}$, and

$$V_t \;=\; ipA_t \;-\; \frac{\alpha(t)}{2}\left(X_t^2+Y_t^2\right) \;+\; \beta(t)\,.$$

Show using Itô's formula, that e^{V_t} is a local martingale provided $\alpha'(t) = \alpha(t)^2 - p^2$ and $\beta'(t) = \alpha(t)$.

2) Let $t_0 \in [0, \infty)$. The solutions of the ordinary differential equations for α and β with $\alpha(t_0) = \beta(t_0) = 0$ are

$$\begin{aligned}\alpha(t) &= p \cdot \tanh(p \cdot (t_0 - t)), \\ \beta(t) &= -\log\cosh(p \cdot (t_0 - t)).\end{aligned}$$

Conclude that

$$E\left[e^{ipA_{t_0}}\right] = \frac{1}{\cosh(pt_0)} \qquad \forall\, p \in \mathbb{R}.$$

3) Show that the distribution of A_t is absolutely continuous with density

$$f_{A_t}(x) = \frac{1}{2t\cosh(\frac{\pi x}{2t})}.$$

12.4 Numerical schemes for SDE

Let X be a solution of the SDE

$$dX_t = b(X_t)\,dt + \sum_{k=1}^{d} \sigma_k(X_t)\,dB_t^k \tag{12.4.1}$$

where we impose the same assumptions on the coefficients as in the last section. By applying the Itô-Doeblin formula to $\sigma_k(X_s)$ and taking into account all terms up to strong order $\mathcal{O}(h^1)$, we obtain the Itô-Taylor expansion

$$\begin{aligned}X_{t+h} - X_t &= b(X_t)\,h + \sum_{k=1}^{d} \sigma_k(X_t)\,(B_{t+h}^k - B_t^h) \qquad (12.4.2)\\ &\quad + \sum_{k,l=1}^{d} (\sigma_l \cdot \nabla\sigma_k)\,(X_t) \int_t^{t+h} \int_t^s dB_r^l\,dB_s^k + \mathcal{O}\left(h^{3/2}\right).\end{aligned}$$

Here the first term on the right hand side has strong L^2 order $\mathcal{O}(h)$, the second term $\mathcal{O}(h^{1/2})$, and the third term $\mathcal{O}(h)$. Taking into account only the first two terms leads to the Euler-Maruyama scheme with step size h, whereas taking into account all terms up to order $\mathcal{O}(h)$ yields the Milstein scheme:

- **Euler-Maruyama scheme with step size h**

$$X^h_{t+h} - X^h_t \;=\; b(X^h_t)\,h \;+\; \sum_{k=1}^{d} \sigma_k(X^h_t)\,(B^k_{t+h} - B^h_t) \qquad (t = 0, h, 2h, 3h, \ldots)$$

- **Milstein scheme with step size h**

$$X^h_{t+h} - X^h_t \;=\; b(X^h_t)\,h + \sum_{k=1}^{d} \sigma_k(X^h_t)\,(B^k_{t+h} - B^h_t) + \sum_{k,l=1}^{d} (\sigma_l \cdot \nabla\sigma_k)\,(X^h_t) \int_t^{t+h}\int_t^s dB^l_r\,dB^k_s$$

The Euler and Milstein scheme provide approximations to the solution of the SDE (12.4.1) that are defined for integer multiples t of the step size h. For a single approximation step, the strong order of accuracy is $\mathcal{O}(h)$ for Euler and $\mathcal{O}(h^{3/2})$ for Milstein. To analyse the total approximation error it is convenient to extend the definition of the approximation schemes to all $t \geq 0$ by considering the delay stochastic differential equations

$$dX^h_s \;=\; b(X^h_{\lfloor s\rfloor_h})\,ds \;+\; \sum_k \sigma_k(X^h_{\lfloor s\rfloor_h})\,dB^k_s, \tag{12.4.3}$$

$$dX^h_s \;=\; b(X^h_{\lfloor s\rfloor_h})\,ds \;+\; \sum_{k,l} \left(\sigma_k(X^h_{\lfloor s\rfloor_h}) + (\sigma_l\nabla\sigma_k)(X^h_{\lfloor s\rfloor_h}) \int_{\lfloor s\rfloor_h}^s dB^l_r\right) dB^k_s \tag{12.4.4}$$

respectively, where

$$\lfloor s\rfloor_h \;:=\; \max\{t \in h\mathbb{Z} : t \leq s\}$$

denotes the next discretization time below s. Notice that indeed, the Euler and Milstein scheme with step size h are obtained by evaluating the solutions of (12.4.3) and (12.4.4) respectively at $t = kh$ with $k \in \mathbb{Z}_+$.

Strong convergence order

Fix $a \in \mathbb{R}^N$, let X be a solution of (12.3.1) with initial condition $X_0 = a$, and let X^h be a corresponding Euler or Milstein approximation satisfying (12.4.3), (12.4.4) respectively with initial condition $X^h_0 = a$.

Theorem 12.15 (Strong order for Euler and Milstein scheme). *Let $t \in [0, \infty)$.*

1) *Suppose that the coefficients b and σ_k are bounded and Lipschitz continuous. Then the Euler-Maruyama approximation on the time interval $[0, t]$ has strong L^2 order $1/2$ in the following sense:*

$$\sup_{s \le t} \left| X_s^h - X_s \right| \;=\; \mathcal{O}(h^{1/2}).$$

2) *If, moreover, the coefficients b and σ_k are C^2 with bounded derivatives then the Milstein approximation on the time interval $[0, t]$ has strong L^2 order 1, i.e.,*

$$\left| X_t^h - X_t \right| \;=\; \mathcal{O}(h).$$

A corresponding uniform in time estimate for the Milstein approximation also holds but the proof is too long for these notes. The assumptions on the coefficients in the theorem are not optimal and can be weakened, see e.g. Milstein and Tretyakov [33]. However, it is well-known that even in the deterministic case a local Lipschitz condition is not sufficient to guarantee convergence of the Euler approximations. The iterated integral in the Milstein scheme can be approximated by a Fourier expansion in such a way that the strong order $\mathcal{O}(h)$ still holds, cf. Kloeden and Platen [26, 33]XXX

Proof. For notational simplicity, we only prove the theorem in the one-dimensional case. The proof in higher dimensions is analogous. The basic idea is to write down an SDE for the approximation error $X - X^h$.

1) By (12.4.3) and since $X_0^h = X_0$, the difference of the Euler approximation and the solution of the SDE satisfies the equation

$$X_t^h - X_t \;=\; \int_0^t \left(b(X^h_{\lfloor s \rfloor_h}) - b(X_s) \right) \, ds \;+\; \int_0^t \left(\sigma(X^h_{\lfloor s \rfloor_h}) - \sigma(X_s) \right) \, dB_s.$$

This enables us to estimate the mean square error

$$\bar{\varepsilon}_t^h \;:=\; E\left[\sup_{s \le t} \left| X_s^h - X_s \right|^2 \right].$$

By the Cauchy-Schwarz inequality and by Doob's L^2 inequality,

$$\begin{aligned}
\bar{\varepsilon}_t^h &\leq 2t\int_0^t E\left[\left|b(X^h_{\lfloor s\rfloor_h}) - b(X_s)\right|^2\right] ds + 8\int_0^t E\left[\left|\sigma(X^h_{\lfloor s\rfloor_h}) - \sigma(X_s)\right|^2\right] ds \\
&\leq (2t+8)\cdot L^2 \cdot \int_0^t E\left[\left|X^h_{\lfloor s\rfloor_h} - X_s\right|^2\right] ds \qquad (12.4.5)\\
&\leq (4t+16)\cdot L^2\cdot\left(\int_0^t \bar{\varepsilon}_s^h\, ds + C_t\, h\right),
\end{aligned}$$

where $t \mapsto C_t$ is an increasing real-valued function, and L is a joint Lipschitz constant for b and σ. Here, we have used that by the triangle inequality,

$$E\left[\left|X^h_{\lfloor s\rfloor_h} - X_s\right|^2\right] \leq 2\,E\left[\left|X^h_{\lfloor s\rfloor_h} - X^h_s\right|^2\right] + 2\,E\left[\left|X^h_s - X_s\right|^2\right],$$

and the first term representing the additional error by the time discretization on the interval $[\lfloor s\rfloor_h, \lfloor s\rfloor_h + h]$ is of order $O(h)$ uniformly on finite time intervals by a similar argument as in Theorem 12.14. By (12.4.5) and Gronwall's inequality, we conclude that

$$\bar{\varepsilon}_t^h \leq (4t+16)L^2 C_t \cdot \exp\left((4t+16)L^2 t\right)\cdot h,$$

and hence $\sqrt{\bar{\varepsilon}_t^h} = O(\sqrt{h})$ for any $t \in (0,\infty)$. This proves the assertion for the Euler scheme.

2) To prove the assertion for the Milstein scheme we have to argue more carefully. We will show that

$$\varepsilon_t^h := \sup_{s\leq t} E\left[\left|X^h_s - X_s\right|^2\right]$$

is of order $O(h^2)$. Notice that now the supremum is in front of the expectation, i.e., we are considering a weaker error than for the Euler scheme. We first derive an equation (and not just an estimate as above) for the mean square error. By (12.4.4), the difference of the Milstein approximation and the solution of the SDE satisfies

$$\begin{aligned}
X_t - X_t^h &= \int_0^t \left(b(X_s) - b(X^h_{\lfloor s\rfloor_h})\right) ds \qquad (12.4.6)\\
&\quad + \int_0^t \left(\sigma(X_s) - \sigma(X^h_{\lfloor s\rfloor_h}) - (\sigma\sigma')(X^h_{\lfloor s\rfloor_h})(B_s - B_{\lfloor s\rfloor_h})\right) dB_s.
\end{aligned}$$

By Itô's formula, we obtain

$$\begin{aligned} |X_t - X_t^h|^2 &= 2\int_0^t (X - X^h)\, d(X - X^h) \,+\, [X - X^h]_t \\ &= 2\int_0^t (X_s - X_s^h)\, \beta_s^h \, ds \,+\, 2\int_0^t (X_s - X_s^h)\, \alpha_s^h \, dB_s \,+\, \int_0^t |\alpha_s^h|^2 \, ds \end{aligned}$$

where $\beta_s^h = b(X_s) - b(X_{\lfloor s \rfloor_h}^h)$ and $\alpha_s^h = \sigma(X_s) - \sigma(X_{\lfloor s \rfloor_h}^h) - (\sigma\sigma')(X_{\lfloor s \rfloor_h}^h)(B_s - B_{\lfloor s \rfloor_h})$ are the integrands in (12.4.6). The assumptions on the coefficients guarantee that the stochastic integral is a martingale. Therefore, we obtain

$$E\left[|X_t - X_t^h|^2\right] \;=\; 2\int_0^t E\left[(X_s - X_s^h)\, \beta_s^h\right]\, ds \,+\, \int_0^t E\left[|\alpha_s^h|^2\right]\, ds. \tag{12.4.7}$$

We will now show that the integrands on the right side of (12.4.7) can be bounded by a constant times $\varepsilon_s^h + h^2$. The assertion then follows similarly as above by Gronwall's inequality.

In order to bound $E[|\alpha_s^h|^2]$ we decompose $\alpha_s^h = \alpha_{s,0}^h + \alpha_{s,1}^h$ where

$$\alpha_{s,1}^h \;=\; \sigma(X_s) - \sigma(X_{\lfloor s \rfloor_h}) - (\sigma\sigma')(X_{\lfloor s \rfloor_h})(B_s - B_{\lfloor s \rfloor_h})$$

is an additional error introduced in the current step, and

$$\alpha_{s,0}^h \;=\; \sigma(X_{\lfloor s \rfloor_h}) - \sigma(X_{\lfloor s \rfloor_h}^h) + \big((\sigma\sigma')(X_{\lfloor s \rfloor_h}) - (\sigma\sigma')(X_{\lfloor s \rfloor_h}^h)\big)\,(B_s - B_{\lfloor s \rfloor_h})$$

is an error carried over from previous steps. By the error estimate in the Itô-Taylor expansion, $\alpha_{s,1}^h$ is of strong order $\mathcal{O}(h)$ uniformly in s, i.e.,

$$E[|\alpha_{s,1}^h|^2] \;\le\; C_1 h^2 \qquad \text{for some finite constant } C_1.$$

Furthermore, since $B_s - B_{\lfloor s \rfloor_h}$ is independent of $\mathcal{F}_{\lfloor s \rfloor_h}^B$,

$$E[|\alpha_{s,0}^h|^2] \;\le\; 2(1+h)L^2\, E\left[|X_{\lfloor s \rfloor_h} - X_{\lfloor s \rfloor_h}^h|^2\right] \;\le\; 2(1+h)L^2\, \varepsilon_s^h,$$

and hence

$$E[|\alpha_s^h|^2] \;\le\; C_2\,(h^2 + \varepsilon_s^h) \qquad \text{for some finite constant } C_2. \tag{12.4.8}$$

It remains to prove an analogue bound for $E[(X_s - X_s^h)\,\beta_s^h]$. Similarly as above, we decompose $\beta_s^h = \beta_{s,0}^h + \beta_{s,1}^h$ where

$$\beta_{s,0}^h = b(X_{\lfloor s\rfloor_h}) - b(X_{\lfloor s\rfloor_h}^h) \qquad \text{and} \qquad \beta_{s,1}^h = b(X_s) - b(X_{\lfloor s\rfloor_h}).$$

By the Cauchy-Schwarz inequality and the Lipschitz continuity of b,

$$E[(X_s - X_s^h)\,\beta_{s,0}^h] \;\leq\; \left(\varepsilon_s^h\right)^{1/2} E\left[|\beta_{s,0}^h|^2\right]^{1/2} \;\leq\; L\,\varepsilon_s^h. \tag{12.4.9}$$

Moreover, there is a finite constant C_3 such that

$$\begin{aligned} E\left[(X_{\lfloor s\rfloor_h} - X_{\lfloor s\rfloor_h}^h)\,\beta_{s,1}^h\right] &= E\left[(X_{\lfloor s\rfloor_h} - X_{\lfloor s\rfloor_h}^h)\,E\left[b(X_s) - b(X_{\lfloor s\rfloor_h})|\mathcal{F}_s^B\right]\right] \\ &\leq C_3 h\left(\varepsilon_s^h\right)^{1/2} \;\leq\; C_3\,(h^2 + \varepsilon_s^h). \end{aligned} \tag{12.4.10}$$

Here we have used that by Kolmogorov's equation,

$$E\left[b(X_s) - b(X_{\lfloor s\rfloor_h})|\mathcal{F}_s^B\right] \;=\; \int_{\lfloor s\rfloor_h}^s E\left[(\mathcal{L}_0 b)(X_r)|\mathcal{F}_s^B\right]\,dr, \tag{12.4.11}$$

and $\mathcal{L}_0 b$ is bounded by the assumptions on b and σ.

Finally, let $Z_s^h := (X_s - X_s^h) - (X_{\lfloor s\rfloor_h} - X_{\lfloor s\rfloor_h}^h)$. By (12.4.6),

$$Z_s^h \;=\; \int_{\lfloor s\rfloor_h}^s \beta_r^h\,dr \;+\; \int_{\lfloor s\rfloor_h}^s \alpha_r^h\,dB_r, \qquad \text{and}$$

$$E\left[|Z_s^h|^2\right] \;\leq\; 2h\int_{\lfloor s\rfloor_h}^s E\left[|\beta_r^h|^2\right]\,dr \;+\; 2\int_{\lfloor s\rfloor_h}^s E\left[|\alpha_r^h|^2\right]\,dr \;\leq\; C_4\,h\,(h^2 + \varepsilon_s^h).$$

Here we have used the decomposition $\beta_s^h = \beta_{s,0}^h + \beta_{s,1}^h$ and (12.4.8). Hence

$$E[Z_s^h\beta_{s,1}^h] \;\leq\; \left\|Z_s^h\right\|_{L^2}\left\|b(X_s) - b(X_{\lfloor s\rfloor_h})\right\|_{L^2} \;\leq\; C_5\,h\,(h^2 + \varepsilon_s^h)^{1/2} \;\leq\; 2\,C_5\,(h^2 + \varepsilon_s^h).$$

By combining this estimate with (12.4.10) and (12.4), we eventually obtain

$$E[(X_s - X_s^h)\,\beta_s^h] \;\leq\; C_6\,(h^2 + \varepsilon_s^h) \qquad \text{for some finite constant } C_6. \tag{12.4.12}$$

□

Weak convergence order

We will now prove under appropriate assumptions on the coefficients that the Euler scheme has weak convergence order h^1. Let

$$\mathcal{L}f \;=\; \frac{1}{2}\sum_{i,j=1}^{N} a^{ij}\frac{\partial^2 f}{\partial x^i \partial x^j} \;+\; b\cdot\nabla f$$

denote the generator of the diffusion process (X_t). We assume that the coefficients $b, \sigma_1, \ldots, \sigma_d$ are in $C_b^3(\mathbb{R}^N, \mathbb{R}^N)$. It can be shown that under these conditions, for $f \in C_b^3(\mathbb{R}^N)$, the Kolmogorov backward equation

$$\frac{\partial u}{\partial t}(t,x) \;=\; (\mathcal{L}u)(t,x), \qquad u(0,x) \;=\; f(x), \tag{12.4.13}$$

has a unique classical solution $u : [0,\infty)\times\mathbb{R}^N \to \mathbb{R}$ such that $u(t,\cdot)\in C_b^3(\mathbb{R}^N)$ for any $t \geq 0$, cf. XXX. Moreover, if (X_t) is the unique strong solution of (12.3.1) with $X_0 = a$, then by Itô's formula,

$$E[f(X_t)] \;=\; u(t,a).$$

Theorem 12.16 (Weak order one for Euler scheme). *Suppose that $b, \sigma_1, \ldots, \sigma_d \in C_b^3(\mathbb{R}^N, \mathbb{R}^N)$, and let (X_t) and (X_t^h) denote the unique solution of* (12.3.1) *with $X_0 = a$ and its Euler approximation, respectively. Then*

$$E[f(X_t^h)] - E[f(X_t)] \;=\; O(h) \qquad \text{for any } t \geq 0 \text{ and } f \in C_b^3(\mathbb{R}^N).$$

Proof. Fix $t \geq 0$. The key idea (that is common with many other proofs) is to consider the "interpolation"

$$A_s \;:=\; u(t-s, X_s^h) \qquad \text{for } s\in[0,t].$$

Notice that $A_t = u(0, X_t^h) = f(X_t^h)$ and $A_0 = u(t,a) = E[f(X_t)]$, whence

$$E[f(X_t^h)] - E[f(X_t)] \;=\; E[A_t - A_0]. \tag{12.4.14}$$

We can now bound the weak error by applying Itô's formula. Indeed, by (12.4.3) and (12.4.13) we obtain

$$\begin{aligned} A_t - A_0 &= M_t + \int_0^t \left[-\frac{\partial u}{\partial t}(t-s, X_s^h) + (\mathcal{L}_s^h u)(t-s, X_{0:s}^h) \right] ds \\ &= M_t + \int_0^t \left[(\mathcal{L}_s^h u)(t-s, X_{0:s}^h) - (\mathcal{L}u)(t-s, X_s^h) \right] ds. \end{aligned}$$

Here M_t is a martingale, $Y_{0:t} := (Y_s)_{s\in[0,t]}$, and

$$(\mathcal{L}_t^h f)(x_{0:t}) = \frac{1}{2}\sum_{i,j=1}^N a^{ij}(x_{\lfloor t\rfloor_h})\frac{\partial^2 f}{\partial x^i \partial x^j}(x_t) + b(x_{\lfloor t\rfloor_h})\cdot \nabla f(x_t)$$

is the generator at time t of the delay equation (12.4.3) satisfied by the Euler scheme. Note that $\mathcal{L}_t^h(x_{0:t})$ is similar to $\mathcal{L}(x_t)$ but the coefficients are evaluated at $x_{\lfloor t\rfloor_h}$ instead of x_t. Taking expectations we conclude

$$E[A_t - A_0] = \int_0^t E\left[(\mathcal{L}_s^h u)(t-s, X_{0:s}^h) - (\mathcal{L}u)(t-s, X_s^h) \right] ds.$$

Thus the proof is complete if we can show that there is a finite constant C such that

$$\left|(\mathcal{L}_s^h u)(t-s, X_{0:s}^h) - (\mathcal{L}u)(t-s, X_s^h)\right| \le C\,h \quad \text{for } s\in[0,t] \text{ and } h\in(0,1]. \tag{12.4.15}$$

This is not difficult to verify by the assumptions on the coefficients. For instance, let us assume for simplicity that $d = 1$ and $b \equiv 0$, and let $a = \sigma^2$. Then

$$\begin{aligned} &\left|(\mathcal{L}_s^h u)(t-s, X_{0:s}^h) - (\mathcal{L}u)(t-s, X_s^h)\right| \\ &\le \frac{1}{2}\left|E\left[\left(a(X_s^h) - a(X_{\lfloor s\rfloor_h}^h)\right) u''(t-s, X_s^h)\right]\right| \\ &\le \frac{1}{2}\left|E\left[E\left[a(X_s^h) - a(X_{\lfloor s\rfloor_h}^h) | \mathcal{F}_{\lfloor s\rfloor_h}^B\right] u''(t-s, X_{\lfloor s\rfloor_h}^h)\right]\right| \\ &\quad + \frac{1}{2}\left|E\left[\left(a(X_s^h) - a(X_{\lfloor s\rfloor_h}^h)\right)\left(u''(t-s, X_s^h) - u''(t-s, X_{\lfloor s\rfloor_h}^h)\right)\right]\right|. \end{aligned}$$

Since u'' is bounded, the first summand on the right hand side is of order $O(h)$, cp. (12.4.11). By the Cauchy-Schwarz inequality, the second summand is also of order $O(h)$. Hence (12.4.15) is satisfied in this case. The proof in the general case is similar.

□

Remark (Generalizations).

1) The Euler scheme is given by

$$\Delta X_t^h \;=\; b(X_t^h)\,h \,+\, \sigma(X_t^h)\,\Delta B_t,\;\; \Delta B_t \text{ independent } \sim N(0,hI_d),\;\; t \in h\mathbb{Z}_+.$$

It can be shown that weak order one still holds if the ΔB_t are replaced by arbitrary i.i.d. random variables with mean zero, covariance hI_d, and third moments of order $O(h^2)$, cf. [26].

2) The Milstein scheme also has weak order h^1, so it does not improve on Euler w.r.t. weak convergence order. Higher weak order schemes are due to Milstein and Talay, see e.g. [33].

12.5 Local time

The occupation time of a Borel set $U \subseteq \mathbb{R}$ by a one-dimensional Brownian motion (B_t) is given by

$$L_t^U \;=\; \int_0^t I_U(B_s)\,ds.$$

Brownian local time is an *occupation time density* for Brownian motion that is informally given by

$$\text{“}\; L_t^a \;=\; \int_0^t \delta_a(B_s)\,ds \;\text{”}$$

for any $a \in \mathbb{R}$. It is a non-decreasing stochastic process satisfying

$$L_t^U \;=\; \int_U L_t^a\,da.$$

We will now apply stochastic integration theory for general predictable integrands to define the local time process $(L_t^a)_{t\geq 0}$ for $a \in \mathbb{R}$ rigorously for Brownian motion, and, more generally, for continuous semimartingales.

Local time of continuous semimartingales

Let (X_t) be a continuous semimartingale on a filtered probability space. Note that by Itô's formula,

$$f(X_t) - f(X_0) \;=\; \int_0^t f'(X_s)\,dX_s + \frac{1}{2}\int_0^t f''(X_s)\,d[X]_s.$$

Informally, if X is a Brownian motion then the last integral on the right hand side should coincide with L^a_t if $f'' = \delta_a$. A convex function with second derivative δ_a is $f(x) = (x-a)^+$. Noting that the left derivative of f is given by $f'_- = I_{(a,\infty)}$, this motivates the following definition:

Definition. *For a continuous semimartingale X and $a \in \mathbb{R}$, the process L^a defined by*

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{(a,\infty)}(X_s)\,dX_s + \frac{1}{2}L^a_t$$

is called the **local time of X at a**.

Remark. 1) By approximating the indicator function by continuous functions it can be easily verified that the process $I_{(a,\infty)}(X_s)$ is predictable and integrable w.r.t. X.

2) Alternatively, we could have defined local time at a by the identity

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{[a,\infty)}(X_s)\,dX_s + \frac{1}{2}\hat{L}^a_t$$

involving the right derivative $I_{[a,\infty)}$ instead of the left derivative $I_{(a,\infty)}$. Note that

$$L^a_t - \hat{L}^a_t \;=\; \int_0^t I_{\{a\}}(X_s)\,dX_s.$$

This difference vanishes almost surely if X is a Brownian motion, or, more generally, a continuous local martingale. For semimartingales, however, the processes L^a and $\hat{L}^a$ may disagree, cf. the example below Lemma 12.17. The choice of L^a in the definition of local time is then just a standard convention that is consistent with the convention of considering left derivatives of convex functions.

Lemma 12.17 (Properties of local time, Tanaka formulae).

1) *Suppose that* $\varphi_n : \mathbb{R} \to [0,\infty)$, $n \in \mathbb{N}$, *is a sequence of continuous functions with* $\int \varphi_n = 1$ *and* $\varphi_n(x) = 0$ *for* $x \notin (a, a+1/n)$. *Then*

$$L_t^a \;=\; \mathrm{ucp}-\lim_{n\to\infty} \int_0^t \varphi_n(X_s)\, d[X]_s.$$

In particular, the process $(L_t^a)_{t\geq 0}$ *is non-decreasing and continuous.*

2) *The process* L^a *grows only when* $X = a$, *i.e.,*

$$\int_0^t I_{\{X_s \neq a\}}\, dL_s^a \;=\; 0 \qquad \text{for any } t \geq 0.$$

3) *The following identities hold:*

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{(a,\infty)}(X_s)\, dX_s \,+\, \frac{1}{2} L_t^a, \tag{12.5.1}$$

$$(X_t - a)^- - (X_0 - a)^- \;=\; -\int_0^t I_{(-\infty,a]}(X_s)\, dX_s \,+\, \frac{1}{2} L_t^a, \tag{12.5.2}$$

$$|X_t - a| - |X_0 - a| \;=\; \int_0^t \operatorname{sgn}(X_s - a)\, dX_s \,+\, L_t^a, \tag{12.5.3}$$

where $\operatorname{sgn}(x) := +1$ *for* $x > 0$, *and* $\operatorname{sgn}(x) := -1$ *for* $x \leq 0$.

Remark. Note that we set $\operatorname{sgn}(0) := -1$. This is related to our convention of using left derivatives as $\operatorname{sgn}(x)$ is the left derivative of $|x|$. There are analogue Tanaka formulae for $\hat{L}^a$ with the intervals (a,∞) and $(-\infty,a]$ replaced by $[a,\infty)$ and $(-\infty,a)$, and the sign function defined by $\widehat{\operatorname{sgn}}(x) := +1$ for $x \geq 0$ and $\operatorname{sgn}(x) := -1$ for $x < 0$.

Proof. 1) For $n \in \mathbb{N}$ let $f_n(x) := \int_{-\infty}^x \int_{-\infty}^y \varphi_n(z)\, dz\, dy$. Then the function f_n is C^2 with $f_n'' = \varphi_n$. By Itô's formula,

$$f_n(X_t) - f_n(X_0) - \int_0^t f_n'(X_s)\, dX_s \;=\; \frac{1}{2}\int_0^t \varphi_n(X_s)\, d[X]_s. \tag{12.5.4}$$

As $n \to \infty$, $f_n'(X_s)$ converges pointwise to $I_{(a,\infty)}(X_s)$. Hence

$$\int_0^t f_n'(X_s)\, dX_s \;\to\; \int_0^t I_{(a,\infty)}(X_s)\, dX_s$$

in the ucp-sense by the Dominated Convergence Theorem 14.34. Moreover,

$$f_n(X_t) - f_n(X_0) \;\to\; (X_t - a)^+ - (X_0 - a)^+.$$

The first assertion now follows from (12.5.4).

2) By 1), the measures $\varphi_n(X_t)\,d[X]_t$ on $\mathbb{R}_+$ converge weakly to the measure dL_t^a with distribution function L^a. Hence by the Portemanteau Theorem, and since $\varphi_n(x) = 0$ for $x \notin (a, a+1/n)$,

$$\int_0^t I_{\{|X_s-a|>\varepsilon\}}\,dL_s^a \;\le\; \liminf_{n\to\infty} \int_0^t I_{\{|X_s-a|>\varepsilon\}}\,\varphi_n(X_s)\,d[X]_s \;=\; 0$$

for any $\varepsilon > 0$. The second assertion of the lemma now follows by the Monotone Convergence Theorem as $\varepsilon \downarrow 0$.

3) The first Tanaka formula (12.5.1) holds by definition of L^a. Moreover, subtracting (12.5.2) from (12.5.1) yields

$$(X_t - a) - (X_0 - a) \;=\; \int_0^t dX_s,$$

which is a valid equation. Therefore, the formulae (12.5.2) and (12.5.1) are equivalent. Finally, (12.5.3) follows by adding (12.5.1) and (12.5.2). □

Remark. In the proof above it is essential that the Dirac sequence (φ_n) approximates δ_a *from the right*. If X is a continuous martingale then the assertion 1) of the lemma also holds under the assumption that φ_n vanishes on the complement of the interval $(a-1/n, a+1/n)$. For semimartingales however, approximating δ_a from the left would lead to an approximation of the process $\hat{L}^a$, which in general may differ from L^a.

Exercise (Brownian local time). Show that the local time of a Brownian motion B in $a \in \mathbb{R}$ is given by

$$L_t^a \;=\; \text{ucp-}\lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^t I_{(a-\varepsilon,a+\varepsilon)}(B_s)\,ds.$$

Example (Reflected Brownian motion). Suppose that $X_t = |B_t|$ where (B_t) is a one-dimensional Brownian motion starting at 0. By Tanaka's formula (12.5.3), X is a semimartingale with decomposition

$$X_t \;=\; W_t \;+\; L_t \tag{12.5.5}$$

where L_t is the local time at 0 of the Brownian motion B and $W_t := \int_0^t \operatorname{sgn}(B_s)\,dB_s$. By Lévy's characterization, the martingale W is also a Brownian motion, cf. Theorem 11.2. We now compute the local time L_t^X of X at 0. By (12.5.2) and Lemma 12.17, 2),

$$\begin{aligned}\frac{1}{2}L_t^X &= X_t^- - X_0^- + \int_0^t I_{(-\infty,0]}(X_s)\,dX_s \qquad (12.5.6)\\ &= \int_0^t I_{\{0\}}(B_s)\,dW_s + \int_0^t I_{\{0\}}(B_s)\,dL_s \;=\; \int_0^t dL_s \;=\; L_t \qquad \text{a.s.,}\end{aligned}$$

i.e., $L_t^X = 2L_t$. Here we have used that $\int_0^t I_{\{0\}}(B_s)\,dW_s$ vanishes almost surely by Itô's isometry, as both W and B are Brownian motions. Notice that on the other hand,

$$\frac{1}{2}\hat{L}_t^X \;=\; X_t^- - X_0^- + \int_0^t I_{(-\infty,0)}(X_s)\,dX_s \;=\; 0 \qquad \text{a.s.,}$$

so the processes L^X and $\hat{L}^X$ do not coincide. By (12.5.5) and (12.5.6), the process X solves the singular SDE

$$dX_t \;=\; dW_t \,+\, \frac{1}{2}\,dL_t^X$$

driven by the Brownian motion W. This justifies thinking of X as *Brownian motion reflected at* 0.

The identity (12.5.5) can be used to compute the law of Brownian local time:

Exercise (The law of Brownian local time).

a) Prove **Skorohod's Lemma**: If $(y_t)_{t\geq 0}$ is a real-valued continuous function with $y_0 = 0$ then there exists a unique pair (x,k) of functions on $[0,\infty)$ such that

   (i) $x = y + k$,

   (ii) x is non-negative, and

   (iii) k is non-decreasing, continuous, vanishing at zero, and the measure dk_t is carried by the set $\{t : x_t = 0\}$.

   The function k is given by $k_t \;=\; \sup_{s\leq t}(-y_s)$.

b) Conclude that the local time process (L_t) at 0 of a one-dimensional Brownian motion (B_t) starting at 0 and the maximum process $S_t := \sup_{s\leq t} B_s$ have the same law. In particular, $L_t \sim |B_t|$ for any $t \geq 0$.

c) More generally, show that the two-dimensional processes $(|B|, L)$ and $(S-B, S)$ have the same law.

Notice that the maximum process $(S_t)_{t\geq 0}$ is the generalized inverse of the Lévy subordinator $(T_a)_{a\geq 0}$ introduced in Section 10.1. Thus we have identified Brownian local time at 0 as the inverse of a Lévy subordinator.

Itô-Tanaka formula

Local time can be used to extend Itô's formula in dimension one from C^2 to general convex functions. Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is **convex** iff

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

For a convex function f, the left derivatives

$$f'_-(x) \;=\; \lim_{h\downarrow 0} \frac{f(x) - f(x-h)}{h}$$

exist, the function f'_- is left-continuous and non-decreasing, and

$$f(b) - f(a) \;=\; \int_a^b f'_-(x)\, dx \qquad \text{for any } a, b \in \mathbb{R}.$$

The second derivative of f in the distributional sense is the positive measure f'' given by

$$f''([a,b)) \;=\; f'_-(b) - f'_-(a) \qquad \text{for any } a, b \in \mathbb{R}.$$

We will prove in Theorem 12.24 below that there is a version $(t, a) \mapsto L_t^a$ of the local time process of a continuous semimartingale X such that $t \mapsto L_t^a$ is continuous and $a \mapsto L_t^a$ is càdlàg. If X is a local martingale then L_t^a is even jointly continuous in t and a. From now on, we fix a corresponding version.

Theorem 12.18 (Itô-Tanaka formula, Meyer). *Suppose that X is a continuous semimartingale, and $f : \mathbb{R} \to \mathbb{R}$ is convex. Then*

$$f(X_t) - f(X_0) \;=\; \int_0^t f'_-(X_s)\, dX_s + \frac{1}{2} \int_{\mathbb{R}} L_t^a\, f''(da). \tag{12.5.7}$$

Proof. We proceed in several steps:

1) Equation (12.5.7) holds for linear functions f.

2) By localization, we may assume that $|X_t| < C$ for a finite constant C. Then both sides of (12.5.7) depend only on the values of f on $(-C, C)$, so we may also assume w.l.o.g. that f is linear on each of the intervals $(-\infty, -C]$ and $[C, \infty)$, i.e.,

$$\mathrm{supp}(f'') \subseteq [-C, C].$$

Moreover, by subtracting a linear function and multiplying f by a constant, we may even assume that f vanishes on $(-\infty, C]$, and f'' is a probability measure. Then

$$f'_-(y) = \mu(-\infty, y) \qquad \text{and} \qquad f(x) = \int_{-\infty}^{x} \mu(-\infty, y)\, dy \tag{12.5.8}$$

where $\mu := f''$.

3) Now suppose that $\mu = \delta_a$ is a Dirac measure. Then $f'_- = I_{(a,\infty)}$ and $f(x) = (x-a)^+$. Hence Equation (12.5.7) holds by definition of L^a. More generally, by linearity, (12.5.7) holds whenever μ has finite support, since then μ is a convex combination of Dirac measures.

4) Finally, if μ is a general probability measure then we approximate μ by measures with finite support. Suppose that Z is a random variable with distribution μ, and let μ_n denote the law of $Z_n := 2^{-n}\lceil 2^n Z \rceil$. By 3), the Itô-Tanaka formula holds for the functions $f_n(x) := \int_{-\infty}^{x} \mu_n(-\infty, y)\, dy$, i.e.,

$$f_n(X_t) - f_n(X_0) = \int_0^t f'_{n-}(X_s)\, dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a\, \mu_n(da) \tag{12.5.9}$$

for any $n \in \mathbb{N}$. As $n \to \infty$, $\mu_n(-\infty, X_s) \to \mu(-\infty, X_s)$, and hence

$$\int_0^t f'_{n-}(X_s)\, dX_s \to \int_0^t f'_-(X_s)\, dX_s$$

in the ucp sense by dominated convergence. Similarly, $f_n(X_t) - f_n(X_0) \to f(X_t) - f(X_0)$. Finally, the right continuity of $a \mapsto L_t^a$ implies that

$$\int_{\mathbb{R}} L_t^a\, \mu_n(da) \to \int_{\mathbb{R}} L_t^a\, \mu(da),$$

since Z_n converges to Z from above. The Itô-Tanaka formula (12.5.7) for f now follows from (12.5.9) as $n \to \infty$. $\square$

Clearly, the Itô-Tanaka formula also holds for functions f that are the difference of two convex functions. If f is C^2 then by comparing the Itô-Tanaka formula and Itô's formula, we can identify the integral $\int L_t^a\, f''(da)$ over a as the stochastic time integral $\int_0^t f''(X_s)\, d[X]_s$. The same remains true whenever the measure $f''(da)$ is absolutely continuous with density denoted by $f''(a)$:

Corollary 12.19. *For any measurable function $V : \mathbb{R} \to [0, \infty)$,*

$$\int_{\mathbb{R}} L_t^a\, V(a)\, da \;=\; \int_0^t V(X_s)\, d[X]_s \qquad \forall\, t \geq 0. \tag{12.5.10}$$

Proof. The assertion holds for any continuous function $V : \mathbb{R} \to [0, \infty)$ as V can be represented as the second derivative of a C^2 function f. The extension to measurable non-negative functions now follows by a monotone class argument. □

Notice that for $V = I_B$, the expression in (12.5.10) is the occupation time of the set B by (X_t), measured w.r.t. the quadratic variation $d[X]_t$.

12.6 Continuous modifications and stochastic flows

Let $\Omega = C_0(\mathbb{R}_+, \mathbb{R}^d)$ endowed with Wiener measure μ_0 and the canonical Brownian motion $W_t(\omega) = \omega(t)$. We consider the SDE

$$dX_t \;=\; b_t(X)\, dt + \sigma_t(X)\, dW_t, \qquad X_0 \;=\; a, \tag{12.6.1}$$

with progressively measurable coefficients $b, \sigma : \mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R}^n) \to \mathbb{R}^n, \mathbb{R}^{n\times d}$ respectively satisfying the global Lipschitz condition

$$|b_t(x) - b_t(\widetilde{x})| + ||\sigma_t(x) - \sigma_t(\widetilde{x})|| \;\leq\; L\, (x - \widetilde{x})_t^\star \qquad \forall\, t, x, \widetilde{x} \tag{12.6.2}$$

for some finite constant $L \in \mathbb{R}_+$, as well as

$$\sup_{s\in[0,t]} \big(|b_s(0)| + ||\sigma_s(0)||\big) \;<\; \infty \qquad \forall\, t. \tag{12.6.3}$$

Then by Itô's existence and uniqueness theorem, there exists a unique global strong solution $(X_t^a)_{t\geq 0}$ of (12.6.1) for any initial condition $a \in \mathbb{R}^n$. Our next goal is to show that there is a continuous modification $(t,a) \mapsto \xi_t^a$ of (X_t^a). The proof is based on the multidimensional version of the Kolmogorov-Čentsov continuity criterion for stochastic processes that is significant in many other contexts as well. Therefore, we start with a derivation of the Kolmogorov-Čentsov criterion from a corresponding regularity result for *deterministic* functions.

Continuous modifications of deterministic functions

Let $x : [0,1)^d \to E$ be a bounded measurable function from the d-dimensional unit cube to a separable Banach space $(E, \|\cdot\|)$. In the applications below, E will either be $\mathbb{R}^n$ or $C([0,t], \mathbb{R}^n)$ endowed with the supremum norm. The average of $x = (x_u)_{u\in[0,1)^d}$ over a smaller cube $Q \subseteq [0,1)^d$ is denoted by x_Q:

$$x_Q \;=\; \fint_Q x_u\,du \;=\; \frac{1}{\mathrm{vol}(Q)}\int_Q x_u\,du.$$

Let $\mathcal{D}_n$ be the collection of all dyadic cubes $Q = \prod_{i=1}^d [(k_i-1)2^{-n}, k_i 2^{-n})$ with $k_1,\dots,k_d \in \{1,2,\dots,2^n\}$. For $u \in [0,1)^d$ and $n \in \mathbb{N}$, we denote the unique cube in $\mathcal{D}_n$ containing u by $Q_n(u)$. Notice that $u \mapsto x_{Q_n(u)}$ is the conditional expectation of x given $\sigma(\mathcal{D}_n)$ w.r.t. the uniform distribution on the unit cube. By the martingale convergence theorem,

$$x_u \;=\; \lim_{n\to\infty} x_{Q_n(u)} \qquad \text{for almost every } u \in [0,1)^d,$$

where the limit is w.r.t. weak convergence if E is infinite dimensional.

Theorem 12.20 (Besov-Hölder embedding). *Let $\beta > 2d$ and $q \geq 1$, and suppose that*

$$B_{\beta,q} \;:=\; \left(\int_{[0,1)^d}\int_{[0,1)^d} \frac{\|x_u - x_v\|^q}{(|u-v|/\sqrt{d})^\beta}\,du\,dv\right)^{1/q} \tag{12.6.4}$$

is finite. Then the limit

$$\widetilde{x}_u \;:=\; \lim_{n\to\infty} x_{Q_n(u)}$$

exists for **every** $u \in [0,1)^d$*, and* $\widetilde{x}$ *is a Hölder continuous modification of* x *satisfying*

$$|\widetilde{x}_u - \widetilde{x}_v\| \;\le\; \frac{8}{\log 2}\frac{\beta}{\beta-2d}\, B_{\beta,q}\, |u-v|^{(\beta-2d)/q}. \tag{12.6.5}$$

For $s = \frac{\beta-d}{q} < 1$, the constant $B_{\beta,q}$ is essentially a *Besov norm* of order (s,q,q), or equivalently, a *Sobolev-Slobodecki norm* of order (s,q). The assertion of the theorem says that the corresponding Besov space is continuously embedded into the Hölder space of order $(\beta-2d)/q$, i.e., there is a finite constant C such that

$$\|\widetilde{x}\|_{\text{Höl}((\beta-2d)/q)} \;\le\; C\cdot \|x\|_{\text{Besov}((\beta-d)/q,q,q)}.$$

Proof. Let $e(Q)$ denote the edge length of a cube Q. The key step in the proof is to show that the inequality

$$\|x_Q - x_{\hat{Q}}\| \;\le\; \frac{4}{\log 2}\frac{\beta}{\beta-2d}\, B_{\beta,q}\, e(\hat{Q})^{(\beta-2d)/q} \tag{12.6.6}$$

holds for arbitrary cubes $Q, \hat{Q} \subseteq (0,1]^d$ such that $Q \subseteq \hat{Q}$. This inequality is proven by a **chaining argument**: Let

$$\hat{Q} = Q_0 \supset Q_1 \supset \cdots \supset Q_n = Q$$

be a decreasing sequence of a subcubes that interpolates between $\hat{Q}$ and Q . We assume that the edge lengths $e_k := e(Q_k)$ satisfy

$$e_{k+1}^{\beta/q} = \frac{1}{2}\, e_k^{\beta/q} \;\text{ for } k \ge 1, \;\text{ and }\; e_1^{\beta/q} \ge \frac{1}{2}\, e_0^{\beta/q}. \tag{12.6.7}$$

Since $\mathrm{vol}(Q_k) = e_k^d$ and $|u-v| \le \sqrt{d}\, e_{k-1}$ for any $u, v \in Q_{k-1}$, we obtain

$$\begin{aligned}
\|x_{Q_k} - x_{Q_{k-1}}\| &= \left\| ⨍_{Q_k} ⨍_{Q_{k-1}} (x_u - x_v)\, du\, dv \right\| \le \left(⨍_{Q_k} ⨍_{Q_{k-1}} \|x_u - x_v\|^q\, du\, dv \right)^{1/q} \\
&\le \left(\int_{Q_k}\int_{Q_{k-1}} \frac{\|x_u - x_v\|^q}{(|u-v|/\sqrt{d})^\beta}\, du\, dv \right)^{1/q} e_k^{-d/q}\, e_{k-1}^{-d/q}\, e_{k-1}^{\beta/q} \\
&\le 2\, B_{\beta,q}\, e_k^{(\beta-2d)/q} \;\le\; 4\, B_{\beta,q}\, e(Q)^{(\beta-2d)/q}\, 2^{-(\beta-2d)k/\beta}.
\end{aligned}$$

In the last two steps, we have used (12.6.7) and $e_{k-1} \geq e_k$. Noting that

$$\sum_{k=1}^{\infty} 2^{-ak} \;=\; 1/(2^a - 1) \;\leq\; 1/(a \log 2),$$

Equation (12.6.6) follows since $\|x_Q - x_{\hat{Q}}\| \leq \sum_{k=1}^{n} \|x_{Q_k} - x_{Q_{k-1}}\|$.

Next, consider arbitrary dyadic cubes $Q_n(u)$ and $Q_m(v)$ with $u, v \in [0,1)^d$ and $n, m \in \mathbb{N}$. Then there is a cube $\hat{Q} \subseteq [0,1)^d$ such that $\hat{Q} \supset Q_n(u) \cup Q_m(v)$ and

$$e(\hat{Q}) \;\leq\; |u - v| + 2^{-n} + 2^{-m}.$$

By (12.6.6) and the triangle inequality, we obtain

$$\begin{aligned} \|x_{Q_n(u)} - x_{Q_m(v)}\| \;&\leq\; \|x_{Q_n(u)} - x_{\hat{Q}}\| + \|x_{\hat{Q}} - x_{Q_m(v)}\| \\ &\leq\; \frac{8}{\log 2}\,\frac{\beta}{\beta - 2d}\, B_{\beta,q}\, \left(|u - v| + 2^{-n} + 2^{-m}\right)^{(\beta-2d)/q}. \end{aligned} \tag{12.6.8}$$

Choosing $v = u$ in (12.6.8), we see that the limit $\widetilde{x}_u = \lim_{n\to\infty} x_{Q_n(u)}$ exists. Moreover, for $v \neq u$, the estimate (12.6.5) follows as $n, m \to \infty$. □

Remark (Garsia-Rodemich-Rumsey). Theorem 12.20 is a special case of a result by Garsia, Rodemich and Rumsey where the powers in the definition of $B_{\beta,q}$ are replaced by more general increasing functions, cf. e.g. the appendix in [19]. This result allows to analyze the modulus of continuity more carefully, with important applications to Gaussian random fields [4].

Continuous modifications of random fields

The Kolmogorov-Čentsov continuity criterion for stochastic processes and random fields is a direct consequence of Theorem 12.20:

Theorem 12.21 (Kolmogorov, Čentsov). *Suppose that $(E, ||\cdot||)$ is a Banach space, $C = \prod_{k=1}^{d} I_k$ is a product of bounded real intervals $I_1, \ldots, I_d \subset \mathbb{R}$, and $X_u : \Omega \to E$,*

$u \in C$, is an E-valued stochastic process (a random field) indexed by C. If there exists constants $q, c, \varepsilon \in \mathbb{R}_+$ such that

$$E\left[||X_u - X_v||^q\right] \quad \leq \quad c|u-v|^{d+\varepsilon} \qquad \textit{for any } \; u, v \in C, \tag{12.6.9}$$

then there exists a modification $(\xi_u)_{u \in C}$ of $(X_u)_{u \in C}$ such that

$$E\left[\left(\sup_{u \neq v} \frac{||\xi_u - \xi_v||}{|u-v|^\alpha}\right)^q\right] \quad < \quad \infty \qquad \textit{for any } \; \alpha \in [0, \varepsilon/q). \tag{12.6.10}$$

In particular, $u \mapsto \xi_u$ is almost surely α-Hölder continuous for any $\alpha < \varepsilon/q$.

A direct proof based on a chaining argument can be found in many textbooks, see e.g. [37, Ch. I, (2.1)]. Here, we deduce the result as a corollary to the Besov-Hölder embedding theorem:

Proof. By rescaling we may assume w.l.o.g. that $C = [0,1)^d$. For $\beta > 0$, the assumption (12.6.9) implies

$$\begin{aligned} E\left[\int_C \int_C \frac{\|X_u - X_v\|^q}{|u-v|^\beta}\, du\, dv\right] \quad &\leq \quad c \int_C \int_C |u-v|^{d+\varepsilon-\beta}\, du\, dv \qquad (12.6.11) \\ &\leq \quad \text{const.} \int_0^{\sqrt{d}} r^{d+\varepsilon-\beta} r^{d-1}\, dr. \end{aligned}$$

Hence the expectation is finite for $\beta < 2d + \varepsilon$, and in this case,

$$\int_C \int_C \frac{\|X_u - X_v\|^q}{|u-v|^\beta}\, du\, dv \;<\; \infty \qquad \text{almost surely.}$$

Thus by Theorem 12.20, $\xi_u = \limsup_{n\to\infty} X_{Q_n(u)}$ defines a modification of (X_u) that is almost surely Hölder continuous with parameter $(\beta - 2d)/q$ for any $\beta < 2d + \varepsilon$. Moreover, the expectation of the q-th power of the Hölder norm is bounded by a multiple of the expectation in (12.6.11). $\square$

Example (Hölder continuity of Brownian motion). Brownian motion satisfies (12.6.9) with $d = 1$ and $\varepsilon = \frac{\gamma}{2} - 1$ for any $\gamma \in (2, \infty)$. Letting γ tend to ∞, we see that almost every Brownian path is α-Hölder continuous for any $\alpha < 1/2$. This result is sharp in the sense that almost every Brownian path is not $\frac{1}{2}$-Hölder-continuous, cf. [14, Thm. 1.20].

In a similar way, one can study the continuity properties of general Gaussian random fields, cf. Adler and Taylor [4]. Another very important application of the Besov-Hölder embedding and the resulting bounds for the modulus of continuity are tightness results for families of stochastic processes or random random fields, see e.g. Stroock and Varadhan [40]. Here, we consider two different applications that concern the continuity of stochastic flows and of local times.

Existence of a continuous flow

We now apply the Kolmogorov-Čentsov continuity criterion to the solution $a \mapsto (X_s^a)$ of the SDE (12.6.1) as a function of its starting point.

Theorem 12.22 (Flow of an SDE). *Suppose that (12.6.2) and (12.6.3) hold.*

1) *There exists a function $\xi : \mathbb{R}^n \times \Omega \to C(\mathbb{R}_+, \mathbb{R}^n)$, $(a, \omega) \mapsto \xi^a(\omega)$ such that*
   - (i) *$\xi^a = (\xi_t^a)_{t\geq 0}$ is a strong solution of (12.6.1) for any $a \in \mathbb{R}^n$, and*
   - (ii) *the map $a \mapsto \xi^a(\omega)$ is continuous w.r.t. uniform convergence on finite time intervals for any $\omega \in \Omega$.*

2) *If $\sigma(t, x) = \widetilde{\sigma}(x_t)$ and $b(t, x) = \widetilde{b}(x_t)$ with Lipschitz continuous functions $\widetilde{\sigma} : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ and $\widetilde{b} : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ then ξ satisfies the* ***cocycle property***

$$\xi_{t+s}^a(\omega) \quad = \quad \xi_s^{\xi_t^a(\omega)}(\Theta_t(\omega)) \qquad \forall\, s, t \geq 0, \;\; a \in \mathbb{R}^n \tag{12.6.12}$$

for μ_0-almost every ω, where

$$\Theta_t(\omega) \quad = \quad \omega(\cdot + t) \quad \in \quad C(\mathbb{R}_+, \mathbb{R}^d)$$

denotes the shifted path, and the definition of ξ has been extended by

$$\xi(\omega) \quad := \quad \xi(\omega - \omega(0)) \tag{12.6.13}$$

to paths $\omega \in C(\mathbb{R}_+, \mathbb{R}^d)$ with starting point $\omega(0) \neq 0$.

Proof. 1) We fix $p > d$. By the a priori estimate in Theorem 12.1 there exists a finite constant $c \in \mathbb{R}_+$ such that

$$E[(X^a - X^{\widetilde{a}})_t^{\star\, p}] \quad \leq \quad c \cdot e^{ct}\, |a - \widetilde{a}|^p \qquad \text{for any } t \geq 0 \text{ and } a, \widetilde{a} \in \mathbb{R}^n, \quad (12.6.14)$$

where X^a denotes a version of the strong solution of (12.6.1) with initial condition a. Now fix $t \in \mathbb{R}_+$. We apply the Kolmogorov-Čentsov Theorem with $E = C([0,t], \mathbb{R}^n)$ endowed with the supremum norm $||X||_t = X_t^\star$. By (12.6.14), there exists a modification ξ of $(X_s^a)_{s \leq t, a \in \mathbb{R}^n}$ such that $a \mapsto (\xi_s^a)_{s \leq t}$ is almost surely α-Hölder continuous w.r.t. $\| \cdot \|_t$ for any $\alpha < \frac{p-n}{p}$. Clearly, for $t_1 \leq t_2$, the almost surely continuous map $(s,a) \mapsto \xi_s^a$ constructed on $[0,t_1] \times \mathbb{R}^n$ coincides almost surely with the restriction of the corresponding map on $[0,t_2] \times \mathbb{R}^n$. Hence we can almost surely extend the definition to $\mathbb{R}_+ \times \mathbb{R}^n$ in a consistent way.

2) Fix $t \geq 0$ and $a \in \mathbb{R}^n$. Then μ_0-almost surely, both sides of (12.6.12) solve the same SDE as a function of s. Indeed,

$$\begin{aligned} \xi_{t+s}^a &= \xi_t^a + \int_t^{t+s} \widetilde{b}(\xi_u^a)\, du + \int_t^{t+s} \widetilde{\sigma}(\xi_u^a)\, dW_u \\ &= \xi_t^a + \int_0^s \widetilde{b}(\xi_{t+r}^a)\, dr + \int_0^s \widetilde{\sigma}(\xi_{t+r}^a)\, d\left(W_r \circ \Theta_t\right), \end{aligned}$$

$$\xi_s^{\xi_t^a} \circ \Theta_t \quad = \quad \xi_t^a + \int_0^s \widetilde{b}\big(\xi_r^{\xi_t^a} \circ \Theta_t\big)\, dr + \int_0^s \widetilde{\sigma}(\xi_r^{\xi_t^a} \circ \Theta_t)\, d(W_r \circ \Theta_t)$$

hold μ_0-almost surely for any $s \geq 0$ where $r \mapsto W_r \circ \Theta_t = W_{r+t}$ is again a Brownian motion, and $\big(\xi_r^{\xi_t^a} \circ \Theta_t\big)(\omega) := \xi_r^{\xi_t^a(\omega)}(\Theta_t(\omega))$. Pathwise uniqueness now implies

$$\xi_{t+s}^a \quad = \quad \xi_s^{\xi_t^a} \circ \Theta_t \qquad \text{for any } s \geq 0, \qquad \text{almost surely.}$$

Continuity of ξ then shows that the cocycle property (12.6.12) holds with probability one for all s, t and a simultaneously. □

Remark (Extensions). 1) *Joint Hölder continuity in t and a:* Since the constant p in the proof above can be chosen arbitrarily large, the argument yields α-Hölder continuity of $a \mapsto \xi^a$ for any $\alpha < 1$. By applying Kolmogorov's criterion in dimension $n+1$, it is also possible to prove joint Hölder continuity in t and a. In Section 13.1 we will prove that under a stronger assumption on the coefficients b and σ, the flow is even continuously differentiable in a.

2) *SDE with jumps:* The first part of Theorem 12.22 extends to solutions of SDE of type (12.1.1) driven by a Brownian motion and a Poisson point process. In that case, under a global Lipschitz condition the same arguments go through if we replace $C([0,t],\mathbb{R}^n)$ by the Banach space $\mathcal{D}([0,t],\mathbb{R}^n)$ when applying Kolmogorov's criterion. Hence in spite of the jumps, the solution depends continuously on the initial value a !

3) *Locally Lipschitz coefficients:* By localization, the existence of a continuous flow can also be shown under local Lipschitz conditions, cf. e.g. [36]. Notice that in this case, the explosion time depends on the initial value.

Above we have shown the existence of a continuous flow for the SDE (12.6.1) on the canonical setup. From this we can obtain strong solutions on other setups:

Exercise. Show that the unique strong solution of (12.6.1) w.r.t. an arbitrary driving Brownian motion B instead of W is given by $X_t^a(\omega) = \xi_t^a(B(\omega))$.

Markov property

In the time-homogeneous diffusion case, the Markov property for solutions of the SDE (12.6.1) is a direct consequence of the cocycle property:

Corollary 12.23. *Suppose that $\sigma(t,x) = \widetilde{\sigma}(x_t)$ and $b(t,x) = \widetilde{b}(x_t)$ with Lipschitz continuous functions $\widetilde{\sigma} : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ and $\widetilde{b} : \mathbb{R}^n \to \mathbb{R}^n$. Then $(\xi_t^a)_{t\geq 0}$ is a time-homogeneous $(\mathcal{F}_t^{W,P})$ Markov process with transition function*

$$p_t(a,B) \quad = \quad P[\xi_t^a \in B], \qquad t \geq 0, \;\; a \in \mathbb{R}^n.$$

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. Then for $0 \leq s \leq t$,

$$\Theta_t(\omega) \quad = \quad \omega(t) + \big(\omega(t+\cdot) - \omega(t)\big),$$

and hence, by the cocycle property and by (12.6.13),

$$f(\xi^a_{s+t}(\omega)) \quad = \quad f\big(\xi_s^{\xi_t^a(\omega)}\big(\omega(t+\cdot)-\omega(t)\big)\big)$$

for a.e. ω. Since $\omega(t+\cdot)-\omega(t)$ is a Brownian motion starting at 0 independent of $\mathcal{F}_t^{W,P}$, we obtain

$$E\big[f(\xi^a_{s+t})|\mathcal{F}_t^{W,P}\big](\omega) \quad = \quad E\big[f(\xi_s^{\xi_t^a(\omega)})\big] \quad = \quad (p_s f)(\xi_t^a(\omega)) \qquad \text{almost surely.}$$

□

Remark. Without pathwise uniqueness, both the cocycle and the Markov property do not hold in general.

Continuity of local time

The Kolmogorov-Čentsov continuity criterion can also be applied to prove the existence of a jointly continuous version $(a,t) \mapsto L_t^a$ of the local time of a continuous local martingale. More generally, recall that the local time of a continuous semimartingale $X = M + A$ is defined by the Tanaka formula

$$\frac{1}{2}L_t^a \;=\; (X_0-a)^+ - (X_t-a)^+ - \int_0^t I_{(a,\infty)}(X_s)\,dM_s - \int_0^t I_{(a,\infty)}(X_s)\,dA_s \quad (12.6.15)$$

almost surely for any $a \in \mathbb{R}$.

Theorem 12.24 (Yor). *There exists a version $(a,t) \mapsto L_t^a$ of the local time process that is continuous in t and càdlàg in a with*

$$L_t^a - L_t^{a-} \;=\; 2\int_0^t I_{\{X_s=a\}}\,dA_s. \qquad (12.6.16)$$

In particular, $(a,t) \mapsto L_t^a$ is jointly continuous if M is a continuous local martingale.

Proof. By localization, we may assume that M is a bounded martingale and A has bounded total variation $V_\infty^{(1)}(A)$. The map $(a,t) \mapsto (X_t - a)^+$ is jointly continuous in t and a. Moreover, by dominated convergence,

$$Z_t^a \ := \ \int_0^t I_{(a,\infty)}(X_s)\, dA_s$$

is continuous in t and càdlàg in a with

$$Z_t^a - Z_t^{a-} \ = \ -\int_0^t I_{\{a\}}(X_s)\, dA_s.$$

Therefore it is sufficient to prove that

$$Y_t^a \ := \ \int_0^t I_{(a,\infty)}(X_s)\, dM_s$$

has a version such that the map $a \mapsto (Y_s^a)_{s \le t}$ from $\mathbb{R}$ to $C([0,t],\mathbb{R}^n)$ is continuous for any $t \in [0,\infty)$.

Hence fix $t \ge 0$ and $p \ge 4$. By Burkholder's inequality,

$$\begin{aligned} E\left[\left(Y^a - Y^b\right)_t^{\star\, p}\right] &= E\left[\sup_{s<t}\left|\int_0^s I_{(a,b]}(X)\, dM\right|^p\right] \qquad (12.6.17)\\ &\le C_1(p)\, E\left[\left|\int_0^t I_{(a,b]}(X)\, d[M]\right|^{p/2}\right] \end{aligned}$$

holds for any $a < b$ with a finite constant $C_1(p)$. The integral appearing on the right hand side is an occupation time of the interval $(a,b]$. To bound this integral, we apply Itô's formula with a function $f \in C^1$ such that $f'(x) = (x \wedge b - a)^+$ and hence $f''_- = I_{(a,b]}$. Although f is not C^2, an approximation of f by smooth functions shows that Itô's formula holds for f, i.e.,

$$\begin{aligned} \int_0^t I_{(a,b]}(X)\, d[M] &= \int_0^t I_{(a,b]}(X)\, d[X] \\ &= -2\left(f(X_t) - f(X_0) - \int_0^t f'(X)\, dX\right) \\ &\le (b-a)^2 \, + \, 2\left|\int_0^t f'(X)\, dM\right| \, + \, |b-a|\, V_t^{(1)}(A) \end{aligned}$$

Here we have used in the last step that $|f'| \leq |b-a|$ and $|f| \leq (b-a)^2/2$. Combining this estimate with 12.6.17 and applying Burkholder's inequality another time, we obtain

$$\begin{aligned} E\left[\left(Y^a - Y^b\right)_t^{\star\, p}\right] &\leq C_2(p,t)\left(|b-a|^{p/2} + E\left[\left(\int_0^t f'(X)^2\, d[M]\right)^{p/4}\right]\right) \\ &\leq C_2(p,t)\,|b-a|^{p/2}\,(1 + [M]_t^{p/4}) \end{aligned}$$

with a finite constant $C_2(p,t)$. The existence of a continuous modification of $a \mapsto (Y_s^a)_{s\leq t}$ now follows from the Kolmogorov-Čentsov Theorem. □

Remark. 1) The proof shows that for a continuous local martingale, $a \mapsto (L_s^a)_{s\leq t}$ is α-Hölder continuous for any $\alpha < 1/2$ and $t \in \mathbb{R}_+$.

2) For a continuous semimartingale, $L_t^{a-} = \hat{L}_t^a$ by (12.6.16).

Chapter 13

Variations of parameters in SDE

In this chapter, we consider variations of parameters in stochastic differential equations. This leads to a first introduction to basic concepts and results of Malliavin calculus. For a more thorough introduction to Malliavin calculus we refer to [35], [34], [41], [23], [32] and [9].

Let μ denote Wiener measure on the Borel σ-algebra $\mathcal{B}(\Omega)$ over the Banach space $\Omega = C_0([0,1],\mathbb{R}^d)$ endowed with the supremum norm $||\omega|| = \sup\{|\omega(t)| : t \in [0,1]\}$, and consider an SDE of type

$$dX_t \quad = \quad b(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad X_0 \;=\; x, \tag{13.0.1}$$

driven by the canonical Brownian motion $W_t(\omega) = \omega(t)$. In this chapter, we will be interested in dependence of strong solutions on the initial condition and other parameters. The existence and uniqueness of strong solutions and of continuous stochastic flows has already been studied in Sections 12.1 and 12.6. We are now going to prove differentiability of the solution w.r.t. variations of the initial condition and the coefficients, see Section 13.1. A main goal will be to establish relations between different types of variations of (13.0.1):

- Variations of the initial condition: $x \to x(\varepsilon)$
- Variations of the coefficients: $b(x) \to b(\varepsilon, x), \quad \sigma(x) \to \sigma(\varepsilon, x)$
- Variations of the driving paths: $W_t \to W_t + \varepsilon H_t, \quad (H_t)$ adapted

- Variations of the underlying probability measure: $\mu \to \mu^\varepsilon = Z^\varepsilon \cdot \mu$

Section 13.2 introduces the Malliavin gradient which is a derivative of a function on Wiener space (e.g. the solution of an SDE) w.r.t. variations of the Brownian path. Bismut's integration by parts formula is an infinitesimal version of the Girsanov Theorem, which relates these variations to variations of Wiener measure. After a digression to representation theorems in Section 9.4, Section 13.3 discusses Malliavin derivatives of solutions of SDE and their connection to variations of the initial condition and the coefficients. As a consequence, we obtain first stability results for SDE from the Bismut integration by parts formula. Finally, Section 13.4 sketches briefly how Malliavin calculus can be applied to prove existence and smoothness of densities of solutions of SDE. This should give a first impression of a powerful technique that eventually leads to impressive results such as Malliavin's stochastic proof of Hörmander's theorem, cf. [21], [34].

13.1 Variations of parameters in SDE

We now consider a stochastic differential equation

$$dX_t^\varepsilon \quad = \quad b(\varepsilon, X_t^\varepsilon)\, dt + \sum_{k=1}^d \sigma_k(\varepsilon, X_t^\varepsilon)\, dW_t^k, \qquad X_0^\varepsilon \;=\; x(\varepsilon), \tag{13.1.1}$$

on $\mathbb{R}^n$ with coefficients and initial condition depending on a parameter $\varepsilon \in U$, where U is a convex neighbourhood of 0 in $\mathbb{R}^m$, $m \in \mathbb{N}$. Here $b, \sigma_k : U \times \mathbb{R}^n \to \mathbb{R}^n$ are functions that are Lipschitz continuous in the second variable, and $x : U \to \mathbb{R}^n$. We already know that for any $\varepsilon \in U$, there exists a unique strong solution $(X_t^\varepsilon)_{t \geq 0}$ of (13.1.1). For $p \in [1, \infty)$ let

$$||X^\varepsilon||_p \quad := \quad E\Big[\sup_{t \in [0,1]} |X_t^\varepsilon|^p\Big]^{1/p}.$$

Exercise (Lipschitz dependence on ε). Prove that if the maps x, b and σ_k are all Lipschitz continuous, then $\varepsilon \mapsto X^\varepsilon$ is also Lipschitz continuous w.r.t. $||\cdot||_p$, i.e., there exists a constant $L_p \in \mathbb{R}_+$ such that

$$||X^{\varepsilon+h} - X^\varepsilon||_p \quad \leq \quad L_p\, |h|, \qquad \text{for any } \varepsilon, h \in \mathbb{R}^m \text{ with } \varepsilon, \varepsilon + h \in U.$$

We now prove a stronger result under additional regularity assumptions.

Differentation of solutions w.r.t. a parameter

Theorem 13.1. *Let $p \in [2, \infty)$, and suppose that x, b and σ_k are C^2 with bounded derivatives up to order 2. Then the function $\varepsilon \mapsto X^\varepsilon$ is differentiable on U w.r.t. $||\cdot||_p$, and the differential $Y^\varepsilon = \frac{dX^\varepsilon}{d\varepsilon}$ is the unique strong solution of the SDE*

$$\begin{aligned} dY_t^\varepsilon &= \left(\frac{\partial b}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial b}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon\right) dt \qquad (13.1.2) \\ &\quad + \sum_{k=1}^d \left(\frac{\partial \sigma_k}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial \sigma_k}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon\right) dW_t^k, \\ Y_0^\varepsilon &= x'(\varepsilon), \qquad (13.1.3) \end{aligned}$$

that is obtained by formally differentiating (13.1.1) w.r.t. ε.

Here and below $\frac{\partial}{\partial \varepsilon}$ and $\frac{\partial}{\partial x}$ denote the differential w.r.t. the ε and x variable, and x' denotes the (total) differential of the function x.

Remark. Note that if (X_t^ε) is given, then (13.1.2) is a linear SDE for (Y_t^ε) (with multiplicative noise). In particular, there is a unique strong solution. The SDE for the derivative process Y^ε is particularly simple if σ is constant: In that case, (13.1.2) is a deterministic ODE with coefficients depending on X^ε.

Proof of 13.1. We prove the stronger statement that there is a constant $M_p \in (0, \infty)$ such that

$$\left|\left|X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h\right|\right|_p \quad \leq \quad M_p\, |h|^2 \qquad (13.1.4)$$

holds for any $\varepsilon, h \in \mathbb{R}^m$ with $\varepsilon, \varepsilon + h \in U$, where Y^ε is the unique strong solution of (13.1.2). Indeed, by subtracting the equations satisfied by $X^{\varepsilon+h}$, X^ε and $Y^\varepsilon h$, we obtain for $t \in [0, 1]$:

$$\left|X_t^{\varepsilon+h} - X_t^\varepsilon - Y_t^\varepsilon h\right| \quad \leq \quad |\mathrm{I}| + \int_0^t |\mathrm{II}|\, ds + \sum_{k=1}^d \left|\int_0^t \mathrm{III}_k\, dW^k\right|,$$

where

$$
\begin{aligned}
\mathrm{I} &= x(\varepsilon+h)-x(\varepsilon)-x'(\varepsilon)h,\\
\mathrm{II} &= b(\varepsilon+h,X^{\varepsilon+h})-b(\varepsilon,X^{\varepsilon})-b'(\varepsilon,X^{\varepsilon})\begin{pmatrix}h\\ Y^{\varepsilon}h\end{pmatrix},\quad\text{and}\\
\mathrm{III}_k &= \sigma_k(\varepsilon+h,X^{\varepsilon+h})-\sigma_k(\varepsilon,X^{\varepsilon})-\sigma_k'(\varepsilon,X^{\varepsilon})\begin{pmatrix}h\\ Y^{\varepsilon}h\end{pmatrix}.
\end{aligned}
$$

Hence by Burkholder's inequality, there exists a finite constant C_p such that

$$
E\big[(X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h)_t^{\star\,p}\big] \quad\leq\quad C_p\cdot\left(|\mathrm{I}|^p+\int_0^t E\big[|\mathrm{II}|^p+\sum_{k=1}^d|\mathrm{III}_k|^p\big]\,ds\right). \tag{13.1.5}
$$

Since x,b and σ_k are C^2 with bounded derivatives, there exist finite constants C_{I}, C_{II}, C_{III} such that

$$
|\mathrm{I}| \quad\leq\quad C_{\mathrm{I}}|h|^2, \tag{13.1.6}
$$

$$
|\mathrm{II}| \quad\leq\quad C_{\mathrm{II}}|h|^2+\big|\frac{\partial b}{\partial x}(\varepsilon,X^{\varepsilon})(X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h)\big|, \tag{13.1.7}
$$

$$
|\mathrm{III}_k| \quad\leq\quad C_{\mathrm{III}}|h|^2+\big|\frac{\partial \sigma_k}{\partial x}(\varepsilon,X^{\varepsilon})(X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h)\big|. \tag{13.1.8}
$$

Hence there exist finite constants $\widetilde{C}_p$, $\hat{C}_p$ such that

$$
E[|\mathrm{II}|^p+\sum_{k=1}^d|\mathrm{III}_k|^p] \quad\leq\quad \widetilde{C}_p\left(|h|^{2p}+E\left[\left|X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h\right|^p\right]\right),
$$

and thus, by (13.1.5) and (13.1.6),

$$
E\big[(X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h)_t^{\star p}\big] \quad\leq\quad \hat{C}_p|h|^{2p}+C_p\widetilde{C}_p\int_0^t E\big[(X^{\varepsilon+h}-X^{\varepsilon}-Y^{\varepsilon}h)_s^{\star p}\big]\,ds
$$

for any $t\leq 1$. The assertion (13.1.4) now follows by Gronwall's lemma. □

Derivative flow and stability of stochastic differential equations

We now apply the general result above to variations of the initial condition, i.e., we consider the flow

$$
d\xi_t^x \quad=\quad b(\xi_t^x)\,dt+\sum_{k=1}^d\sigma_k(\xi_t^x)\,dW_t^k,\qquad \xi_0^x\;=\;x. \tag{13.1.9}
$$

Assuming that b and σ_k ($k = 1, \dots, d$) are C^2 with bounded derivatives, Theorem 13.1 shows that the **derivative flow**

$$Y_t^x \quad := \quad \xi_t'(x) \quad = \quad \left(\frac{\partial}{\partial x^k} \xi_t^{x,l}\right)_{1 \le k,l \le n}$$

exists w.r.t. $||\cdot||_p$ and $(Y_t^x)_{t \ge 0}$ satisfies the SDE

$$dY_t^x \quad = \quad b'(\xi_t^x)\, Y_t^x\, dt + \sum_{k=1}^d \sigma_k'(\xi_t^x)\, Y_t^x\, dW_t^k, \qquad Y_0^x \;=\; I_n. \tag{13.1.10}$$

Note that again, this is a linear SDE for Y if ξ is given, and Y is the fundamental solution of this SDE.

Remark (Flow of diffeomorphisms). One can prove that $x \mapsto \xi_t^x(\omega)$ is a diffeomorphism on $\mathbb{R}^n$ for any t and ω, cf. [27] or [15].

In the sequel, we will denote the directional derivative of the flow ξ_t in direction $v \in \mathbb{R}^n$ by $Y_{v,t}$:

$$Y_{v,t} \quad = \quad Y_{v,t}^x \quad = \quad Y_t^x v \quad = \quad \partial_v \xi_t^x.$$

(i) Constant diffusion coefficients. Let us now first assume that $d = n$ and $\sigma(x) = I_n$ for any $x \in \mathbb{R}^n$. Then the SDE reads

$$d\xi^x \quad = \quad b(\xi^x)\, dt + dW, \qquad \xi_0^x \;=\; x;$$

and the derivative flow solves the ODE

$$dY^x \quad = \quad b'(\xi^x) Y\, dt, \qquad Y_0 \;=\; I_n.$$

This can be used to study the stability of solutions w.r.t. variations of initial conditions pathwise:

Theorem 13.2 (Exponential stability I). *Suppose that $b : \mathbb{R}^n \to \mathbb{R}^n$ is C^2 with bounded derivatives, and let*

$$\kappa \quad = \quad \sup_{x \in \mathbb{R}^n} \sup_{\substack{v \in \mathbb{R}^n \\ |v|=1}} v \cdot b'(x) v.$$

Then for any $t \geq 0$ and $x, y, v \in \mathbb{R}^n$,

$$|\partial_v \xi_t^x| \quad \leq \quad e^{\kappa t}|v|, \qquad \textit{and} \qquad |\xi_t^x - \xi_t^y| \quad \leq \quad e^{\kappa t}|x-y|.$$

The theorem shows in particular that exponential stability holds if $\kappa < 0$.

Proof. The derivative $Y_{v,t}^x = \partial_v \xi_t^x$ satisfies the ODE

$$dY_v \quad = \quad b'(\xi)Y_v\,dt.$$

Hence

$$d|Y_v|^2 \quad = \quad 2Y_v \cdot b'(\xi)Y_v\,dt \quad \leq \quad 2\kappa|Y_v|^2\,dt,$$

which implies

$$|\partial_v \xi_t^x|^2 \quad = \quad |Y_{v,t}^x|^2 \quad \leq \quad e^{2\kappa t}|v|^2, \qquad \text{and thus}$$

$$|\xi_t^x - \xi_t^y| \quad = \quad \left|\int_0^1 \partial_{x-y}\xi_t^{(1-s)x+sy}\,ds\right| \quad \leq \quad e^{\kappa t}|x-y|.$$

□

Example (Ornstein-Uhlenbeck process). Let $A \in \mathbb{R}^{n\times n}$. The generalized Ornstein-Uhlenbeck process solving the SDE

$$d\xi_t \quad = \quad A\xi_t\,dt + dW_t$$

is exponentially stable if $\kappa = \sup\{v \cdot Av \ : \ v \in S^{n-1}\} < 0$.

(ii) Non-constant diffusion coefficients. If the diffusion coefficients are not constant, the noise term in the SDE for the derivative flow does not vanish. Therefore, the derivative flow can not be bounded pathwise. Nevertheless, we can still obtain stability in an L^2 sense.

Lemma 13.3. *Suppose that $b, \sigma_1, \ldots, \sigma_d : \mathbb{R}^n \to \mathbb{R}^n$ are C^2 with bounded derivatives. Then for any $t \geq 0$ and $x, v \in \mathbb{R}^n$, the derivative flow $Y_{v,t}^x = \partial_v \xi_t^x$ is in $L^2(\Omega, \mathcal{A}, P)$, and*

$$\frac{d}{dt}E[|Y_{v,t}^x|^2] \quad = \quad 2E[Y_{v,t}^x \cdot K(\xi_t^x)Y_{v,t}^x]$$

where

$$K(x) \quad = \quad b'(x) + \frac{1}{2}\sum_{k=1}^{d} \sigma_k'(x)^T \sigma_k'(x).$$

Proof. Let $Y_v^{(k)}$ denote the k-the component of Y_v. The Itô product rule yields

$$\begin{aligned} d|Y_v|^2 \quad &= \quad 2Y_v \cdot dY_v + \sum_k d[Y_v^{(k)}] \\ &\overset{(13.1.10)}{=} \quad 2Y_v \cdot b'(\xi)Y_v\, dt + 2\sum_k Y_v \cdot \sigma_k'(\xi)\, dW^k + \sum_k |\sigma_k'(\xi)Y_v|^2\, dt. \end{aligned}$$

Noting that the stochastic integrals on the right-hand side stopped at
$T_n = \inf\{t \geq 0 : |Y_{v,t}| \geq n\}$ are martingales, we obtain

$$E\big[|Y_{v,t\wedge T_n}|^2\big] \quad = \quad |v|^2 + 2E\Big[\int_0^{t\wedge T_n} Y_v \cdot K(\xi) Y_v\, ds\Big].$$

The assertion follows as $n \to \infty$. □

Theorem 13.4 (Exponential stability II). *Suppose that the assumptions in Lemma 13.3 hold, and let*

$$\kappa \quad := \quad \sup_{x\in\mathbb{R}^n} \sup_{\substack{v\in\mathbb{R}^n \\ |v|=1}} v \cdot K(x)v. \tag{13.1.11}$$

Then for any $t \geq 0$ and $x, y, v \in \mathbb{R}^n$,

$$E[|\partial_v \xi_t^x|^2] \quad \leq \quad e^{2\kappa t}|v|^2, \quad \textit{and} \tag{13.1.12}$$

$$E[|\xi_t^x - \xi_t^y|^2]^{1/2} \quad \leq \quad e^{\kappa t}|x - y|. \tag{13.1.13}$$

Proof. Since $K(x) \leq \kappa I_n$ holds in the form sense for any x, Lemma 13.3 implies

$$\frac{d}{dt}E[|Y_{v,t}|^2] \quad \leq \quad 2\kappa E[|Y_{v,t}|^2].$$

(13.1.12) now follows immediately by Gronwell's lemma, and (13.1.13) follows from (13.1.12) since $\xi_t^x - \xi_t^y = \int_0^1 \partial_{x-y}\xi_t^{(1-s)x+sy}\, ds$. □

Remark. (Curvature) The quantity $-\kappa$ can be viewed as a lower curvature bound w.r.t. the geometric structure defined by the diffusion process. In particular, exponential stability w.r.t. the L^2 norm holds if $\kappa < 0$, i.e., if the curvature is bounded from below by a strictly positive constant.

Consequences for the transition semigroup

We still consider the flow (ξ_t^x) of the SDE (13.0.1) with assumptions as in Lemma 13.3 and Theorem 13.4. Let

$$p_t(x, B) \quad = \quad P[\xi_t^x \in B], \qquad x \in \mathbb{R}^n, \;\; B \in \mathcal{B}(\mathbb{R}^n),$$

denote the transition function of the diffusion process on $\mathbb{R}^n$. For two probability measures μ, ν on $\mathbb{R}^n$, we define the L^2 Wasserstein distance

$$\mathcal{W}_2(\mu, \nu) \quad = \quad \inf_{\substack{(X,Y) \\ X\sim\mu, Y\sim\nu}} E[|X-Y|^2]^{1/2}$$

as the infimum of the L^2 distance among all couplings of μ and ν. Here a coupling of μ and ν is defined as a pair (X, Y) of random variables on a joint probability space with distributions $X \sim \mu$ and $Y \sim \nu$. Let κ be defined as in (13.1.11).

Corollary 13.5. *For any $t \geq 0$ and $x, y \in \mathbb{R}^n$,*

$$\mathcal{W}_2\big(p_t(x, \cdot\,), p_t(y, \cdot\,)\big) \quad \leq \quad e^{\kappa t}|x - y|.$$

Proof. The flow defines a coupling between $p_t(x, \cdot\,)$ and $p_t(y, \cdot\,)$ for any t, x and y:

$$\xi_t^x \quad \sim \quad p_t(x, \cdot\,), \qquad \xi_t^y \;\sim\; p_t(y, \cdot\,).$$

Therefore,

$$\mathcal{W}_2\big(p_t(x, \cdot\,), p_t(y, \cdot\,)\big)^2 \quad \leq \quad E\big[|\xi_t^x - \xi_t^y|^2\big].$$

The assertion now follows from Theorem 13.4. ☐

Exercise (Exponential convergence to equilibrium). Suppose that μ is a stationary distribution for the diffusion process, i.e., μ is a probability measure on $\mathcal{B}(\mathbb{R}^n)$ satisfying $\mu p_t = \mu$ for every $t \geq 0$. Prove that if $\kappa < 0$ and $\int |x|^2 \, \mu(dx) < \infty$, then for any $x \in \mathbb{R}^d$, $\mathcal{W}_2\big(p_t(x, \cdot\,), \mu\big) \to 0$ exponentially fast with rate κ as $t \to \infty$.

Besides studying convergence to a stationary distribution, the derivative flow is also useful for computing and controlling derivatives of transtion functions. Let

$$(p_t f)(x) = \int p_t(x, dy) f(y) = E[f(\xi_t^x)]$$

denote the transition semigroup acting on functions $f : \mathbb{R}^n \to \mathbb{R}$. We still assume the conditions from Lemma 13.3.

Exercise (Lipschitz bound). Prove that for any Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$,

$$||p_t f||_{\mathrm{Lip}} \quad \leq \quad e^{\kappa t} ||f||_{\mathrm{Lip}} \qquad \forall\, t \geq 0,$$

where $||f||_{\mathrm{Lip}} = \sup\{|f(x) - f(y)|/|x - y| : x, y \in \mathbb{R}^n \text{ s.t. } x \neq y\}$.

For continuously differentiable functions f, we even obtain an explicit formula for the gradient of $p_t f$:

Corollary 13.6 (First Bismut-Elworthy Formula). *For any function $f \in C_b^1(\mathbb{R}^n)$ and $t \geq 0$, $p_t f$ is differentiable with*

$$v \cdot \nabla_x \, p_t f \quad = \quad E\big[Y_{v,t}^x \cdot \nabla_{\xi_t^x} f\big] \qquad \forall\, x, v \in \mathbb{R}^n. \tag{13.1.14}$$

Here $\nabla_x p_t f$ denotes the gradient evaluated at x. Note that $Y_{t,v}^x \cdot \nabla_{\xi_t^x} f$ is the directional derivative of f in the direction of the derivative flow $Y_{t,v}^x$.

Proof of 13.6. For $\lambda \in \mathbb{R} \setminus \{0\}$,

$$\frac{(p_t f)(x + \lambda v) - (p_t f)(x)}{\lambda} = \frac{1}{\lambda} E\big[f(\xi_t^{x+\lambda v}) - f(\xi_t^x)\big] = \frac{1}{\lambda} \int_0^\lambda E\big[Y_{v,t}^{x+sv} \cdot \nabla_{\xi_t^{x+sv}} f\big] \, ds.$$

The assertion now follows since $x \mapsto \xi_t^x$ and $x \mapsto Y_{v,t}^x$ are continuous, ∇f is continuous and bounded, and the derivative flow is bounded in L^2. $\square$

The first Bismut-Elworthy Formula shows that the gradient of $p_t f$ can be controlled by the gradient of f for all $t \geq 0$. In Section 13.3, we will see that by applying an integration by parts on the right hand side of (13.1.14), for $t > 0$ it is even possible to control the gradient of $p_t f$ in terms of the supremum norm of f, provided a non-degeneracy condition holds, cf. (**??**).

13.2 Malliavin gradient and Bismut integration by parts formula

Let $W_t(\omega) = \omega_t$ denote the canonical Brownian motion on $\Omega = C_0([0,1], \mathbb{R}^d)$ endowed with Wiener measure. In the sequel, we denote Wiener measure by P, expectation values w.r.t. Wiener measure by $E[\,\cdot\,]$, and the supremum norm by $||\cdot||$.

Definition. *Let $\omega \in \Omega$. A function $F : \Omega \to \mathbb{R}$ is called **Fréchet differentiable at ω** iff there exists a continuous linear functional $d_\omega F : \Omega \to \mathbb{R}$ such that*

$$||F(\omega + h) - F(\omega) - (d_\omega F)(h)|| \quad = \quad o(||h||) \qquad \textit{for any } \; h \in \Omega.$$

If a function F is Fréchet differentiable at ω then the directional derivatives

$$\frac{\partial F}{\partial h}(\omega) \quad = \quad \lim_{\varepsilon \to 0} \frac{F(\omega + \varepsilon h) - F(\omega)}{\varepsilon} \quad = \quad (d_\omega F)(h)$$

exist for all directions $h \in \Omega$. For applications in stochastic analysis, Fréchet differentiability is often too restrictive, because Ω contains "too many directions". Indeed, solutions of SDE are typically not Fréchet differentiable as the following example indicates:

Example. Let $F = \int_0^1 W_t^1 \, dW_t^2$ where $W_t = (W_t^1, W_t^2)$ is a two dimensional Brownian motion. A formal computation of the derivative of F in a direction $h = (h^1, h^2) \in \Omega$ yields

$$\frac{\partial F}{\partial h} \quad = \quad \int_0^1 h_t^1 \, dW_t^2 + \int_0^1 W_t^1 \, dh_t^2.$$

Clearly, this expression is NOT CONTINUOUS in h w.r.t. the supremum norm.

A more suitable space of directions for computing derivatives of stochastic integrals is the **Cameron-Martin space**

$$H_{CM} \quad = \quad \Big\{ h : [0,1] \to \mathbb{R}^d \;:\; h_0 = 0, \; h \text{ abs. contin. with } \; h' \in L^2([0,1], \mathbb{R}^d)] \Big\}.$$

Recall that H_{CM} is a Hilbert space with inner product

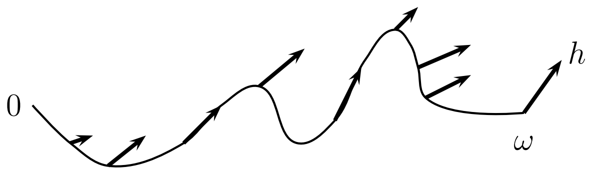
$$(h,g)_H \quad = \quad \int_0^1 h'_t \cdot g'_t \, dt, \qquad h,g \in H_{CM}.$$

The map $h \mapsto h'$ is an isometry from H_{CM} onto $L^2([0,1],\mathbb{R}^d)$. Moreover, H_{CM} is **continuously embedded** into Ω, since

$$||h|| \quad = \quad \sup_{t\in[0,1]} |h_t| \quad \leq \quad \int_0^1 |h'_t| \, dt \quad \leq \quad (h,h)_H^{1/2}$$

for any $h \in H_{CM}$ by the Cauchy Schwarz inequality.

As we will consider variations and directional derivatives in directions in H_{CM}, it is convenient to think of the Cameron-Martin space as a **tangent space** to Ω at a given path $\omega \in \Omega$. We will now define a gradient corresponding to the Cameron-Martin inner product in two steps: at first for smooth functions $F : \Omega \to \mathbb{R}$, and then for functions that are only weakly differentiable in a sense to be specified.



Gradient and integration by parts for smooth functions

Let $C_b^1(\Omega)$ denote the linear space consisting of all functions $F : \Omega \to \mathbb{R}$ that are everywhere Fréchet differentiable with continuous bounded derivative $dF : \Omega \to \Omega'$, $\omega \mapsto d_\omega F$. Here Ω' denotes the space of continuous linear functionals $l : \Omega \to \mathbb{R}$ endowed with the dual norm of the supremum norm, i.e.,

$$||l||_{\Omega'} \quad = \quad \sup \{ l(h) \; : \; h \in \Omega \;\text{ with }\; ||h|| \leq 1 \} \, .$$

Definition (Malliavin Gradient I). *Let $F \in C_b^1(\mathbb{R})$ and $\omega \in \Omega$.*

*1) The **H-gradient** $(D^H F)(\omega)$ is the unique element in H_{CM} satisfying*

$$\big((D^H F)(\omega), h\big)_H \quad = \quad \frac{\partial F}{\partial h}(\omega) \quad = \quad (d_\omega F)(h) \qquad \text{for any } h \in H_{CM}. \tag{13.2.1}$$

*2) The **Malliavin gradient** $(DF)(\omega)$ is the function $t \mapsto (D_t F)(\omega)$ in $L^2([0,1], \mathbb{R}^d)$ defined by*

$$(D_t F)(\omega) \quad = \quad \frac{d}{dt}(D^H F)(\omega)(t) \qquad \text{for a.e. } t \in [0,1]. \tag{13.2.2}$$

In other words, $D^H F$ is the usual gradient of F w.r.t. the Cameron-Martin inner product, and $(DF)(\omega)$ is the element in $L^2\big([0,1], \mathbb{R}^d\big)$ identified with $(D^H F)(\omega)$ by the canonical isometry $h \mapsto h'$ between H_{CM} and $L^2([0,1], \mathbb{R}^d)$. In particular, for any $h \in H_{CM}$ and $\omega \in \Omega$,

$$\begin{aligned} \frac{\partial F}{\partial h}(\omega) \quad &= \quad \big(h, (D^H F)(\omega)\big)_H \quad = \quad (h', (DF)(\omega))_{L^2} \\ &= \quad \int_0^1 h'_t \cdot (D_t F)(\omega)\, dt, \end{aligned} \tag{13.2.3}$$

and this identity characterizes DF completely. The examples given below should help to clarify the definitions.

Remark.

1) The existence of the H-gradient is guaranteed by the Riesz Representation Theorem. Indeed, for $\omega \in \Omega$ and $F \in C_b^1(\Omega)$, the Fréchet differential $d_\omega F$ is a continuous linear functional on Ω. Since H_{CM} is continuously embedded into Ω, the restriction to H_{CM} is a continuous linear functional on H_{CM} w.r.t. the H-norm. Hence there exists a unique element $(D^H F)(\omega)$ in H_{CM} such that (13.2.1) holds.

2) By definition of the Malliavin gradient,

$$||D^H F(\omega)||_H^2 \quad = \quad \int_0^1 |D_t F(\omega)|^2\, dt.$$

3) Informally, one may think of D_tF as a directional derivative of F in direction $I_{(t,1]}$, because

$$\text{“}\ D_tF \quad = \quad \frac{d}{dt}D^HF(t) \quad = \quad \int_0^1 (D^HF)'\, I'_{(t,1]} \quad = \quad \partial_{I_{(t,1]}}F\ \text{”}.$$

Of course, this is a purely heuristic representation, since $I_{(t,1]}$ is not even continuous.

Example (Linear functions on Wiener space).

1) **Brownian motion**: Consider the function $F(\omega) = W_s^i(\omega) = \omega_s^i$, where $s \in (0,1]$ and $i \in \{1,\dots,d\}$. Clearly, F is in $C_b^1(\Omega)$ and

$$\frac{\partial}{\partial h}W_s^i \quad = \quad \frac{d}{d\varepsilon}\big(W_s^i + \varepsilon h_s^i\big)\big|_{\varepsilon=0} \quad = \quad h_s^i \quad = \quad \int_0^1 h_t' \cdot e_i\, I_{(0,s)}(t)\, dt$$

for any $h \in H_{CM}$. Therefore, by the characterization in (13.2.3), the Malliavin gradient of F is given by

$$\big(D_tW_s^i\big)(\omega) \quad = \quad e_i\, I_{(0,s)}(t) \qquad \text{for every } \omega \in \Omega \text{ and a.e. } t \in (0,1).$$

Since the function $F : \Omega \to \mathbb{R}$ is linear, the gradient is deterministic. The H-gradient is obtained by integrating DW_s^i:

$$D_t^HW_s^i \quad = \quad \int_0^t D_rW_s^i\, dr \quad = \quad \int_0^t e_i\, I_{(0,s)} \quad = \quad (s \wedge t)\, e_i.$$

2) **Wiener integrals**: More generally, let

$$F \quad = \quad \int_0^1 g_s \cdot dW_s$$

where $g : [0,1] \to \mathbb{R}^d$ is a C^1 function. Integration by parts shows that

$$F \quad = \quad g_1 \cdot W_1 - \int_0^1 g_s' \cdot W_s\, ds \qquad \text{almost surely.} \tag{13.2.4}$$

The function on the right hand side of (13.2.4) is defined for **every** ω, and it is Fréchet differentiable. Taking this expression as a pointwise definition for the stochastic integral F, we obtain

$$\frac{\partial F}{\partial h} \quad = \quad g_1 \cdot h_1 - \int_0^1 g_s' \cdot h_s\, ds \quad = \quad \int_0^1 g_s \cdot h_s'\, ds$$

for any $h \in H_{CM}$. Therefore, by (13.2.3),

$$D_t F = g_t \quad \text{and} \quad D_t^H F = \int_0^t g_s \, ds.$$

Theorem 13.7 (Integration by parts, Bismut). *Let* $F \in C_b^1(\Omega)$ *and* $G \in \mathcal{L}_a^2(\Omega \times [0,1] \to \mathbb{R}^d, P \otimes \lambda)$. *Then*

$$E\Big[\int_0^1 D_t F \cdot G_t \, dt\Big] = E\Big[F \int_0^1 G_t \cdot dW_t\Big]. \tag{13.2.5}$$

To recognize (13.2.5) as an integration by parts identity on Wiener space let $H_t = \int_0^t G_s, ds$. Then

$$\int_0^1 D_t F \cdot G_t \, dt = \big(D^H F, H\big)_H = \partial_H F.$$

Replacing F in (13.2.5) by $F \cdot \widetilde{F}$ with $F, \widetilde{F} \in C_b^1(\Omega)$, we obtain the equivalent identity

$$E[F \, \partial_H \widetilde{F}] = -E[\partial_H F \, \widetilde{F}] + E\Big[F\widetilde{F} \int_0^1 G_t \cdot dW_t\Big] \tag{13.2.6}$$

by the product rule for the directional derivative.

Proof of Theorem 13.7. The formula (13.2.6) is an infinitesimal version of Girsanov's Theorem. Indeed, suppose first that G is bounded. Then, by Novikov's criterion,

$$Z_t^\varepsilon = \exp\Big(\varepsilon \int_0^t G_s \cdot dW_s - \frac{\varepsilon^1}{2} \int_0^t |G_s|^2 \, ds\Big)$$

is a martingale for any $\varepsilon \in \mathbb{R}$. Hence for $H_t = \int_0^t G_s \, ds$,

$$E[F(W + \varepsilon H)] = E[F(W) Z_1^\varepsilon].$$

The equation (13.2.6) now follows formally by taking the derivative w.r.t. ε at $\varepsilon = 0$. Rigorously, we have

$$E\Big[\frac{F(W + \varepsilon H) - F(W)}{\varepsilon}\Big] = E\Big[F(W) \frac{Z_1^\varepsilon - 1}{\varepsilon}\Big]. \tag{13.2.7}$$

As $\varepsilon \to 0$, the right hand side in (13.2.7) converges to $E\big[F(W)\int_0^t G \cdot dW\big]$, since

$$\frac{1}{\varepsilon}(Z_1^\varepsilon - 1) \quad = \quad \int_0^1 Z^\varepsilon G \cdot dW \quad \longrightarrow \quad \int_0^1 G \cdot dW \qquad \text{in } L^2(P).$$

Similarly, by the Dominated Convergence Theorem, the left hand side in (13.2.7) converges to the left hand side in (13.2.6):

$$E\Big[\frac{1}{\varepsilon}(F(W+\varepsilon H) - F(W))\Big] \;=\; E\Big[\int_0^\varepsilon (\partial_H F)(W+sH)\,ds\Big] \;\longrightarrow\; E[(\partial_H F)(W)]$$

as $\varepsilon \to 0$ since $F \in C_b^1(\Omega)$. We have shown that (13.2.6) holds for bounded adapted G. Moreover, the identity extends to any $G \in \mathcal{L}_a^2(P \otimes \lambda)$ because both sides of (13.2.6) are continuous in G w.r.t. the $L^2(P \otimes \lambda)$ norm. □

Remark. Adaptedness of G is essential for the validity of the integration by parts identity.

Skorokhod integral

The Bismut integration by parts formula shows that the adjoint of the Malliavin gradient coincides with the Itô integral on adapted processes. Indeed, the Malliavin gradient

$$\begin{aligned} D : C_b^1(\Omega) \subseteq L^2(\Omega, \mathcal{A}, P) &\longrightarrow L^2(\Omega \times [0,1] \to \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}, P \otimes \lambda), \\ F \qquad &\longmapsto \qquad (D_t F)_{0 \le t \le 1}, \end{aligned}$$

is a densely defined linear operator from the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ to the Hilbert space $L^2(\Omega \times [0,1] \to \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}, P \otimes \lambda)$. Let

$$\delta :\; \mathrm{Dom}(\delta) \subseteq L^2(\Omega \times [0,1] \to \mathbb{R}^d, \mathcal{A} \otimes B, P \otimes \lambda) \quad \longrightarrow \quad L^2(\Omega, \mathcal{A}, P)$$

denote the adjoint operator (i.e., the **divergence operator** corresponding to the Malliavin gradient). By (13.2.6), any adapted process $G \in \mathcal{L}^2(\Omega \times [0,1] \in \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}, P \otimes \lambda)$ is contained in the domain of δ, and

$$\delta G \quad = \quad \int_0^1 G_t \cdot dW_t \qquad \text{for any } G \in \mathcal{L}_a^2.$$

Hence the divergence operator δ defines an extension of the Itô integral $G \mapsto \int_0^1 G_t \cdot dW_t$ to not necessarily adapted square integrable processes $G : \Omega \times [0,1] \to \mathbb{R}^d$. This extension is called the **Skorokhod integral** .

Exercise (Product rule for divergence). Suppose that $(G_t)_{t\in[0,1]}$ is adapted and bounded, and $F \in C_b^1(\Omega)$. Prove that the process $(F \cdot G_t)_{t\in[0,1]}$ is contained in the domain of δ, and

$$\delta(FG) \quad = \quad F\delta(G) - \int_0^1 D_t F \cdot G_t \, dt.$$

Definition of Malliavin gradient II

So far we have defined the Malliavin gradient only for continuously Fréchet differentiable functions F on Wiener space. We will now extend the definition to the Sobolev spaces $\mathbb{D}^{1,p}$, $1 < p < \infty$, that are defined as closures of $C_b^1(\Omega)$ in $L^p(\Omega, \mathcal{A}, P)$ w.r.t. the norm

$$||F||_{1,p} \quad = \quad E\big[|F|^p + ||D^H F||_H^p\big]^{1/p}.$$

In particular, we will be interested in the case $p = 2$ where

$$||F||_{1,2}^2 \quad = \quad E\Big[F^2 + \int_0^1 |D_t F|^2 \, dt\Big].$$

Theorem 13.8 (Closure of the Malliavin gradient).

1) *There exists a unique extension of D^H to a continuous linear operator*

$$D^H : \quad \mathbb{D}^{1,p} \quad \longrightarrow \quad L^P(\Omega \to H, P)$$

.

2) *The Bismut integration by parts formula holds for any $F \in \mathbb{D}^{1,2}$.*

Proof for $p = 2$. 1) Let $F \in \mathbb{D}^{1,2}$ and let $(F_n)_{n\in\mathbb{N}}$ be a Cauchy sequence w.r.t. the $(1, 2)$ norm of functions in $C_b^1(\Omega)$ converging to F in $L^2(\Omega, P)$. We would like to define

$$D^H F \quad := \quad \lim_{n\to\infty} D^H F_n \tag{13.2.8}$$

w.r.t. convergence in the Hilbert space $L^2(\Omega \to H, P)$. The non-trivial fact to be shown is that $D^H F$ is **well-defined** by (13.2.8), i.e., independently of the approximating sequence. In functional analytic terms, this is the **closability** of the operator D^H.

To verify closability, we apply the integration by parts identity. Let (F_n) and $(\widetilde{F}_n)$ be approximating sequences as above, and let $L = \lim F_n$ and $\widetilde{L} = \lim \widetilde{F}_n$ in $L^2(\Omega, P)$. We have to show $L = \widetilde{L}$. To this end, it suffices to show

$$(L - \widetilde{L}, h)_H \quad = \quad 0 \qquad \text{almost surely for any } h \in H. \tag{13.2.9}$$

Hence fix $h \in H$, and let $\phi \in C_b^2(\Omega)$. Then by (13.2.6),

$$\begin{aligned} E[(L - \widetilde{L}, h)_H \cdot \phi] \quad &= \quad \lim_{n\to\infty} E[\partial_h (F_n - \widetilde{F}_n) \cdot \phi] \\ &= \quad \lim_{n\to\infty} \Big\{ E\big[(F_n - \widetilde{F}_n)\phi \int_0^1 h' \cdot dW\big] - E\Big[(F_n - \widetilde{F}_n)\partial_h \phi\Big]\Big\} \\ &= \quad 0 \end{aligned}$$

since $F_n - \widetilde{F}_n \to 0$ in L^2. As $C_b^1(\Omega)$ is dense in $L^2(\Omega, \mathcal{A}, P)$ we see that (13.2.9) holds.

2) To extend the Bismut integration by parts formula to functions $F \in \mathbb{D}^{1,2}$ let (F_n) be an approximating sequence of C_b^1 functions w.r.t. the $(1,2)$ norm. Then for any process $G \in \mathcal{L}_a^2$ and $H_t = \int_0^t G_s\, ds$, we have

$$E\Big[\int_0^1 D_t F_n \cdot G_t\, dt\Big] \quad = \quad E\Big[(D^H F_n, H)_H\Big] \quad = \quad E\Big[F_n \int_0^1 G \cdot dW\Big].$$

Clearly, both sides are continuous in F_n w.r.t. the $(1,2)$ norm, and hence the identity extends to F as $n \to \infty$. □

The next lemma is often useful to verify Malliavin differentiability:

Lemma 13.9. *Let $F \in L^2(\Omega, \mathcal{A}, P)$, and let $(F_n)_{n\in\mathbb{N}}$ be a sequence of functions in $\mathbb{D}^{1,2}$ converging to F w.r.t. the L^2 norm. If*

$$\sup_{n\in\mathbb{N}} E[||D^H F_n||_H^2] \quad < \quad \infty \tag{13.2.10}$$

then F is in $\mathbb{D}^{1,2}$, and there exists a subsequence $(F_{n_i})_{i\in\mathbb{N}}$ of (F_n) such that

$$\frac{1}{k}\sum_{i=1}^k F_{n_i} \quad \to \quad F \qquad \text{w.r.t. the (1,2) norm.} \tag{13.2.11}$$

The functional analytic proof is based on the theorems of Banach-Alaoglu and Banach-Saks, cf. e.g. the appendix in [30].

Proof. By (13.2.10), the sequence $(D^H F_n)_{n\in\mathbb{N}}$ of gradients is bounded in $L^2(\Omega \to H; P)$, which is a Hilbert space. Therefore, by the Banach-Alaoglu theorem , there exists a weakly convergent subsequence $(D^H F_{k_i})_{i\in\mathbb{N}}$. Moreover, by the Banach-Saks Theorem, there exists a subsequence $(D^H F_{n_i})_{i\in\mathbb{N}}$ of the first subsequence such that the averages $\frac{1}{k}\sum_{i=1}^k D^H F_{n_i}$ are even strongly convergent in $L^2(\Omega \to H; P)$. Hence the corresponding averages $\frac{1}{k}\sum_{i=1}^k F_{n_i}$ converge in $\mathbb{D}^{1,2}$. The limit is F since $F_{n_i} \to F$ in L^2 and the $\mathbb{D}^{1,2}$ norm is stronger than the L^2 norm. □

Product and chain rule

Lemma 13.9 can be used to extend the product and the chain rule to functions in $\mathbb{D}^{1,2}$.

Theorem 13.10. *1) If F and G are bounded functions in $\mathbb{D}^{1,2}$ then the product FG is again in $\mathbb{D}^{1,2}$, and*

$$D(FG) \quad = \quad F\,DG + G\,DF \qquad \text{a.s.}$$

2) Let $m \in \mathbb{N}$ and $F^{(1)},\dots,F^{(m)} \in \mathbb{D}^{1,2}$. If $\phi : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable with bounded derivatives then $\phi(F^{(1)},\dots,F^{(m)})$ is in $\mathbb{D}^{1,2}$, and

$$D\,\phi(F^{(1)},\dots,F^{(m)}) \quad = \quad \sum_{i=1}^m \frac{\partial\phi}{\partial x_i}(F^{(1)},\dots,F^{(m)})\,DF^{(i)}.$$

Proof. We only prove the product rule, whereas the proof of the chain rule is left as an exercise. Suppose that (F_n) and (G_n) are sequences of C_b^1 functions converging to F and G respectively in $\mathbb{D}^{1,2}$. If F and G are bounded then one can show that the approximating sequences (F_n) and (G_n) can be chosen uniformly bounded. In particular, $F_nG_n \to FG$ in L^2. By the product rule for the Fréchet differential,

$$\begin{aligned} D^H(F_nG_n) &= F_nD^HG_n + G_nD^HF_n \qquad \text{for any } n \in \mathbb{N}, \quad \text{and} \\ ||D^H(F_nG_n)||_H &\leq |F_n|\,||D^HG_n||_H + |G_n|\,||D^HF_n||_H. \end{aligned} \tag{13.2.12}$$

Thus the sequence $(D^H(F_nG_n))_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega \to H; P)$. By Lemma 13.9, we conclude that FG is in $\mathbb{D}^{1,2}$ and

$$D^H(FG) \quad = \quad L^2\text{-}\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} D^H(F_{n_i}G_{n_i})$$

for an appropriate subsequence. The product rule for FG now follows by (13.2.12). □

Clark-Ocone formula

Recall that by Itô's Representation Theorem, any function $F \in L^2(\Omega, \mathcal{A}, P)$ on Wiener space that is measurable w.r.t. the σ-algebra $\mathcal{F}_1 = \mathcal{F}_1^{W,P}$ can be represented as a stochastic integral of an $(\mathcal{F}_t)$ adapted process:

$$F - E[F] \quad = \quad \int_0^1 G_s \cdot dW_s \qquad \text{for some } G \in L_a^2(0,1).$$

If F is in $\mathbb{D}^{1,2}$ then the integration by parts formula on Wiener space can be used to identify the process G explicitly:

Theorem 13.11 (Clark-Ocone). *For any $F \in \mathbb{D}^{1,2}$,*

$$F - E[F] \quad = \quad \int_0^1 G \cdot dW$$

where

$$G_t \quad = \quad E[D_tF \mid \mathcal{F}_t].$$

Proof. It remains to identify the process G in the Itô representation. We assume w.l.o.g. that $E[F] = 0$. Let $H \in L^1_a([0,1],\mathbb{R}^d)$. Then by Itô's isometry and the integration by parts identity,

$$\begin{aligned} E\Big[\int_0^1 G_t \cdot H_t \, dt\Big] &= E\Big[\int_0^1 G \cdot dW \int_0^1 H \, dW\Big] = E\Big[\int_0^1 D_t F \cdot H_t \, dt\Big] \\ &= E\Big[\int_0^1 E[D_t F|\mathcal{F}_t] \cdot H_t \, dt\Big] \end{aligned}$$

for all Setting $H_t := G_t - E[D_t F|\mathcal{F}_t]$ we obtain

$$G_t(\omega) = E[D_t F|\mathcal{F}_t](\omega) \qquad P \otimes \lambda - \text{ a.e.}$$

□

13.3 First applications to stochastic differential equations

13.4 Existence and smoothness of densities

Chapter 14

Stochastic calculus for semimartingales with jumps

Our aim in this chapter is to develop a stochastic calculus for functions of finitely many real-valued stochastic processes $X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(d)}$. In particular, we will make sense of stochastic differential equations of type

$$dY_t \quad = \quad \sum_{k=1}^{d} \sigma_k(t, Y_{t-})\, dX_t^{(k)}$$

with continuous time-dependent vector fields $\sigma_1, \ldots, \sigma_d : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$. The sample paths of the driving processes $(X_t^{(k)})$ and of the solution (Y_t) may be discontinuous, but we will always assume that they are *càdlàg*, i.e., right-continuous with left limits. In most relevant cases this can be assured by choosing an appropriate modification. For example, a martingale or a Lévy process w.r.t. a right-continuous complete filtration always has a càdlàg modification, cf. [37, Ch.II, §2] and [36, Ch.I Thm.30].

An adequate class of stochastic processes for which a stochastic calculus can be developed are *semimartingales*, i.e., sums of local martingales and adapted finite variation processes with càdlàg trajectories. To understand why this is a reasonable class of processes to consider, we first briefly review the discrete time case.

Semimartingales in discrete time

If $(\mathcal{F}_n)_{n=0,1,2,...}$ is a discrete-time filtration on a probability space $(\Omega, \mathcal{A}, P)$ then any $(\mathcal{F}_n)$ adapted integrable stochastic process (X_n) has a unique Doob decomposition

$$X_n = X_0 + M_n + A_n^{\nearrow} - A_n^{\searrow} \tag{14.0.1}$$

into an $(\mathcal{F}_n)$ martingale (M_n) and non-decreasing predictable processes $(A_n^{\nearrow})$ and $(A_n^{\searrow})$ such that $M_0 = A_0^{\nearrow} = A_0^{\searrow} = 0$, cf. [14, Thm. 2.4]. The decomposition is determined by choosing

$$M_n - M_{n-1} \quad = \quad X_n - X_{n-1} - E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}],$$

$$A_n^{\nearrow} - A_{n-1}^{\nearrow} = E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}]^+, \quad \text{and} \quad A_n^{\searrow} - A_{n-1}^{\searrow} = E[X_n - X_{n-1} \mid \mathcal{F}_{n-1}]^-.$$

In particular, (X_n) is a sub- or supermartingale if and only if $A_n^{\searrow} = 0$ for any n, or $A_n^{\nearrow} = 0$ for any n, respectively. The discrete stochastic integral

$$(G_\bullet X)_n \quad = \quad \sum_{k=1}^n G_k\,(X_k - X_{k-1})$$

of a bounded predictable process (G_n) w.r.t. (X_n) is again a martingale if (X_n) is a martingale, and an increasing (decreasing) process if $G_n \geq 0$ for any n, and (X_n) is increasing (respectively decreasing). For a bounded adapted process (H_n), we can define correspondingly the integral

$$(H_{-\,\bullet} X)_n \quad = \quad \sum_{k=1}^n H_{k-1}\,(X_k - X_{k-1})$$

of the predictable process $H_- = (H_{k-1})_{k\in\mathbb{N}}$ w.r.t. X.

The Taylor expansion of a function $F \in C^2(\mathbb{R})$ yields a primitive version of the *Itô formula* in discrete time. Indeed, notice that for $k \in \mathbb{N}$,

$$\begin{aligned} F(X_k) - F(X_{k-1}) &= \int_0^1 F'(X_{k-1} + s\Delta X_k)\,ds\ \ \Delta X_k \\ &= F'(X_{k-1})\,\Delta X_k\ +\int_0^1\int_0^s F''(X_{k-1} + r\Delta X_k)\,dr\,ds\ \ (\Delta X_k)^2\,. \end{aligned}$$

where $\Delta X_k := X_k - X_{k-1}$. By summing over k, we obtain

$$F(X_n) \;=\; F(X_0) + (F'(X)_{-\bullet}X)_n + \sum_{k=1}^{n} \int_0^1 \int_0^s F''(X_{k-1} + r\Delta X_k)\, dr\, ds \;\; (\Delta X_k)^2 \,.$$

Itô's formula for a semimartingale (X_t) in continuous time will be derived in Theorem 14.22 below. It can be rephrased in a way similar to the formula above, where the last term on the right-hand side is replaced by an integral w.r.t. the quadratic variation process $[X]_t$ of X, cf. (XXX).

Semimartingales in continuous time

In continuous time, it is no longer true that any adapted process can be decomposed into a local martingale and an adapted process of finite variation (i.e., the sum of an increasing and a decreasing process). A counterexample is given by fractional Brownian motion, cf. Section 2.3 below. On the other hand, a large class of relevant processes has a corresponding decomposition.

Definition. *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. A real-valued $(\mathcal{F}_t)$-adapted stochastic process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$ is called an* ***$(\mathcal{F}_t)$ semimartingale*** *if and only if it has a decomposition*

$$X_t \;=\; X_0 + M_t + A_t, \qquad t \geq 0, \tag{14.0.2}$$

into a strict local $(\mathcal{F}_t)$-martingale (M_t) with càdlàg paths, and an $(\mathcal{F}_t)$-adapted process (A_t) with càdlàg finite-variation paths such that $M_0 = A_0 = 0$.

Here a **strict local martingale** is a process that can be localized by martingales with uniformly bounded jumps, see Section 2.2 for the precise definition. Any continuous local martingale is strict. In general, it can be shown that if the filtration is right continuous and complete then any local martingale can be decomposed into a strict local martingale and an adapted finite variation process ("Fundamental Theorem of Local Martingales", cf. [36]). Therefore, the notion of a semimartingale defined above is not changed if the word “strict” is dropped in the definition. Since the non-trivial proof of the Fundamental Theorem of Local Martingales is not included in these notes, we nevertheless stick to the definition above.

Remark. (Assumptions on path regularity). Requiring (A_t) to be càdlàg is just a standard convention ensuring in particular that $t \mapsto A_t(\omega)$ is the distribution function of a signed measure. The existence of right and left limits holds for any monotone function, and, therefore, for any function of finite variation. Similarly, every local martingale w.r.t. a right-continuous complete filtration has a càdlàg modification.

Without additional conditions on (A_t), the semimartingale decomposition in (14.0.2) is *not unique*, see the example below. Uniqueness holds if, in addition, (A_t) is assumed to be predictable, cf. [7, 36]. Under the extra assumption that (A_t) is continuous, uniqueness is a consequence of Corollary 14.15 below.

Example (Semimartingale decompositions of a Poisson process). An $(\mathcal{F}_t)$ Poisson process (N_t) with intensity λ has the semimartingale decompositions

$$N_t \;=\; \widetilde{N}_t + \lambda t \;=\; 0 + N_t$$

into a martingale and an adapted finite variation process. Only in the first decomposition, the finite variation process is predictable and continuous respectively.

The following examples show that semimartingales form a sufficiently rich class of stochastic processes.

Example (Stochastic integrals). Let (B_t) and (N_t) be a d-dimensional $(\mathcal{F}_t)$ Brownian motion and an $(\mathcal{F}_t)$ Poisson point process on a σ-finite measure space $(S, \mathcal{S}, \nu)$ respectively. Then any process of the form

$$X_t = \int_0^t H_s \cdot dB_s + \int_{(0,t]\times S} G_s(y)\widetilde{N}(ds\, dy) + \int_0^t K_s\, ds + \int_{(0,t]\times S} L_s(y) N(ds\, dy) \tag{14.0.3}$$

is a semimartingale provided the integrands H, G, K, L are predictable, H and G are (locally) square integrable w.r.t. $P \otimes \lambda$, $P \otimes \lambda \otimes \nu$ respectively, and K and L are (locally) integrable w.r.t. these measures. In particular, by the Lévy-Itô decomposition, *every Lévy process is a semimartingale.* Similarly, the components of *solutions of SDE driven by Brownian motions and Poisson point processes are semimartingales.* More generally, Itô's formula yields an explicit semimartingale decomposition of $f(t, X_t)$ for an arbitrary function $f \in \mathcal{C}^2(\mathbb{R}_+ \times \mathbb{R}^n)$ and (X_t) as above, cf. Section 14.4 below.

Example (Functions of Markov processes). If (X_t) is a time-homogeneous $(\mathcal{F}_t)$ Markov process on a probability space $(\Omega, \mathcal{A}, P)$, and f is a function in the domain of the generator $\mathcal{L}$, then $f(X_t)$ is a semimartingale with decomposition

$$f(X_t) = \text{local martingale } + \int_0^t (\mathcal{L}f)(X_s)\, ds, \tag{14.0.4}$$

cf. e.g. [12] or [16]. Indeed, it is possible to define the generator $\mathcal{L}$ of a Markov process through a solution to a martingale problem as in (14.0.4).

Many results for continuous martingales carry over to the càdlàg case. However, there are some important differences and pitfalls to be noted:

Exercise (Càdlàg processes).

1) A stopping time is called *predictable* iff there exists an increasing sequence (T_n) of stopping times such that $T_n < T$ on $\{T > 0\}$ and $T = \sup T_N$. Show that for a càdlàg stochastic process $(X_t)_{t\geq 0}$, the first hitting time

$$T_A = \inf\{t \geq 0 : X_t \in A\}$$

of a closed set $A \subset \mathbb{R}$ is *not predictable* in general.

2) Prove that for a right continuous $(\mathcal{F}_t)$ martingale $(M_t)_{t\geq 0}$ and an $(\mathcal{F}_t)$ stopping time T, the stopped process $(M_{t\wedge T})_{t\geq 0}$ is again an $(\mathcal{F}_t)$ martingale.

3) Prove that a càdlàg local martingale (M_t) can be localized by a sequence $(M_{t\wedge T_n})$ of bounded martingales provided the jumps of (M_t) are uniformly bounded, i.e.,

$$\sup\{|\Delta M_t(\omega)| : t \geq 0, \omega \in \Omega\} < \infty.$$

4) Give an example of a càdlàg local martingale that can not be localized by bounded martingales.

Our next goal is to define the stochastic integral $G_\bullet X$ w.r.t. a semimartingale X for the left limit process $G = (H_{t-})$ of an adapted càdlàg process H, and to build up a corresponding stochastic calculus. Before studying integration w.r.t. càdlàg martingales in Section 14.2, we will consider integrals and calculus w.r.t. finite variation processes in Section 14.1.

14.1 Finite variation calculus

In this section we extend Stieltjes calculus to càdlàg paths of finite variation. The results are completely deterministic. They will be applied later to the sample paths of the finite variation part of a semimartingale.

Fix $u \in (0, \infty]$, and let $A : [0, u) \to \mathbb{R}$ be a right-continuous function of finite variation. In particular, A is càdlàg. We recall that there is a σ-finite measure μ_A on $(0, u)$ with distribution function A, i.e.,

$$\mu_A\left((s,t]\right) \quad = \quad A_t - A_s \qquad \text{for any } 0 \le s \le t < u. \tag{14.1.1}$$

The function A has the decomposition

$$A_t \quad = \quad A_t^c + A_t^d \tag{14.1.2}$$

into the pure jump function

$$A_t^d \quad := \quad \sum_{s \le t} \Delta A_s \tag{14.1.3}$$

and the continuous function $A_t^c = A_t - A_t^d$. Indeed, the series in (14.1.3) converges absolutely since

$$\sum_{s\le t} |\Delta A_s| \quad \le \quad V_t^{(1)}(A) \quad < \infty \qquad \text{for any } t \in [0, u).$$

The measure μ_A can be decomposed correspondingly into

$$\mu_A \quad = \quad \mu_{A^c} + \mu_{A^d}$$

where

$$\mu_{A^d} \quad = \quad \sum_{\substack{s\in(0,u)\\ \Delta A_s\neq 0}} \Delta A_s \cdot \delta_s$$

is the atomic part, and μ_{A^c} does not contain atoms. Note that μ_{A^c} is not necessarily absolutely continuous!

Lebesgue-Stieltjes integrals revisited

Let $\mathcal{L}^1_{\text{loc}}([0,u),\mu_A) := \mathcal{L}^1_{\text{loc}}([0,u),|\mu_A|)$ where $|\mu_A|$ denotes the positive measure with distribution function $V^{(1)}_t(A)$. For $G \in \mathcal{L}^1_{\text{loc}}([0,u),\mu_A)$, the Lebesgue-Stieltjes integral of H w.r.t. A is defined as

$$\int_s^u G_r\, dA_r \quad = \quad \int G_r\, I_{(s,t]}(r)\, \mu_A(dr) \qquad \text{for } 0 \leq s \leq t < u.$$

A crucial observation is that the function

$$I_t \quad := \quad \int_0^t G_r\, dA_r \quad = \quad \int_{(0,t]} G_r\, \mu_A(dr) \quad , \quad t \in [0,u),$$

is the distribution function of the measure

$$\mu_I(dr) \quad = \quad G_r\, \mu_A(dr)$$

with density G w.r.t. μ_A. This has several important consequences:

1) The function I is again càdlàg and of finite variation with

$$V^{(1)}_t(I) \quad = \quad \int_0^t |G_r|\, |\mu_A|(dr) \quad = \quad \int_0^t |G_r|\, dV^{(1)}_r(A).$$

2) I decomposes into the continuous and pure jump parts

$$I^c_t \quad = \quad \int_0^t G_r\, dA^c_r \quad , \qquad I^d_t \quad = \quad \int_0^t G_r\, dA^d_r \quad = \quad \sum_{s\leq t} G_s\, \Delta A_s.$$

3) For any $\widetilde{G} \in \mathcal{L}^1_{\text{loc}}(\mu_I)$,

$$\int_0^t \widetilde{G}_r\, dI_r \quad = \quad \int_0^t \widetilde{G}_r G_r\, dA_r,$$

i.e., if "$dI = G\, dA$" then also "$\widetilde{G}\, dI = \widetilde{G}G\, dA$".

Theorem 14.1 (Riemann sum approximations for Lebesgue-Stieltjes integrals). *Suppose that $H : [0,u) \to \mathbb{R}$ is a càdlàg function. Then for any $a \in [0,u)$ and for any sequence (π_n) of partitions with* $\mathrm{mesh}(\pi_n) \to 0$*,*

$$\lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} H_s(A_{s'\wedge t} - A_s) \quad = \quad \int_0^t H_{s-}\, dA_s \qquad \textit{uniformly for } t \in [0,a].$$

Remark. If (A_t) is continuous then

$$\int_0^t H_{s-}\, dA_s \quad = \quad \int_0^t H_s\, dA_s,$$

because $\int_0^t \Delta H_s dA_s = \sum_{s\leq t} \Delta H_s \Delta A_s = 0$ for any càdlàg function H. In general, however, the limit of the Riemann sums in Theorem 14.1 takes the modified form

$$\int_0^t H_{s-}\, dA_s \quad = \quad \int_0^t H_s\, dA_s^c + \sum_{s\leq t} H_{s-}\Delta A_s.$$

Proof. For $n \in \mathbb{N}$ and $t \geq 0$,

$$\sum_{\substack{s\in\pi_n \\ s<t}} H_s(A_{s'\wedge t} - A_s) \quad = \quad \sum_{\substack{s\in\pi_n \\ s<t}} \int_{(s,s'\wedge t]} H_s\, dA_r \quad = \quad \int_{(0,t]} H_{\lfloor r\rfloor_n} dA_r$$

where $\lfloor r\rfloor_n := \max\{s \in \pi_n : s < r\}$ is the next partition point strictly below r. As $n \to \infty$, $\lfloor r\rfloor_n \to r$ from below, and thus $H_{\lfloor r\rfloor_n} \to H_{r-}$. Since the càdlàg function H is uniformly bounded on the compact interval $[0,a]$, we obtain

$$\sup_{t\leq a} \left| \int_0^t H_{\lfloor r\rfloor_n}\, dA_r - \int_0^t H_{r-}\, dA_r \right| \quad \leq \quad \int_{(0,a]} \left| H_{\lfloor r\rfloor_n} - H_{r-} \right| \, |\mu_A|(dr) \to 0$$

as $n \to \infty$ by dominated convergence. □

Product rule

The covariation $[H,A]$ of two functions $H, A : [0,u) \to \mathbb{R}$ w.r.t. a sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$ is defined by

$$[H,A]_t \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} (H_{s'\wedge t} - H_s)(A_{s'\wedge t} - A_s), \tag{14.1.4}$$

provided the limit exists. For finite variation functions, $[H, A]$ can be represented as a countable sum over the common jumps of H and A:

Lemma 14.2. *If H and A are càdlàg and A has finite variation then the covariation exists and is independently of (π_n) given by*

$$[H, A]_t \quad = \quad \sum_{0<s\leq t} \Delta H_s \Delta A_s$$

Proof. We again represent the sums as integrals:

$$\sum_{\substack{s\in\pi_n \\ s<t}} (H_{s'\wedge t} - H_s)(A_{s'\wedge t} - A_s) \quad = \quad \int_0^t (H_{\lceil r\rceil_n \wedge t} - H_{\lfloor r\rfloor_n})\, dA_r$$

with $\lfloor r\rfloor_n$ as above, and $\lceil r\rceil_n := \min\{s \in \pi_n \ : \ s \geq r\}$. As $n \to \infty$, $H_{\lceil r\rceil_n \wedge t} - H_{\lfloor r\rfloor_n}$ converges to $H_r - H_{r-}$, and hence the integral on the right hand side converges to

$$\int_0^t (H_r - H_{r-})\, dA_r \quad = \quad \sum_{r\leq t} \Delta H_r \Delta A_r$$

by dominated convergence. ☐

Remark. 1) If H or A is continuous then $[H, A] = 0$.
2) In general, the proof above shows that

$$\int_0^t H_s\, dA_s \quad = \quad \int_0^t H_{s-}\, dA_s + [H, A]_t,$$

i.e., $[H, A]$ is the difference between limits of right and left Riemann sums.

Theorem 14.3 (Integration by parts, product rule). *Suppose that $H, A : [0, u) \to \mathbb{R}$ are right continuous functions of finite variation. Then*

$$H_t A_t - H_0 A_0 \;=\; \int_0^t H_{r-}\, dA_r + \int_0^t A_{r-} dH_r + [H, A]_t \qquad \textit{for any } t \in [0, u). \quad (14.1.5)$$

In particular, the covariation $[H, A]$ is a càdlàg function of finite variation, and for $a < u$, the approximations in (14.1.4) converge uniformly on $[0, a]$ w.r.t. any sequence (π_n) such that $\mathrm{mesh}(\pi_n) \to 0$.

In differential notation, (14.1.5) reads

$$d(HA)_r \quad = \quad H_{r-}dA_r + A_{r-}dH_r + d[H,A]_r.$$

As special cases we note that if H and A are continuous then HA is continuous with

$$d(HA)_r \quad = \quad H_r\, dA_r + A_r\, dH_r,$$

and if H and A are pure jump functions (i.e. $H^c = A^c = 0$) then HA is a pure jump function with

$$\Delta(HA)_r \quad = \quad H_{r-}\Delta A_r + A_{r-}\Delta H_r + \Delta A_r \Delta H_r \qquad .$$

In the latter case, (14.1.5) implies

$$H_tA_t - H_0A_0 = \sum_{r\leq t} \Delta(HA)_r.$$

Note that this statement is not completely trivial, as it holds even when the jump times of HA form a countable dense subset of $[0,t]$!

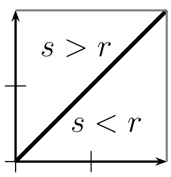
Since the product rule is crucial but easy to prove, we give two proofs of Theorem 14.3:

Proof 1. For (π_n) with $\mathrm{mesh}(\pi_n) \to 0$, we have

$$\begin{aligned} H_tA_t - H_0A_0 \quad &= \quad \sum_{\substack{s\in\pi_n \\ s<t}} (H_{s'\wedge t}A_{s'\wedge t} - H_sA_s) \\ = \sum H_s(A_{s'\wedge t} - A_s) &+ \sum A_s(H_{s'\wedge t} - H_s) + \sum (A_{s'\wedge t} - A_s)(H_{s'\wedge t} - H_s). \end{aligned}$$

As $n\to\infty$, (14.1.5) follows by Theorem 14.1 above. Moreover, the convergence of the covariation is uniform for $t \in [0,a]$, $a < u$, since this holds true for the Riemann sum approximations of $\int_0^t H_{s-}\, dA_s$ and $\int_0^t A_{s-}\, dH_s$ by Theorem 14.1. $\square$

Proof 2. Note that for $t \in [0,u)$,



$$(H_t - H_0)(A_t - A_0) \quad = \quad \int_{(0,t]\times(0,t]} \mu_H(dr)\, \mu_A(ds)$$

is the area of $(0,t] \times (0,t]$ w.r.t. the product measure $\mu_H \otimes \mu_A$. By dividing the square $(0,t]\times(0,t]$ into the parts $\{(s,r) \mid s < r\}$, $\{(s,r) \mid s > r\}$ and the diagonal $\{(s,r) \mid s = r\}$ we see that this area is given by

$$\int_{s<r} + \int_{s>r} + \int_{s=r} \quad = \quad \int_0^t (A_{r-} - A_0)\, dH_r + \int_0^t (H_{s-} - H_0)\, dA_s + \sum_{s\le t} \Delta H_s \Delta A_s,$$

The assertion follows by rearranging terms in the resulting equation. □

Chain rule

The chain rule can be deduced from the product rule by iteration and approximation of $\mathcal{C}^1$ functions by polynomials:

Theorem 14.4 (Change of variables, chain rule, Itô formula for finite variation functions). *Suppose that $A : [0,u) \to \mathbb{R}$ is right continuous with finite variation, and let $F \in \mathcal{C}^1(\mathbb{R})$. Then for any $t \in [0,u)$,*

$$F(A_t) - F(A_0) \quad = \quad \int_0^t F'(A_{s-})\, dA_s + \sum_{s\le t} \left(F(A_s) - F(A_{s-}) - F'(A_{s-})\Delta A_s\right), \tag{14.1.6}$$

or, equivalently,

$$F(A_t) - F(A_0) \quad = \quad \int_0^t F'(A_{s-})\, dA_s^c + \sum_{s\le t} \left(F(A_s) - F(A_{s-})\right). \tag{14.1.7}$$

If A is continuous then $F(A)$ is also continuous, and (14.1.6) reduces to the standard chain rule

$$F(A_t) - F(A_0) \quad = \quad \int_0^t F'(A_s)\, dA_s.$$

If A is a pure jump function then the theorem shows that $F(A)$ is also a pure jump function (this is again not completely obvious!) with

$$F(A_t) - F(A_0) \quad = \quad \sum_{s \le t} \left(F(A_s) - F(A_{s-})\right).$$

Remark. Note that by Taylor's theorem, the sum in (14.1.6) converges absolutely whenever $\sum_{s\le t}(\Delta A_s)^2 < \infty$. This observation will be crucial for the extension to Itô's formula for processes with finite quadratic variation, cf. Theorem 14.22 below.

Proof of Theorem 2.4. Let $\mathcal{A}$ denote the linear space consisting of all functions $F \in \mathcal{C}^1(\mathbb{R})$ satisfying (14.1.6). Clearly the constant function 1 and the identity $F(t) = t$ are in $\mathcal{A}$. We now prove that $\mathcal{A}$ is an algebra: Let $F, G \in \mathcal{A}$. Then by the integration by parts identity and by (14.1.7),

$$\begin{aligned}
(FG)(A_t) - &(FG)(A_0) \\
&= \int_0^t F(A_{s-})\, dG(A)_s + \int_0^t G(A_{s-})\, dF(A)_s + \sum_{s\le t} \Delta F(A)_s \Delta G(A)_s \\
&= \int_0^t \left(F(A_{s-})G'(A_{s-}) + G(A_{s-})F'(A_{s-})\right) dA_s^c \\
&\quad + \sum_{s\le t} \left(F(A_{s-})\Delta G(A)_s + G(A_{s-})\Delta F(A)_s + \Delta F(A)_s \Delta G(A)_s\right) \\
&= \int_0^t (FG)'(A_{s-})\, dA_s^c + \sum_{s\le t} \left((FG)(A_s) - (FG)(A_{s-})\right)
\end{aligned}$$

for any $t \in [0, u)$, i.e., FG is in $\mathcal{A}$.

Since $\mathcal{A}$ is an algebra containing 1 and t, it contains all polynomials. Moreover, if F is an arbitrary $\mathcal{C}^1$ function then there exists a sequence (p_n) of polynomials such that $p_n \to F$ and $p_n' \to F'$ uniformly on the bounded set $\{A_s \mid s \le t\}$. Since (14.1.7) holds for the polynomials p_n, it also holds for F. □

Exponentials of finite variation functions

Let $A : [0,\infty) \to \mathbb{R}$ be a right continuous finite variation function. The **exponential of A** is defined as the right-continuous finite variation function $(Z_t)_{t\geq 0}$ solving the equation

$$\begin{aligned} dZ_t &= Z_{t-}\, dA_t \quad , \qquad Z_0 = 1 \quad , \qquad \text{i.e.,} \\ Z_t &= 1 + \int_0^t Z_{s-}\, dA_s \qquad \text{for any } t \geq 0. \end{aligned} \tag{14.1.8}$$

If A is continuous then $Z_t = \exp(A_t)$ solves (14.1.8) by the chain rule. On the other hand, if A is piecewise constant with finitely many jumps then $Z_t = \prod_{s\leq t}(1+\Delta A_s)$ solves (14.1.8), since

$$Z_t \quad = \quad Z_0 + \sum_{s\leq t} \Delta Z_s \quad = \quad 1 + \sum_{s\leq t} Z_{s-}\Delta A_s \quad = \quad 1 + \int_{(0,t]} Z_{s-}\, dA_s.$$

In general, we obtain:

Theorem 14.5. *The unique càdlàg function solving (14.1.8) is*

$$Z_t \quad = \quad \exp(A_t^c) \cdot \prod_{s\leq t}(1+\Delta A_s), \tag{14.1.9}$$

where the product converges for any $t \geq 0$.

Proof. 1) We first show convergence of the product

$$P_t \quad = \quad \prod_{s\leq t}(1+\Delta A_s).$$

Recall that since A is càdlàg, there are only finitely many jumps with $|\Delta A_s| > 1/2$. Therefore, we can decompose

$$P_t \quad = \quad \exp\left(\sum_{\substack{s\leq t \\ |\Delta A_s|\leq 1/2}} \log(1+\Delta A_s) \right) \cdot \prod_{\substack{s\leq t \\ |\Delta A_s|>1/2}} (1+\Delta A_s) \tag{14.1.10}$$

in the sense that the product P_t converges if and only if the series converges. The series converges indeed absolutely for A with finite variation, since $\log(1+x)$ can be bounded by a constant times $|x|$ for $|x| \leq 1/2$. The limit S_t of the series defines a pure jump function with variation $V_t^{(1)}(S) \leq$ const. $\cdot\, V_t^{(1)}(A)$ for any $t \geq 0$.

2) *Equation for P_t*: The chain and product rule now imply by (14.1.10) that $t \mapsto P_t$ is also a finite variation pure jump function. Therefore,

$$P_t = P_0 + \sum_{s \leq t} \Delta P_s = 1 + \sum_{s \leq t} P_{s-} \Delta A_s = 1 + \int_0^t P_{s-}\, dA_s^d, \quad \forall t \geq 0, \tag{14.1.11}$$

i.e., P is the exponential of the pure jump part $A_t^d = \sum_{s \leq t} \Delta A_s$.

3) *Equation for Z_t*: Since $Z_t = \exp(A_t^c) P_t$ and $\exp(A^c)$ is continuous, the product rule and (14.1.11) imply

$$\begin{aligned} Z_t - 1 &= \int_0^t e^{A_s^c}\, dP_s + \int_0^t P_{s-}\, e^{A_s^c}\, dA_s^c \\ &= \int_0^t e^{A_s^c} P_{s-}\, d(A^d + A^c)_s = \int_0^t Z_{s-}\, dA_s. \end{aligned}$$

4) *Uniqueness*: Suppose that $\widetilde{Z}$ is another càdlàg solution of (14.1.8), and let $X_t := Z_t - \widetilde{Z}_t$. Then X solves the equation

$$X_t = \int_0^t X_{s-}\, dA_s \quad \forall\, t \geq 0$$

with zero initial condition. Therefore,

$$|X_t| \leq \int_0^t |X_{s-}|\, dV_t \leq M_t V_t \qquad \forall\, t \geq 0,$$

where $V_t := V_t^{(1)}(A)$ is the variation of A and $M_t := \sup_{s \leq t} |X_s|$. Iterating the estimate yields

$$|X_t| \leq M_t \int_0^t V_{s-}\, dV_s \leq M_t V_t^2/2$$

by the chain rule, and

$$|X_t| \leq \frac{M_t}{n!} \int_0^t V_{s-}^n\, dV_s \leq \frac{M_t}{(n+1)!} V_t^{n+1} \quad \forall\, t \geq 0, n \in \mathbb{N}. \tag{14.1.12}$$

Note that the correction terms in the chain rule are non-negative since $V_t \geq 0$ and $[V]_t \geq 0$ for all t. As $n \to \infty$, the right hand side in (14.1.12) converges to 0 since M_t and V_t are finite. Hence $X_t = 0$ for each $t \geq 0$. □

From now on we will denote the unique exponential of (A_t) by $(\mathcal{E}_t^A)$.

Remark (Taylor expansion). By iterating the equation (14.1.8) for the exponential, we obtain the convergent Taylor series expansion

$$\mathcal{E}_t^A \quad = \quad 1 \; + \; \sum_{k=1}^{n} \int_{(0,t]} \int_{(0,s_1)} \cdots \int_{(0,s_{n-1})} dA_{s_k} dA_{s_{k-1}} \cdots dA_{s_1} \; + \; R_t^{(n)},$$

where the remainder term can be estimated by

$$|R_t^{(n)}| \quad \leq \quad M_t V_t^{n+1}/(n+1)!.$$

If A is continuous then the iterated integrals can be evaluated explicitly:

$$\int_{(0,t]} \int_{(0,s_1)} \cdots \int_{(0,s_{k-1})} dA_{s_k} dA_{s_{k-1}} \cdots dA_{s_1} \quad = \quad (A_t - A_0)^k/k!.$$

If A is increasing but not necessarily continuous then the right hand side still is an upper bound for the iterated integral.

We now derive a formula for $\mathcal{E}_t^A \cdot \mathcal{E}_t^B$ where A and B are right-continuous finite variation functions. By the product rule and the exponential equation,

$$\begin{aligned}
\mathcal{E}_t^A \mathcal{E}_t^B - 1 \quad &= \quad \int_0^t \mathcal{E}_{s-}^A \, d\mathcal{E}_s^B + \int_0^t \mathcal{E}_{s-}^B \, d\mathcal{E}_s^A + \sum_{s \leq t} \Delta \mathcal{E}_s^A \Delta \mathcal{E}_s^B \\
&= \quad \int_0^t \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \, d(A+B)_s + \sum_{s \leq t} \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \Delta A_s \Delta B_s \\
&= \quad \int_0^t \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \, d(A+B+[A,B])_s
\end{aligned}$$

for any $t \geq 0$. This shows that in general, $\mathcal{E}^A \mathcal{E}^B \neq \mathcal{E}^{A+B}$.

Theorem 14.6. *If* $A, B : [0, \infty) \to \mathbb{R}$ *are right continuous with finite variation then*

$$\mathcal{E}^A \mathcal{E}^B \quad = \quad \mathcal{E}^{A+B+[A,B]}.$$

Proof. The left hand side solves the defining equation for the exponential on the right hand side. □

In particular, choosing $B = -A$, we obtain:

$$\frac{1}{\mathcal{E}^A} \quad = \quad \mathcal{E}^{-A+[A]}$$

Example (Geometric Poisson process). A **geometric Poisson process** with parameters $\lambda > 0$ and $\sigma, \alpha \in \mathbb{R}$ is defined as a solution of a stochastic differential equation of type

$$dS_t \quad = \quad \sigma S_{t-}\, dN_t + \alpha S_t\, dt \tag{14.1.13}$$

w.r.t. a Poisson process (N_t) with intensity λ. Geometric Poisson processes are relevant for financial models, cf. e.g. [39]. The equation (14.1.13) can be interpreted pathwise as the Stieltjes integral equation

$$S_t \quad = \quad S_0 + \sigma \int_0^t S_{r-}\, dN_r + \alpha \int_0^t S_r dr \qquad , \quad t \geq 0.$$

Defining $A_t = \sigma N_t + \alpha t$, (14.1.13) can be rewritten as the exponential equation

$$dS_t \quad = \quad S_{t-}\, dA_t \qquad ,$$

which has the unique solution

$$S_t \quad = \quad S_0 \cdot \mathcal{E}_t^A \quad = \quad S_0 \cdot e^{\alpha t} \prod_{s \leq t} (1 + \sigma \Delta N_s) \quad = \quad S_0 \cdot e^{\alpha t} (1 + \sigma)^{N_t}.$$

Note that for $\sigma > -1$, a solution (S_t) with positive initial value S_0 is positive for all t, whereas in general the solution may also take negative values. If $\alpha = -\lambda\sigma$ then (A_t) is a martingale. We will show below that this implies that (S_t) is a local martingale. Indeed, it is a true martingale which for $S_0 = 1$ takes the form

$$S_t \quad = \quad (1 + \sigma)^{N_t} e^{-\lambda \sigma t} \qquad .$$

Corresponding exponential martingales occur as "likelihood ratio" when the intensity of a Poisson process is modified, cf. Chapter 11 below.

Example (Exponential martingales for compound Poisson processes). For compound Poisson processes, we could proceed as in the last example. To obtain a different point of view, we go in the converse direction: Let

$$X_t \quad = \quad \sum_{j=1}^{K_t} \eta_j$$

be a compound Poisson process on $\mathbb{R}^d$ with jump intensity measure $\nu = \lambda\mu$ where $\lambda \in (0,\infty)$ and μ is a probability measure on $\mathbb{R}^d \setminus \{0\}$. Hence the η_j are i.i.d. $\sim \mu$, and (K_t) is an independent Poisson process with intensity λ. Suppose that we would like to change the jump intensity measure to an absolutely continuous measure $\bar{\nu}(dy) = \varrho(y)\nu(dy)$ with relative density $\varrho \in \mathcal{L}^1(\nu)$, and let $\bar{\lambda} = \bar{\nu}(\mathbb{R}^d \setminus \{0\})$. Intuitively, we could expect that the change of the jump intensity is achieved by changing the underlying probability measure P on $\mathcal{F}_t^X$ with relative density ("likelihood ratio")

$$Z_t \quad = \quad e^{(\lambda-\bar{\lambda})t} \prod_{j=1}^{K_t} \varrho(\eta_j) \quad = \quad e^{(\lambda-\bar{\lambda})t} \prod_{\substack{s \le t \\ \Delta X_s \neq 0}} \varrho(\Delta X_s).$$

In Chapter 11, as an application of Girsanov's Theorem, we will prove rigorously that this heuristics is indeed correct. For the moment, we identify (Z_t) as an exponential martingale. Indeed, $Z_t = \mathcal{E}_t^A$ with

$$\begin{aligned} A_t \quad &= \quad (\lambda - \bar{\lambda})t + \sum_{\substack{s \le t \\ \Delta X_s \neq 0}} (\varrho(\Delta X_s) - 1) \\ &= \quad -(\bar{\lambda} - \lambda)t + \int (\varrho(y) - 1)\ N_t(dy). \end{aligned} \tag{14.1.14}$$

Here $N_t = \sum_{j=1}^{K_t} \delta_{\eta_j}$ denotes the corresponding Poisson point process with intensity measure ν. Note that (A_t) is a martingale, since it is a compensated compound Poisson process

$$A_t \quad = \quad \int (\varrho(y) - 1)\ \widetilde{N}_t(dy) \quad , \quad \text{where } \widetilde{N}_t := N_t - t\nu.$$

By the results in the next section, we can then conclude that the exponential (Z_t) is a local martingale. We can write down the SDE

$$Z_t \quad = \quad 1 + \int_0^t Z_{s-}\, dA_s \tag{14.1.15}$$

in the equivalent form

$$Z_t \quad = \quad 1 + \int_{(0,t]\times\mathbb{R}^d}^t Z_{s-}\, (\varrho(y)-1)\, \widetilde{N}(ds\, dy) \tag{14.1.16}$$

where $\widetilde{N}(ds\, dy) := N(ds\, dy) - ds\, \nu(dy)$ is the random measure on $\mathbb{R}^+ \times \mathbb{R}^d$ with $\widetilde{N}((0,t]\times B) = \widetilde{N}_t(B)$ for any $t \geq 0$ and $B \in \mathcal{B}(\mathbb{R}^d)$. In differential notation, (14.1.16) is an SDE driven by the compensated Poisson point process $(\widetilde{N}_t)$:

$$dZ_t \quad = \quad \int_{y\in\mathbb{R}^d} Z_{t-}\, (\varrho(y)-1)\, \widetilde{N}(dt\, dy).$$

Example (Stochastic calculus for finite Markov chains). Functions of continuous time Markov chains on finite sets are semimartingales with finite variation paths. Therefore, we can apply the tools of finite variation calculus. Our treatment follows Rogers & Williams [38] where more details and applications can be found.

Suppose that (X_t) on $(\Omega, \mathcal{A}, \mathcal{P})$ is a continuous-time, time-homogeneous Markov process with values in a finite set S and càdlàg paths. We denote the transition matrices by p_t and the generator (Q-matrix) by $\mathcal{L} = (\mathcal{L}(a,b))_{a,b\in S}$. Thus $\mathcal{L} = \lim_{t\downarrow 0} t^{-1}(p_t - I)$, i.e., for $a \neq b$, $\mathcal{L}(a,b)$ is the jump rate from a to b, and $\mathcal{L}(a,a) = -\sum_{b\in S, b\neq a} \mathcal{L}(a,b)$ is the total (negative) intensity for jumping away from a. In particular,

$$(\mathcal{L}f)(a) \quad := \quad \sum_{b\in S} \mathcal{L}(a,b) f(b) \quad = \quad \sum_{b\in S, b\neq a} \mathcal{L}(a,b)(f(b) - f(a))$$

for any real-valued function $f = (f(a))_{a\in S}$ on S. It is a standard fact that $((X_t), P)$ solves the martingale problem for $\mathcal{L}$, i.e., the process

$$M_t^{[f]} \quad = \quad f(X_t) - \int_0^t (\mathcal{L}f)(X_s)\, ds \quad , \quad t \geq 0, \tag{14.1.17}$$

is an $(\mathcal{F}_t^X)$ martingale for any $f : S \to \mathbb{R}$. Indeed, this is a direct consequence of the Markov property and the Kolmogorov forward equation, which imply

$$\begin{aligned} E[M_t^{[f]} - M_s^{[f]} \mid \mathcal{F}_s^X] &= E[f(X_t) - f(X_s) - \int_s^t (\mathcal{L}f)(X_r)\, dr \mid \mathcal{F}_s] \\ &= (p_{t-s}f)(X_s) - f(X_s) - \int_s^t (p_{r-s}\mathcal{L}f)(X_s)\, ds = 0 \end{aligned}$$

for any $0 \leq s \leq t$. In particular, choosing $f = I_{\{b\}}$ for $b \in S$, we see that

$$M_t^b = I_{\{b\}}(X_t) - \int_0^t \mathcal{L}(X_s, b)\, ds \tag{14.1.18}$$

is a martingale, and, in differential notation,

$$dI_{\{b\}}(X_t) = \mathcal{L}(X_t, b)\, dt + dM_t^b. \tag{14.1.19}$$

Next, we note that by the results in the next section, the stochastic integrals

$$N_t^{a,b} = \int_0^t I_{\{a\}}(X_{s-})\, dM_s^b \quad , \; t \geq 0,$$

are martingales for any $a, b \in S$. Explicitly, for any $a \neq b$,

$$\begin{aligned} N_t^{a,b} &= \sum_{s \leq t} I_{\{a\}}(X_{s-}) \left(I_{S\setminus\{b\}}(X_{s-}) I_{\{b\}}(X_s) - I_{\{b\}}(X_{s-}) I_{S\setminus\{b\}}(X_s) \right) \\ &\quad - \int_0^t I_{\{a\}}(X_s)\, \mathcal{L}(X_s, b)\, ds \quad , \quad \text{i.e.,} \end{aligned}$$

$$N_t^{a,b} = J_t^{a,b} - \mathcal{L}(a, b)\, L_t^a \tag{14.1.20}$$

where $J_t^{a,b} = |\{s \leq t \, : \, X_{s-} = a, X_s = b\}|$ is the number of jumps from a to b until time t, and

$$L_t^a = \int_0^t I_a(X_s)\, ds$$

is the amount of time spent at a before time t (**"local time at a"**). In the form of an SDE,

$$dJ_t^{a,b} = \mathcal{L}(a, b)\, dL_t^a + dN_t^{a,b} \quad \text{for any } a \neq b. \tag{14.1.21}$$

More generally, for any function $g : S \times S \to \mathbb{R}$, the process

$$N_t^{[g]} = \sum_{a,b \in S} g(a, b) N_t^{a,b}$$

is a martingale. If $g(a,b) = 0$ for $a = b$ then by (14.1.20),

$$N_t^{[g]} \quad = \quad \sum_{s \le t} g(X_{s-}, X_s) - \int_0^t (\mathcal{L} g^T)(X_s, X_s)\, ds \tag{14.1.22}$$

Finally, the exponentials of these martingales are again local martingales. For example, we find that

$$\mathcal{E}_t^{\alpha N^{a,b}} \quad = \quad (1+\alpha)^{\mathcal{J}_t^{a,b}} \exp(-\alpha \mathcal{L}(a,b) L_t^a)$$

is an exponential martingale for any $\alpha \in \mathbb{R}$ and $a, b \in S$. These exponential martingales appear again as likelihood ratios when changing the jump rates of the Markov chains.

Exercise (Change of measure for finite Markov chains). Let (X_t) on $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ be a continuous time Markov chain with finite state space S and generator (Q-matrix) $\mathcal{L}$, i.e.,

$$M_t^{[f]} \ := \ f(X_t) - f(X_0) - \int_0^t (\mathcal{L} f)(X_s)\, ds$$

is a martingale w.r.t. P for each function $f : S \to \mathbb{R}$. We assume $\mathcal{L}(a,b) > 0$ for $a \neq b$. Let

$$g(a,b) \ := \ \widetilde{\mathcal{L}}(a,b)/\mathcal{L}(a,b) - 1 \quad \text{for } a \neq b, \qquad g(a,a) \ := \ 0,$$

where $\widetilde{\mathcal{L}}$ is another Q-matrix.

1) Let $\lambda(a) = \sum_{b \neq a} \mathcal{L}(a,b) = -\mathcal{L}(a,a)$ and $\widetilde{\lambda}(a) = -\widetilde{\mathcal{L}}(a,a)$ denote the total jump intensities at a. We define a "likelihood quotient" for the trajectories of Markov chains with generators $\widetilde{\mathcal{L}}$ and $\mathcal{L}$ by $Z_t = \widetilde{\zeta}_t / \zeta_t$ where

$$\widetilde{\zeta}_t \ = \ \exp\left(- \int_0^t \widetilde{\lambda}(X_s)\, ds \right) \prod_{s \le t : X_{s-} \neq X_s} \widetilde{\mathcal{L}}(X_{s-}, X_s),$$

and ζ_t is defined correspondingly. Prove that (Z_t) is the exponential of $(N_t^{[g]})$, and conclude that (Z_t) is a martingale with $E[Z_t] = 1$ for any t.

2) Let $\widetilde{P}$ denote a probability measure on $\mathcal{A}$ that is absolutely continuous w.r.t. P on $\mathcal{F}_t$ with relative density Z_t for every $t \ge 0$. Show that for any $f : S \to \mathbb{R}$,

$$\widetilde{M}_t^{[f]} \ := \ f(X_t) - f(X_0) - \int_0^t (\widetilde{\mathcal{L}} f)(X_s)\, ds$$

is a martingale w.r.t. $\widetilde{P}$. Hence under the new probability measure $\widetilde{P}$, (X_t) is a Markov chain with generator $\widetilde{\mathcal{L}}$.

Hint: You may assume without proof that $(\widetilde{M}_t^{[f]})$ *is a local martingale w.r.t.* $\widetilde{P}$ *if and only if* $(Z_t \widetilde{M}_t^{[f]})$ *is a local martingale w.r.t.* P. *A proof of this fact is given in Section 3.3.*

14.2 Stochastic integration for semimartingales

Throughout this section we fix a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$. We now define the stochastic integral of the left limit of an adapted càdlàg process w.r.t. a semimartingale in several steps. The key step is the first, where we prove the existence for the integral $\int H_{s-}\, dM_s$ of a *bounded* adapted càdlàg process H w.r.t. a *bounded* martingale M.

Integrals with respect to bounded martingales

Suppose that $M = (M_t)_{t\geq 0}$ is a uniformly bounded càdlàg $(\mathcal{F}_t^{\mathcal{P}})$ martingale, and $H = (H_t)_{t\geq 0}$ is a uniformly bounded càdlàg $(\mathcal{F}_t^{\mathcal{P}})$ adapted process. In particular, the left limit process

$$H_- \quad := \quad (H_{t-})_{t\geq 0}$$

is left continuous with right limits and $(\mathcal{F}_t^{\mathcal{P}})$ adapted. For a partition π of $\mathbb{R}_+$ we consider the elementary processes

$$H_t^\pi \quad := \quad \sum_{s\in\pi} H_s\, I_{[s,s')}(t), \quad \text{and} \quad H_{t-}^\pi \quad = \quad \sum_{s\in\pi} H_s\, I_{(s,s']}(t).$$

The process H^π is again càdlàg and adapted, and the left limit H_-^π is left continuous and (hence) predictable . We consider the Riemann sum approximations

$$I_t^\pi \quad := \quad \sum_{\substack{s\in\pi \\ s<t}} H_s(M_{s'\wedge t} - M_s)$$

to the integral $\int_0^t H_{s-}\, dM_s$ to be defined. Note that if we define the stochastic integral of an elementary process in the obvious way then

$$I_t^\pi \quad = \quad \int_0^t H_{s-}^\pi \, dM_s \qquad .$$

We remark that a straightforward pathwise approach for the existence of the limit of $I^\pi(\omega)$ as $\mathrm{mesh}(\pi) \to 0$ is doomed to fail, if the sample paths are not of finite variation:

Exercise. Let $\omega \in \Omega$ and $t \in (0, \infty)$, and suppose that (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. Prove that if $\sum_{\substack{s\in\pi \\ s<t}} h_s(M_{s'\wedge t}(\omega) - M_s(\omega))$ converges for every deterministic continuous function $h : [0, t] \to \mathbb{R}$ then $V_t^{(1)}(M(\omega)) < \infty$ (*Hint: Apply the Banach-Steinhaus theorem from functional analysis*).

The assertion of the exercise is just a restatement of the standard fact that the dual space of $\mathcal{C}([0, t])$ consists of measures with finite total variation. There are approaches to extend the pathwise approach by restricting the class of integrands further or by assuming extra information on the relation of the paths of the integrand and the integrator (Young integrals, rough paths theory, cf. [29], [19]). Here, following the standard development of stochastic calculus, we also restrict the class of integrands further (to predictable processes), but at the same time, we give up the pathwise approach. Instead, we consider stochastic modes of convergence.

For H and M as above, the process I^π is again a bounded càdlàg $(\mathcal{F}_t^{\mathcal{P}})$ martingale as is easily verified. Therefore, it seems natural to study convergence of the Riemann sum approximations in the space $M_d^2([0, a])$ of equivalence classes of càdlàg L^2-bounded $(\mathcal{F}_t^{\mathcal{P}})$ martingales defined up to a finite time a. The following fundamental theorem settles this question completely:

Theorem 14.7 (Convergence of Riemann sum approximations to stochastic integrals). *Let $a \in (0, \infty)$ and let M and H be as defined above. Then for every $\gamma > 0$ there exists a constant $\Delta > 0$ such that*

$$||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \quad < \quad \gamma \tag{14.2.1}$$

holds for any partitions π and $\widetilde{\pi}$ of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi) < \Delta$ and $\mathrm{mesh}(\widetilde{\pi}) < \Delta$.

The constant Δ in the theorem depends on M, H and a. The proof of the theorem for discontinuous processes is not easy, but it is worth the effort. For continuous processes, the proof simplifies considerably. The theorem can be avoided if one assumes existence of the quadratic variation of M. However, proving the existence of the quadratic variation requires the same kind of arguments as in the proof below (cf. [16]), or, alternatively, a lengthy discussion of general semimartingale theory (cf. [38]).

Proof of Theorem 14.7. Let $C \in (0,\infty)$ be a common uniform upper bound for the processes (H_t) and (M_t). To prove the estimate in (14.2.1), we assume w.l.o.g. that both partitions π and $\widetilde{\pi}$ contain the end point a, and π is a refinement of $\widetilde{\pi}$. If this is not the case, we may first consider a common refinement and then estimate by the triangle inequality. Under the additional assumption, we have

$$I_a^\pi - I_a^{\widetilde{\pi}} \quad = \quad \sum_{s\in\pi}(H_s - H_{\lfloor s\rfloor})(M_{s'} - M_s) \tag{14.2.2}$$

where from now on, we only sum over partition points less than a, s' denotes the successor of s in the fine partition π, and

$$\lfloor s\rfloor \quad := \quad \max\{t \in \widetilde{\pi} \,:\, t \le s\}$$

is the next partition point of the rough partition $\widetilde{\pi}$ below s. Now fix $\varepsilon > 0$. By (14.2.2), the martingale property for M, and the adaptedness of H, we obtain

$$\begin{aligned} ||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \quad &= \quad E\big[(I_a^\pi - I_a^{\widetilde{\pi}})^2\big] \\ = E\Big[\sum_{s\in\pi}(H_s - H_{\lfloor s\rfloor})^2(M_{s'} - M_s)^2\Big] & \qquad\qquad (14.2.3) \\ \le \varepsilon^2 E\Big[\sum_{s\in\pi}(M_{s'} - M_s)^2\Big] + (2C)^2 E\Big[\sum_{t\in\widetilde{\pi}} &\sum_{\substack{s\in\pi \\ \tau_t(\varepsilon)\le s<\lceil t\rceil}} (M_{s'} - M_s)^2\Big] \end{aligned}$$

where $\lceil t\rceil := \min\{u \in \widetilde{\pi} \,:\, u > t\}$ is the next partition point of the rough partition, and

$$\tau_t(\varepsilon) \quad := \quad \min\{s \in \pi, s > t \,:\, |H_s - H_t| > \varepsilon\} \wedge \lceil t\rceil.$$

is the first time after t where H deviates substantially from H_s. Note that τ_t is a random variable.

The summands on the right hand side of (14.2.3) are now estimated separately. Since M is a bounded martingale, we can easily control the first summand:

$$E\big[\sum(M_{s'}-M_s)^2\big] = \sum E\big[M_{s'}^2 - M_s^2\big] = E\big[M_a^2 - M_0^2\big] \le C^2. \qquad (14.2.4)$$

The second summand is more difficult to handle. Noting that

$$E\big[(M_{s'}-M_s)^2 \mid \mathcal{F}_{\tau_t}\big] \quad = \quad E\big[M_{s'}^2 - M_s^2 \mid \mathcal{F}_{\tau_t}\big] \qquad \text{on} \quad \{\tau_t \le s\}\,,$$

we can rewrite the expectation value as

$$\begin{aligned} &\sum_{t\in\widetilde{\pi}} E\big[\sum_{\tau_t \le s < \lceil t \rceil} E\big[(M_{s'}-M_s)^2 \mid \mathcal{F}_{\tau_t}\big]\big] && (14.2.5)\\ &\quad = \sum_{t\in\widetilde{\pi}} E\big[E\big[M_{\lceil t\rceil}^2 - M_{\tau_t}^2 \mid \mathcal{F}_{\tau_t}\big]\big] = E\big[\sum_{t\in\widetilde{\pi}}(M_{\lceil t\rceil} - M_{\tau_t})^2\big] \quad =: \quad B \end{aligned}$$

Note that $M_{\lceil t\rceil} - M_{\tau_t} \neq 0$ only if $\tau_t < \lceil t \rceil$, i.e., if H oscillates more than ε in the interval $[t, \tau_t]$. We can therefore use the càdlàg property of H and M to control (14.2.5). Let

$$D_{\varepsilon/2} \quad := \quad \{r \in [0,a] \ : \ |H_r - H_{r-}| > \varepsilon/2\}$$

denote the (random) set of "large" jumps of H. Since H is càdlàg, $D_{\varepsilon/2}$ contains only finitely many elements. Moreover, for given $\varepsilon, \bar{\varepsilon} > 0$ there exists a random variable $\delta(\omega) > 0$ such that for $u, v \in [0,a]$,

(i) $|u-v| \le \delta \quad \Rightarrow \quad |H_u - H_v| \le \varepsilon \quad$ or $\quad (u,v] \cap D_{\varepsilon/2} \neq \emptyset$,

(ii) $r \in D_{\varepsilon/2}\,, \quad u, v \in [r, r+\delta] \quad \Rightarrow \quad |M_u - M_v| \le \bar{\varepsilon}$.

Here we have used that H is càdlàg, $D_{\varepsilon/2}$ is finite, and M is right continuous.

Let $\Delta > 0$. By (i) and (ii), the following implication holds on $\{\Delta \le \delta\}$:

$$\tau_t < \lceil t\rceil \quad \Rightarrow \quad |H_{\tau_t} - H_t| > \varepsilon \quad \Rightarrow \quad [t,\tau_t] \cap D_{\varepsilon/2} \neq \emptyset \quad \Rightarrow \quad |M_{\lceil t\rceil} - M_{\tau_t}| \le \bar{\varepsilon},$$

i.e., if $\tau_t < \lceil t\rceil$ and $\Delta \le \delta$ then the increment of M between τ_t and $\lceil t\rceil$ is small.

Now fix $k \in \mathbb{N}$ and $\bar{\varepsilon} > 0$. Then we can decompose $B = B_1 + B_2$ where

$$B_1 \quad = \quad E\big[\sum_{t\in\widetilde{\pi}}(M_{\lceil t\rceil} - M_{\tau_t})^2 \,;\, \Delta \le \delta,\ |D_{\varepsilon/2}| \le k\big] \quad \le \quad k\bar{\varepsilon}^2, \tag{14.2.6}$$

$$\begin{aligned} B_2 \quad &= \quad E\big[\sum_{t\in\widetilde{\pi}}(M_{\lceil t\rceil} - M_{\tau_t})^2 \,;\, \Delta > \delta \text{ or } |D_{\varepsilon/2}| > k\big] \\ &\le \quad E\big[(\sum_{t\in\widetilde{\pi}}(M_{\lceil t\rceil} - M_{\tau_t})^2)^2\big]^{1/2}\, P\big[\Delta > \delta \text{ or } |D_{\varepsilon/2}| > k\big]^{1/2} \qquad (14.2.7) \\ &\le \quad \sqrt{6}\, C^2\big(P\big[\Delta > \delta\big] + P\big[|D_{\varepsilon/2}| > k\big]\big)^{1/2} \quad . \end{aligned}$$

In the last step we have used the following upper bound for the martingale increments $\eta_t := M_{\lceil t\rceil} - M_{\tau_t}$:

$$\begin{aligned} E\big[\big(\sum_{t\in\widetilde{\pi}}\eta_t^2\big)^2\big] \quad &= \quad E\big[\sum_t \eta_t^4\big] + 2E\big[\sum_t\sum_{u>t}\eta_t^2\eta_u^2\big] \\ &\le \quad 4C^2E\big[\sum_t \eta_t^2\big] + 2E\big[\sum_t \eta_t^2 E\big[\sum_{u>t}\eta_u^2 \mid \mathcal{F}_t\big]\big] \\ &\le 6C^2E\big[\sum_t \eta_t^2\big] \quad \le \quad 6C^2E\big[M_a^2 - M_0^2\big] \quad \le \quad 6C^4. \end{aligned}$$

This estimate holds by the Optional Sampling Theorem, and since $E[\sum_{u>t}\eta_u^2 \mid \mathcal{F}_t] \le E[M_u^2 - M_t^2 \mid \mathcal{F}_t] \le C^2$ by the orthogonality of martingale increments $M_{T_{i+1}} - M_{T_i}$ over disjoint time intervals $(T_i, T_{i+1}]$ bounded by stopping times.

We now summarize what we have shown. By (14.2.3), (14.2.4) and (14.2.5),

$$||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \quad \le \quad \varepsilon^2C^2 + 4C^2(B_1 + B_2) \tag{14.2.8}$$

where B_1 and B_2 are estimated in (14.2.6) and (14.2.7). Let $\gamma > 0$ be given. To bound the right hand side of (14.2.8) by γ we choose the constants in the following way:

1. Choose $\varepsilon > 0$ such that $C^2\varepsilon^2 < \gamma/4$.
2. Choose $k \in \mathbb{N}$ such that $4\sqrt{6}\, C^4P\big[|D_{\varepsilon/2}| > k\big]^{1/2} < \gamma/4$,
3. Choose $\bar{\varepsilon} > 0$ such that $4C^2k\bar{\varepsilon}^2 < \gamma/4$, then choose the random variable δ depending on ε and $\bar{\varepsilon}$ such that (i) and (ii) hold.
4. Choose $\Delta > 0$ such that $4\sqrt{6}\, C^4P\big[\Delta > \delta\big]^{1/2} < \gamma/4$.

Then for this choice of Δ we finally obtain

$$||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \quad < \quad 4\cdot\frac{\gamma}{4} \quad = \quad \gamma$$

whenever $\mathrm{mesh}(\widetilde{\pi}) \leq \Delta$ and π is a refinement of $\widetilde{\pi}$. □

The theorem proves that the stochastic integral $H_{-\bullet}M$ is well-defined as an M^2 limit of the Riemann sum approximations:

Definition (Stochastic integral for left limits of bounded adapted càdlàg processes w.r.t. bounded martingales). *For H and M as above, the stochastic integral $H_{-\bullet}M$ is the unique equivalence class of càdlàg $(\mathcal{F}_t^P)$ martingales on $[0,\infty)$ such that*

$$H_{-\bullet}M\big|_{[0,a]} \quad = \quad \lim_{n\to\infty} H_{-\bullet}^{\pi_n}M\big|_{[0,a]} \qquad \text{in } M_d^2([0,a])$$

for any $a \in (0,\infty)$ and for any sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$.

Note that the stochastic integral is defined uniquely only up to càdlàg modifications. We will often denote versions of $H_{-\bullet}M$ by $\int_0^\bullet H_{s-}\,dM_s$, but we will not always distinguish between equivalence classes and their representatives carefully. Many basic properties of stochastic integrals with left continuous integrands can be derived directly from the Riemann sum approximations:

Lemma 14.8 (Elementary properties of stochastic integrals). *For H and M as above, the following statements hold:*

1) *If $t \mapsto M_t$ has almost surely finite variation then $H_{-\bullet}M$ coincides almost surely with the pathwise defined Lebesgue-Stieltjes integral $\int_0^\bullet H_{s-}\,dM_s$.*
2) *$\Delta(H_{-\bullet}M) = H_-\Delta M$ almost surely.*
3) *If $T : \Omega \to [0,\infty]$ is a random variable, and H, $\widetilde{H}$, M, $\widetilde{M}$ are processes as above such that $H_t = \widetilde{H}_t$ for any $t < T$ and $M_t = \widetilde{M}_t$ for any $t \leq T$ then, almost surely,*

$$H_{-\bullet}M \quad = \quad \widetilde{H}_{-\bullet}\widetilde{M} \qquad \text{on } [0,T].$$

Proof. The statements follow easily by Riemann sum approximation. Indeed, let (π_n) be a sequence of partitions of $\mathbb{R}_+$ such that $\mathrm{mesh}(\pi_n) \to 0$. Then almost surely along a subsequence $(\widetilde{\pi}_n)$,

$$(H_{-\bullet}M)_t \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\le t\\ s\in\widetilde{\pi}_n}} H_s(M_{s'\wedge t} - M_s)$$

w.r.t. uniform convergence on compact intervals. This proves that $H_{-\bullet}M$ coincides almost surely with the Stieltjes integral if M has finite variation. Moreover, for $t > 0$ it implies

$$\Delta(H_{-\bullet}M)_t \quad = \quad \lim_{n\to\infty} H_{\lfloor t\rfloor_n}(M_t - M_{t-}) \quad = \quad H_{t-}\Delta M_t \tag{14.2.9}$$

almost surely, where $\lfloor t\rfloor_n$ denotes the next partition point of $(\widetilde{\pi}_n)$ below t. Since both $H_{-\bullet}M$ and M are càdlàg, (14.2.9) holds almost surely simultaneously for all $t > 0$. The third statement can be proven similarly. □

Localization

We now extend the stochastic integral to local martingales. It turns out that unbounded jumps can cause substantial difficulties for the localization. Therefore, we restrict ourselves to local martingales that can be localized by martingales with bounded jumps. Remark 2 below shows that this is not a substantial restriction.

Suppose that $(M_t)_{t\ge 0}$ is a càdlàg $(\mathcal{F}_t)$ adapted process, where $(\mathcal{F}_t)$ is an arbitrary filtration. For an $(\mathcal{F}_t)$ stopping time T, the stopped process M^T is defined by

$$M_t^T \quad := \quad M_{t\wedge T} \qquad \text{for any } t \ge 0.$$

Definition (Local martingale, Strict local martingale). *A* ***localizing sequence*** *for M is a non-decreasing sequence $(T_n)_{n\in\mathbb{N}}$ of $(\mathcal{F}_t)$ stopping times such that* $\sup T_N = \infty$, *and the stopped process M^{T_n} is an $(\mathcal{F}_t)$ martingale for each n. The process M is called a* ***local $(\mathcal{F}_t)$ martingale*** *iff there exists a localizing sequence. Moreover, M is called a*

strict local $(\mathcal{F}_t)$ ***martingale*** *iff there exists a localizing sequence* (T_n) *such that* M^{T_n} *has uniformly bounded jumps for each* n*, i.e.,*

$$\sup\{|\Delta M_t(\omega)|\;:\;0\le t\le T_n(\omega)\,,\;\omega\in\Omega\}\quad<\quad\infty\quad\forall\,n\in\mathbb{N}.$$

Remark. 1) Any continuous local martingale is a strict local martingale.

2) In general, any local martingale is the sum of a strict local martingale and a local martingale of finite variation. This is the content of the "Fundamental Theorem of Local Martingales", cf. [36]. The proof of this theorem, however, is not trivial and is omitted here.

The next example indicates how (local) martingales can be decomposed into strict (local) martingales and finite variation processes:

Example (Lévy martingales). Suppose that $X_t=\int y\,(N_t(dy)-t\nu(dy))$ is a compensated Lévy jump process on $\mathbb{R}^1$ with intensity measure ν satisfying $\int(|y|\wedge|y|^2)\,\nu(dy)<\infty$. Then (X_t) is a martingale but, in general, not a strict local martingale. However, we can easily decompose $X_t=M_t+A_t$ where $A_t=\int y\;I_{\{|y|>1\}}\;(N_t(dy)-t\;\nu(dy))$ is a finite variation process, and $M_t=\int yI_{\{|y|\le1\}}\;(N_t(dy)-t\nu(dy))$ is a strict (local) martingale.

Strict local martingales can be localized by bounded martingales:

Lemma 14.9. *M is a strict local martingale if and only if there exists a localizing sequence (T_n) such that M^{T_n} is a bounded martingale for each n.*

Proof. If M^{T_n} is a bounded martingale then also the jumps of M^{T_n} are uniformly bounded. To prove the converse implication, suppose that (T_n) is a localizing sequence such that ΔM^{T_n} is uniformly bounded for each n. Then

$$S_n\quad:=\quad T_n\wedge\inf\{t\ge0\;:\;|M_t|\ge n\}\quad,\quad n\in\mathbb{N},$$

is a non-decreasing sequence of stopping times with $\sup S_n=\infty$, and the stopped processes M^{S_n} are uniformly bounded, since

$$|M_{t\wedge S_n}|\quad\le\quad n+|\Delta M_{S_n}|\quad=\quad n+|\Delta M^{T_n}_{S_n}|\qquad\text{for any }\;t\ge0.$$

□

Definition (Stochastic integrals of left limits of adapted càdlàg processes w.r.t. strict local martingales). *Suppose that* $(M_t)_{t\geq 0}$ *is a strict local* $(\mathcal{F}_t^{\mathcal{P}})$ *martingale, and* $(H_t)_{t\geq 0}$ *is càdlàg and* $(\mathcal{F}_t^{\mathcal{P}})$ *adapted. Then the stochastic integral* $H_{-\bullet}M$ *is the unique equivalence class of local* $(\mathcal{F}_t^{\mathcal{P}})$ *martingales satisfying*

$$H_{-\bullet}M\big|_{[0,T]} \quad = \quad \widetilde{H}_{-\bullet}\widetilde{M}\big|_{[0,T]} \quad a.s., \tag{14.2.10}$$

whenever T *is an* $(\mathcal{F}_t^{\mathcal{P}})$ *stopping time,* $\widetilde{H}$ *is a bounded càdlàg* $(\mathcal{F}_t^{\mathcal{P}})$ *adapted process with* $H|_{[0,T)} = \widetilde{H}|_{[0,T)}$ *almost surely, and* $\widetilde{M}$ *is a bounded càdlàg* $(\mathcal{F}_t^{\mathcal{P}})$ *martingale with* $M\big|_{[0,T]} = \widetilde{M}\big|_{[0,T]}$ *almost surely.*

You should convince yourself that the integral $H_{-\bullet}M$ is well defined by (14.2.10) because of the local dependence of the stochastic integral w.r.t. bounded martingales on H and M (Lemma 14.8, 3). Note that $\widetilde{H}_t$ and H_t only have to agree for $t < T$, so we may choose $\widetilde{H}_t = H_t \cdot I_{\{t<T\}}$. This is crucial for the localization. Indeed, we can always find a localizing sequence (T_n) for M such that both $H_t \cdot I_{\{t<T_n\}}$ and M^{T_n} are bounded, whereas the process H^T stopped at an exit time from a bounded domain is not bounded in general!

Remark (Stochastic integrals of càdlàg integrands w.r.t. strict local martingales are again strict local martingales). This is a consequence of Lemma 14.9 and Lemma 14.8, 2: If (T_n) is a localizing sequence for M such that both $H^{(n)} = H \cdot I_{[0,T_n)}$ and M^{T_n} are bounded for every n then

$$H_{-\bullet}M \quad = \quad H^{(n)}_{-\bullet}M^{T_n} \qquad \text{on} \quad [0, T_n],$$

and, by Lemma 14.8, $\Delta(H^{(n)}_{-\bullet}M^{T_n}) = H^{(n)}_{-}\Delta M^{T_n}$ is uniformly bounded for each n.

Integration w.r.t. semimartingales

The stochastic integral w.r.t. a semimartingale can now easily be defined via a semimartingale decomposition. Indeed, suppose that X is an $(\mathcal{F}_t^{\mathcal{P}})$ semimartingale with decomposition

$$X_t \quad = \quad X_0 + M_t + A_t \quad , \quad t \geq 0,$$

into a strict local $(\mathcal{F}_t^{\mathcal{P}})$ martingale M and an $(\mathcal{F}_t^{\mathcal{P}})$ adapted process A with càdlàg finite-variation paths $t \mapsto A_t(\omega)$.

Definition (Stochastic integrals of left limits of adapted càdlàg processes w.r.t. semimartingales). *For any $(\mathcal{F}_t^{\mathcal{P}})$ adapted process $(H_t)_{t\geq 0}$ with càdlàg paths, the stochastic integral of H w.r.t. X is defined by*

$$H_{-\bullet}X \quad = \quad H_{-\bullet}M + H_{-\bullet}A,$$

where M and A are the strict local martingale part and the finite variation part in a semimartingale decomposition as above, $H_{-\bullet}M$ is the stochastic integral of H_- w.r.t. M, and $(H_{-\bullet}A)_t = \int_0^t H_{s-}\, dA_s$ is the pathwise defined Stieltjes integral of H_- w.r.t. A.

Note that the semimartingale decomposition of X is not unique. Nevertheless, the integral $H_{-\bullet}X$ is uniquely defined up to modifications:

Theorem 14.10. *Suppose that (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\operatorname{mesh}(\pi_n) \to 0$. Then for any $a \in [0, \infty)$,*

$$(H_{-\bullet}X)_t \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} H_s(X_{s'\wedge t} - X_s)$$

w.r.t. uniform convergence for $t \in [0, a]$ in probability, and almost surely along a subsequence. In particular:

1) *The definition of $H_{-\bullet}X$ does not depend on the chosen semimartingale decomposition.*
2) *The definition does not depend on the choice of a filtration $(\mathcal{F}_t)$ such that X is an $(\mathcal{F}_t^{\mathcal{P}})$ semimartingale and H is $(\mathcal{F}_t^{\mathcal{P}})$ adapted.*
3) *If X is also a semimartingale w.r.t. a probability measure Q that is absolutely continuous w.r.t. $\mathcal{P}$ then each version of the integral $(H_{-\bullet}X)_{\mathcal{P}}$ defined w.r.t. $\mathcal{P}$ is a version of the integral $(H_{-\bullet}X)_Q$ defined w.r.t. Q.*

The proofs of this and the next theorem are left as exercises to the reader.

Theorem 14.11 (Elementary properties of stochastic integrals).

1) ***Semimartingale decomposition**: The integral $H_{-\bullet}X$ is again an $(\mathcal{F}_t^{\mathcal{P}})$ semimartingale with decomposition $H_{-\bullet}X = H_{-\bullet}M + H_{-\bullet}A$ into a strict local martingale and an adapted finite variation process.*
2) ***Linearity**: The map $(H, X) \mapsto H_{-\bullet}X$ is bilinear.*
3) ***Jumps**: $\Delta(H_{-\bullet}X) = H_-\Delta X$ almost surely.*
4) ***Localization**: If T is an $(\mathcal{F}_t^{\mathcal{P}})$ stopping time then*

$$(H_{-\bullet}X)^T \quad = \quad H_{-\bullet}X^T \quad = \quad (H \cdot I_{[0,T)})_{-\bullet}X.$$

14.3 Quadratic variation and covariation

From now on we fix a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with a filtration $(\mathcal{F}_t)$. The vector space of (equivalence classes of) **strict** local $(\mathcal{F}_t^{\mathcal{P}})$ martingales and of $(\mathcal{F}_t^{\mathcal{P}})$ adapted processes with càdlàg finite variation paths are denoted by M_{loc} and FV respectively. Moreover,

$$\mathcal{S} \quad = \quad M_{\text{loc}} + \text{FV}$$

denotes the vector space of $(\mathcal{F}_t^{\mathcal{P}})$ semimartingales. If there is no ambiguity, we do not distinguish carefully between equivalence classes of processes and their representatives.

The stochastic integral induces a bilinear map $\mathcal{S} \times \mathcal{S} \to \mathcal{S}$, $(H, X) \mapsto H_{-\bullet}X$ on the equivalence classes that maps $\mathcal{S} \times M_{\text{loc}}$ to M_{loc} and $\mathcal{S} \times \text{FV}$ to FV.
A suitable notion of convergence on (equivalence classes of) semimartingales is uniform convergence in probability on compact time intervals:

Definition (ucp-convergence). *A sequence of semimartingales $X_n \in \mathcal{S}$ converges to a limit $X \in \mathcal{S}$* ***uniformly on compact intervals in probability*** *iff*

$$\sup_{t \le a} |X_t^n - X_t| \xrightarrow{\mathcal{P}} 0 \qquad \text{as } n \to \infty \qquad \text{for any } \; a \in \mathbb{R}_+.$$

By Theorem (14.10), for $H, X \in \mathcal{S}$ and any sequence of partitions with $\text{mesh}(\pi_n) \to 0$, the stochastic integral $\int H_- \, dX$ is a ucp-limit of predictable Riemann sum approximations, i.e., of the integrals of the elementary predictable processes $H_-^{\pi_n}$.

Covariation and integration by parts

The covariation is a symmetric bilinear map $\mathcal{S} \times \mathcal{S} \to \text{FV}$. Instead of going once more through the Riemann sum approximations, we can use what we have shown for stochastic integrals and define the covariation by the integration by parts identity

$$X_t Y_t - X_0 Y_0 \quad = \quad \int_0^t X_{s-} \, dY_s + \int_0^t Y_{s-} \, dX_s + [X, Y]_t.$$

The approximation by sums is then a direct consequence of Theorem 14.10.

Definition (Covariation of semimartingales). *For $X, Y \in \mathcal{S}$,*

$$[X, Y] \quad := \quad XY - X_0 Y_0 - \int X_- \, dY - \int Y_- \, dX.$$

Clearly, $[X, Y]$ is again an $(\mathcal{F}_t^{\mathcal{P}})$ adapted càdlàg process. Moreover, $(X, Y) \mapsto [X, Y]$ is symmetric and bilinear, and hence the polarization identity

$$[X, Y] \quad = \quad \frac{1}{2} ([X + Y] - [X] - [Y])$$

holds for any $X, Y \in \mathcal{S}$ where

$$[X] = [X, X]$$

denotes the **quadratic variation** of X. The next corollary shows that $[X, Y]$ deserves the name "covariation":

Corollary 14.12. *For any sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$,*

$$[X, Y]_t = \mathrm{ucp} - \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s). \tag{14.3.1}$$

In particular, the following statements hold almost surely:

1) $[X]$ is non-decreasing, and $[X, Y]$ has finite variation.

2) $\Delta[X, Y] = \Delta X \Delta Y$.

3) $[X, Y]^T = [X^T, Y] = [X, Y^T] = [X^T, Y^T]$.

4) $|[X, Y]| \leq [X]^{1/2}[Y]^{1/2}$.

Proof. (14.3.1) is a direct consequence of Theorem 14.10, and 1) follows from (14.3.1) and the polarization identity. 2) follows from Theorem 14.11, which yields

$$\begin{aligned}
\Delta[X, Y] &= \Delta(XY) - \Delta(X_{-\bullet}Y) - \Delta(Y_{-\bullet}X) \\
&= X_-\Delta Y + Y_-\Delta X + \Delta X\Delta Y - X_-\Delta Y - Y_-\Delta X \\
&= \Delta X\Delta Y.
\end{aligned}$$

3) follows similarly and is left as an exercise and 4) holds by (14.3.1) and the Cauchy-Schwarz formula for sums. □

Statements 1) and 2) of the corollary show that $[X, Y]$ is a finite variation process with decomposition

$$[X, Y]_t = [X, Y]^c_t + \sum_{s\leq t} \Delta X_s \Delta Y_s \tag{14.3.2}$$

into a continuous part and a pure jump part.

If Y has finite variation then by Lemma 14.2,

$$[X,Y]_t \quad = \quad \sum_{s\leq t} \Delta X_s \Delta Y_s.$$

Thus $[X,Y]^c = 0$ and if, moreover, X or Y is continuous then $[X,Y] = 0$.

More generally, if X and Y are semimartingales with decompositions $X = M + A$, $Y = N + B$ into $M, N \in M_{\text{loc}}$ and $A, B \in \text{FV}$ then by bilinearity,

$$[X,Y]^c \quad = \quad [M,N]^c + [M,B]^c + [A,N]^c + [A,B]^c \quad = [M,N]^c.$$

It remains to study the covariations of the local martingale parts which turn out to be the key for controlling stochastic integrals effectively.

Quadratic variation and covariation of local martingales

If M is a strict local martingale then by the integration by parts identity, $M^2 - [M]$ is a strict local martingale as well. By localization and stopping we can conclude:

Theorem 14.13. *Let $M \in \mathcal{M}_{loc}$ and $a \in [0,\infty)$. Then $M \in \mathcal{M}_d^2([0,a])$ if and only if $M_0 \in \mathcal{L}^2$ and $[M]_a \in \mathcal{L}^1$. In this case, $M_t^2 - [M]_t$ $(0 \leq t \leq a)$ is a martingale, and*

$$||M||^2_{M^2([0,a])} \quad = \quad E\big[M_0^2\big] + E\big[[M]_a\big]. \tag{14.3.3}$$

Proof. We may assume $M_0 = 0$; otherwise we consider $\widetilde{M} = M - M_0$. Let (T_n) be a joint localizing sequence for the local martingales M and $M^2 - [M]$ such that M^{T_n} is bounded. Then by optional stopping,

$$E\big[M^2_{t\wedge T_n}\big] \quad = \quad E\big[[M]_{t\wedge T_n}\big] \qquad \text{for any } \ t \geq 0 \quad \text{and any } n \in \mathbb{N}. \tag{14.3.4}$$

Since M^2 is a submartingale, we have

$$E[M_t^2] \quad \leq \quad \liminf_{n\to\infty} E[M^2_{t\wedge T_n}] \quad \leq \quad E[M_t^2] \tag{14.3.5}$$

by Fatou's lemma. Moreover, by the Monotone Convergence Theorem,

$$E\big[[M]_t\big] \quad = \quad \lim_{n\to\infty} E\big[[M]_{t\wedge T_n}\big].$$

Hence by (14.3.5), we obtain

$$E[M_t^2] \quad = \quad E\big[[M]_t\big] \qquad \text{for any } t \geq 0.$$

For $t \leq a$, the right-hand side is dominated from above by $E\big[[M]_a\big]$, Therefore, if $[M]_a$ is integrable then M is in $M_d^2([0,a])$ with M^2 norm $E\big[[M]_a\big]$. Moreover, in this case, the sequence $\left(M_{t\wedge T_n}^2 - [M]_{t\wedge T_n}\right)_{n\in\mathbb{N}}$ is uniformly integrable for each $t \in [0,a]$, because,

$$\sup_{t\leq a} |M_t^2 - [M]_t| \quad \leq \quad \sup_{t\leq a} |M_t|^2 + [M]_a \quad \in \mathcal{L}^1,$$

Therefore, the martingale property carries over from the stopped processes $M_{t\wedge T_n}^2 - [M]_{t\wedge T_n}$ to $M_t^2 - [M]_t$. □

Remark. The assertion of Theorem 14.13 also remains valid for $a = \infty$ in the sense that if M_0 is in $\mathcal{L}^2$ and $[M]_\infty = \lim_{t\to\infty}[M]_t$ is in $\mathcal{L}^1$ then M extends to a square integrable martingale $(M_t)_{t\in[0,\infty]}$ satisfying (14.3.4) with $a = \infty$. The existence of the limit $M_\infty = \lim_{t\to\infty} M_t$ follows in this case from the L^2 Martingale Convergence Theorem.

The next corollary shows that the M^2 norms also control the covariations of square integrable martingales.

Corollary 14.14. *The map $(M,N) \mapsto [M,N]$ is symmetric, bilinear and continuous on $M_d^2([0,a])$ in the sense that*

$$E[\sup_{t\leq a} |[M,N]|_t] \quad \leq \quad ||M||_{M^2([0,a])}||N||_{M^2([0,a])}.$$

Proof. By the Cauchy-Schwarz inequality for the covariation (Cor. 14.12,4),

$$|[M,N]_t| \quad \leq \quad [M]_t^{1/2}[N]_t^{1/2} \quad \leq \quad [M]_a^{1/2}[N]_a^{1/2} \quad \forall\, t \leq a.$$

Applying the Cauchy-Schwarz inequality w.r.t. the L^2-inner product yields

$$E[\sup_{t\leq a}|[M,N]_t|] \quad \leq \quad E\big[[M]_a\big]^{1/2}E\big[[N]_a\big]^{1/2} \quad \leq \quad ||M||_{M^2([0,a])}||N||_{M^2([0,a])}$$

by Theorem 14.13. □

Corollary 14.15. *Let* $M \in \mathcal{M}_{loc}$ *and suppose that* $[M]_a = 0$ *almost surely for some* $a \in [0,\infty]$. *Then almost surely,*

$$M_t \quad = \quad M_0 \qquad \text{for any } t \in [0,a].$$

In particular, continuous local martingales of finite variation are almost surely constant.

Proof. By Theorem 14.13, $||M - M_0||_{M^2([0,a])} = E\big[[M]_a\big] = 0$. □

The assertion also extends to the case when a is replaced by a stopping time. Combined with the existence of the quadratic variation, we have now proven:

»Non-constant strict local martingales have non-trivial quadratic variation«

Example (Fractional Brownian motion is not a semimartingale). Fractional Brownian motion with Hurst index $H \in (0,1)$ is defined as the unique continuous Gaussian process $(B_t^H)_{t\geq 0}$ satisfying

$$E\big[B_t^H\big] \quad = \quad 0 \qquad \text{and} \qquad \mathrm{Cov}\big[B_s^H, B_t^H\big] \quad = \quad \frac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big)$$

for any $s,t \geq 0$. It has been introduced by Mandelbrot as an example of a self-similar process and is used in various applications, cf. [2]. Note that for $H = 1/2$, the covariance is equal to $\min(s,t)$, i.e., $B^{1/2}$ is a standard Brownian motion. In general, one can prove that fractional Brownian motion exists for any $H \in (0,1)$, and the sample paths $t \mapsto B_t^H(\omega)$ are almost surely α-Hölder continuous if and only if $\alpha < H$, cf. e.g. [19]. Furthermore,

$$V_t^{(1)}(B^H) \quad = \quad \infty \qquad \text{for any } t > 0 \text{ almost surely}, \quad \text{and}$$

$$[B^H]_t \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} \left(B^H_{s'\wedge t} - B^H_s\right)^2 \quad = \quad \begin{cases} 0 & \text{if } H > 1/2 \ , \\ t & \text{if } H = 1/2 \ , \\ \infty & \text{if } H < 1/2 \ . \end{cases}$$

Since $[B^H]_t = \infty$, fractional Brownian motion is **not a semimartingale** for $H < 1/2$. Now suppose that $H > 1/2$ and assume that there is a decomposition $B^H_t = M_t + A_t$ into a continuous local martingale M and a continuous finite variation process A. Then

$$[M] \quad = \quad [B^H] \quad = \quad 0 \qquad \text{almost surely} \ ,$$

so by Corollary 14.1.11, M is almost surely constant, i.e., B^H has finite variation paths. Since this is a contradiction, we see that also for $H > 1/2$, B^H is **not a continuous semimartingale**, i.e., the sum of a continuous local martingale and a continuous adapted finite variation process. It is possible (but beyond the scope of these notes) to prove that any semimartingale that is continuous is a continuous semimartingale in the sense above (cf. [36]). Hence for $H \neq 1/2$, fractional Brownian motion is not a semimartingale and classical Itô calculus is not applicable. Rough paths theory provides an alternative way to develop a calculus w.r.t. the paths of fractional Brownian motion, cf. [19].

The covariation $[M, N]$ of local martingales can be characterized in an alternative way that is often useful for determining $[M, N]$ explicitly.

Theorem 14.16 (Martingale characterization of covariation). *For $M, N \in M_{loc}$, the covariation $[M, N]$ is the unique process $A \in \mathrm{FV}$ such that*

(i) $MN - A \quad \in M_{loc} \quad , \quad$ *and*

(ii) $\Delta A \quad = \quad \Delta M\, \Delta N \quad , \quad A_0 = 0 \quad$ *almost surely* .

Proof. Since $[M, N] = MN - M_0N_0 - \int M_-\, dN - \int N_-\, dM$, (i) and (ii) are satisfied for $A = [M, N]$. Now suppose that $\widetilde{A}$ is another process in FV satisfying (i) and (ii). Then $A - \widetilde{A}$ is both in M_{loc} and in FV, and $\Delta(A - \widetilde{A}) = 0$ almost surely. Hence $A - \widetilde{A}$ is a continuous local martingale of finite variation, and thus $A - \widetilde{A} = A_0 - \widetilde{A}_0 = 0$ almost surely by Corollary 14.15. $\square$

The covariation of two local martingales M and N yields a semimartingale decomposition for MN:

$$MN \quad = \quad \text{local martingale} \; + \; [M, N].$$

However, such a decomposition is not unique. By Corollary 14.15 it is unique if we assume in addition that the finite variation part A is continuous with $A_0 = 0$ (which is not the case for $A = [M, N]$ in general).

Definition. *Let $M, N \in M_{loc}$. If there exists a continuous process $A \in \mathrm{FV}$ such that*

(i) $MN - A \quad \in M_{loc}$, *and*

(ii) $\Delta A \quad = \quad 0 \quad , \quad A_0 \quad = \quad 0$ *almost surely,*

then $\langle M, N \rangle = A$ is called the ***conditional covariance process of M and N****.*

In general, a conditional covariance process as defined above need not exist. General martingale theory (Doob-Meyer decomposition) yields the existence under an additional assumption if continuity is replaced by predictability, cf. e.g. [36]. For applications it is more important that in many situations the conditional covariance process can be easily determined explicitly, see the example below.

Corollary 14.17. *Let $M, N \in M_{loc}$.*

1) *If M is continuous then $\langle M, N \rangle = [M, N]$ almost surely.*
2) *In general, if $\langle M, N \rangle$ exists then it is unique up to modifications.*
3) *If $\langle M \rangle$ exists then the assertions of Theorem 14.13 hold true with $[M]$ replaced by $\langle M \rangle$.*

Proof. 1) If M is continuous then $[M, N]$ is continuous.
2) Uniqueness follows as in the proof of 14.16.
3) If (T_n) is a joint localizing sequence for $M^2 - [M]$ and $M^2 - \langle M \rangle$ then, by monotone convergence,

$$E\big[\langle M \rangle_t\big] \quad = \quad \lim_{n\to\infty} E\big[\langle M \rangle_{t\wedge T_n}\big] \quad = \quad \lim_{n\to\infty} E\big[[M]_{t\wedge T_n}\big] \quad = \quad E\big[[M]_t\big]$$

for any $t \geq 0$. The assertions of Theorem 14.13 now follow similarly as above. $\square$

Examples (Covariations of Lévy processes).

1) **Brownian motion**: If (B_t) is a Brownian motion in $\mathbb{R}^d$ then the components (B_t^k) are independent one-dimensional Brownian motions. Therefore, the processes $B_t^k B_t^l - \delta_{kl} t$ are martingales, and hence almost surely,

$$[B^k, B^l]_t \quad = \quad \langle B^k, B^l \rangle_t \quad = \quad t \cdot \delta_{kl} \quad \text{for any } \ t \geq 0.$$

2) **Lévy processes without diffusion part**: Let

$$X_t \quad = \quad \int_{\mathbb{R}^d \setminus \{0\}} y \left(N_t(dy) - t\, I_{\{|y| \leq 1\}} \nu(dy) \right) + bt$$

with $b \in \mathbb{R}^d$, a σ-finite measure ν on $\mathbb{R}^d \setminus \{0\}$ satisfying $\int (|y|^2 \wedge 1)\, \nu(dy) < \infty$, and a Poisson point process (N_t) of intensity ν. Suppose first that $\operatorname{supp}(\nu) \subset \left\{ y \in \mathbb{R}^d : |y| \geq \varepsilon \right\}$ for some $\varepsilon > 0$. Then the components X^k are finite variation processes, and hence

$$[X^k, X^l]_t \quad = \quad \sum_{s \leq t} \Delta X_s^k \Delta X_s^l \quad = \quad \int y^k y^l \, N_t(dy). \tag{14.3.6}$$

In general, (14.3.6) still holds true. Indeed, if $X^{(\varepsilon)}$ is the corresponding Lévy process with intensity measure $\nu^{(\varepsilon)}(dy) = I_{\{|y| \geq \varepsilon\}}\, \nu(dy)$ then $||X^{(\varepsilon),k} - X^k||_{M^2([0,a])} \to 0$ as $\varepsilon \downarrow 0$ for any $a \in \mathbb{R}_+$ and $k \in \{1, \ldots, d\}$, and hence by Corollary 14.14,

$$\left[X^k, X^l\right]_t \quad = \quad \text{ucp-}\lim_{\varepsilon \downarrow 0} \left[X^{(\varepsilon),k}, X^{(\varepsilon),l}\right]_t \quad = \quad \sum_{s \leq t} \Delta X_s^k \Delta X_s^l.$$

On the other hand, we know that if X is square integrable then $M_t = X_t - it\nabla\psi(0)$ and $M_t^k M_t^l - t\frac{\partial^2 \psi}{\partial p_k \partial p_l}(0)$ are martingales, and hence

$$\langle X^k, X^l \rangle_t \quad = \quad \langle M^k, M^l \rangle_t \quad = \quad t \cdot \frac{\partial^2 \psi}{\partial p_k \partial p_l}(0).$$

3) **Covariations of Brownian motion and Lévy jump processes**: For B and X as above we have

$$\langle B^k, X^l \rangle \quad = \quad [B^k, X^l] \quad = \quad 0 \qquad \text{almost surely for any } k \text{ and } l. \tag{14.3.7}$$

Indeed, (14.3.7) holds true if X^l has finite variation paths. The general case then follows once more by approximating X^l by finite variation processes. Note that **independence of B and X has not been assumed**! We will see in Section 3.1 that (14.3.7) implies that a Brownian motion and a Lévy process without diffusion term defined on the same probability space are always independent.

Covariation of stochastic integrals

We now compute the covariation of stochastic integrals. This is not only crucial for many computations, but it also yields an alternative characterization of stochastic integrals w.r.t. local martingales, cf. Corollary 14.19 below.

Theorem 14.18. *Suppose that X and Y are $(\mathcal{F}_t^{\mathcal{P}})$ semimartingales, and H is $(\mathcal{F}_t^{\mathcal{P}})$ adapted and càdlàg. Then*

$$[\int H_- \, dX, Y] \quad = \quad \int H_- \, d[X,Y] \qquad \textit{almost surely.} \tag{14.3.8}$$

Proof. 1. We first note that (14.3.8) holds if X or Y has finite variation paths. If, for example, $X \in \mathrm{FV}$ then also $\int H_- \, dX \in \mathrm{FV}$, and hence

$$[\int H_- \, dX, Y] \;=\; \sum \Delta(H_{-\bullet} X)\Delta Y \;=\; \sum H_- \Delta X \Delta Y \;=\; \int H_- \, d[X,Y] \ .$$

The same holds if $Y \in \mathrm{FV}$.

2. Now we show that (14.3.8) holds if X and Y are bounded martingales, and H is bounded. For this purpose, we fix a partition π, and we approximate H_- by the elementary process $H_-^\pi = \sum_{s\in\pi} H_s \cdot I_{(s,s']}$. Let

$$I_t^\pi \quad = \quad \int_{(0,t]} H_-^\pi \, dX \quad = \quad \sum_{s\in\pi} H_s (X_{s'\wedge t} - X_s).$$

We can easily verify that

$$[I^\pi, Y] \quad = \quad \int H_-^\pi \, d[X,Y] \qquad \text{almost surely.} \tag{14.3.9}$$

Indeed, if $(\widetilde{\pi}_n)$ is a sequence of partitions such that $\pi \subset \widetilde{\pi}_n$ for any n and $\mathrm{mesh}(\widetilde{\pi}_n) \to 0$ then

$$\sum_{\substack{r\in\widetilde{\pi}_n\\ r<t}} (I^\pi_{r'\wedge t} - I^\pi_r)(Y_{r'\wedge t} - Y_r) \quad = \quad \sum_{s\in\pi} H_s \sum_{\substack{r\in\widetilde{\pi}_n\\ s\le r<s'\wedge t}} (X_{r'\wedge t} - X_r)(Y_{r'\wedge t} - Y_r).$$

Since the outer sum has only finitely many non-zero summands, the right hand side converges as $n \to \infty$ to

$$\sum_{s\in\pi} H_s\big([X,Y]_{s'\wedge t} - [X,Y]_s\big) \quad = \quad \int_{(0,t]} H^\pi_- \, d[X,Y],$$

in the ucp sense, and hence (14.3.9) holds.
Having verified (14.3.9) for any fixed partition π, we choose again a sequence (π_n) of partitions with $\mathrm{mesh}(\pi_n) \to 0$. Then

$$\int H_- \, dX \quad = \quad \lim_{n\to\infty} I^{\pi_n} \quad \text{in } M^2([0,a]) \text{ for any } a \in (0,\infty),$$

and hence, by Corollary 14.14 and (14.3.9),

$$\Big[\int H_- \, dX, Y\Big] \quad = \quad \mathrm{ucp}\text{-}\lim_{n\to\infty} [I^{\pi_n}, Y] \quad = \quad \int H_- \, d[X,Y].$$

3. Now suppose that X and Y are strict local martingales. If T is a stopping time such that X^T and Y^T are bounded martingales, and $HI_{[0,T)}$ is bounded as well, then by Step 2, Theorem 14.11 and Corollary 14.12,

$$\begin{aligned}\Big[\int H_- \, dX, Y\Big]^T \quad &= \quad \Big[\Big(\int H_- \, dX\Big)^T, Y^T\Big] \quad = \quad \Big[\int (H_- \, I_{[0,T)}) \, dX^T, Y^T\Big] \\ &= \quad \int H_- \, I_{[0,T)} \, d[X^T, Y^T] \quad = \quad \Big(\int H_- \, d[X,Y]\Big)^T.\end{aligned}$$

Since this holds for all localizing stopping times as above, (14.3.9) is satisfied as well.

4. Finally, suppose that X and Y are arbitrary semimartingales. Then $X = M + A$ and $Y = N + B$ with $M, N \in M_{\mathrm{loc}}$ and $A, B \in \mathrm{FV}$. The assertion (14.3.8) now follows by Steps 1 and 3 and by the bilinearity of stochastic integral and covariation. $\square$

Perhaps the most remarkable consequences of Theorem 14.18 is:

Corollary 14.19 (Kunita-Watanabe characterization of stochastic integrals).
Let $M \in M_{loc}$ and $G = H_$with H $(\mathcal{F}_t^{\mathcal{P}})$ adapted and càdlàg. Then $G_\bullet M$ is the unique element in M_{loc} satisfying

(i) $(G_\bullet M)_0 \quad = \quad 0 \quad ,$ *and*

(ii) $[G_\bullet M, N] \quad = \quad G_\bullet[M, N]$ *for any* $N \in M_{loc}$.

Proof. By Theorem 14.18, $G_\bullet M$ satisfies (i) and (ii). It remains to prove uniqueness. Let $L \in M_{\text{loc}}$ such that $L_0 = 0$ and

$$[L, N] \quad = \quad G_\bullet[M, N] \quad \text{for any } N \in M_{\text{loc}}.$$

Then $[L - G_\bullet M, N] = 0$ for any $N \in M_{\text{loc}}$. Choosing $N = L - G_\bullet M$, we conclude that $[L - G_\bullet M] = 0$. Hence $L - G_\bullet M$ is almost surely constant, i.e.,

$$L - G_\bullet M \quad \equiv \quad L_0 - (G_\bullet M)_0 \quad = \quad 0.$$

□

Remark. Localization shows that it is sufficient to verify Condition (ii) in the Kunita-Watanabe characterization for bounded martingales N.

The corollary tells us that in order to identify stochastic integrals w.r.t. local martingales it is enough to "test" with other (local) martingales via the covariation. This fact can be used to give an **alternative definition of stochastic integrals** that applies to general predictable integrands. Recall that a stochastic process $(G_t)_{t\geq 0}$ is called $(\mathcal{F}_{\mathbf{t}}^{\mathcal{P}})$ **predictable** iff the function $(\omega, t) \to G_t(\omega)$ is measurable w.r.t. the σ-algebra $\mathcal{P}$ on $\Omega \times [0, \infty)$ generated by all $(\mathcal{F}_t^{\mathcal{P}})$ adapted left-continuous processes.

Definition (Stochastic integrals with general predictable integrands).
Let $M \in M_{loc}$, and suppose that G is an $(\mathcal{F}_t^{\mathcal{P}})$ predictable process satisfying

$$\int_0^t G_s^2 \, d[M]_s \; < \; \infty \qquad \textit{almost surely for any } t \geq 0.$$

If there exists a local martingale $G_\bullet M \in M_{loc}$ such that conditions (i) and (ii) in Corollary 14.19 hold, then $G_\bullet M$ is called the ***stochastic integral of G w.r.t. M****.*

Many properties of stochastic integrals can be deduced directly from this definition, see e.g. Theorem 14.21 below.

The Itô isometry for stochastic integrals w.r.t. martingales

Of course, Theorem 14.18 can also be used to compute the covariation of two stochastic integrals. In particular, if M is a semimartingale and $G = H_-$ with H càdlàg and adapted then

$$[G_\bullet M, G_\bullet M] \quad = \quad G_\bullet[M, G_\bullet M] \quad = \quad G^2_\bullet[M].$$

Corollary 14.20 (Itô isometry for martingales). *Suppose that $M \in M_{loc}$. Then also $\left(\int G\, dM\right)^2 - \int G^2\, d[M] \in M_{loc}$, and*

$$\Big|\Big|\int G\, dM\Big|\Big|^2_{M^2([0,a])} \;=\; E\Big[\Big(\int_0^a G\, dM\Big)^2\Big] \;=\; E\Big[\int_0^a G^2\, d[M]\Big] \qquad \forall\, a \geq 0, \quad a.s\,.$$

Proof. If $M \in M_{\text{loc}}$ then $G_\bullet M \in M_{\text{loc}}$, and hence $(G_\bullet M)^2 - [G_\bullet M] \in M_{\text{loc}}$. Moreover, by Theorem 14.13,

$$||G_\bullet M||^2_{M^2([0,a])} \quad = \quad E\big[[G_\bullet M]_a\big] \quad = \quad E\big[(G^2_\bullet[M])_a\big].$$

□

The Itô isometry for martingales states that the $M^2([0,a])$ norm of the stochastic integral $\int G\, dM$ coincides with the $L^2\big(\Omega \times (0,a], P_{[M]}\big)$ norm of the integrand $(\omega, t) \mapsto G_t(\omega)$, where $P_{[M]}$ is the measure on $\Omega \times \mathbb{R}_+$ given by

$$P_{[M]}(d\omega\; dt) \quad = \quad P(d\omega)\; [M](\omega)(dt).$$

This can be used to prove the existence of the stochastic integral for general predictable integrands $G \in L^2(P_{[M]})$, cf. Section 2.5 below.

14.4 Itô calculus for semimartingales

We are now ready to prove the two most important rules of Itô calculus for semimartingales: The so-called "Associative Law" which tells us how to integrate w.r.t. processes that are stochastic integrals themselves, and the change of variables formula.

Integration w.r.t. stochastic integrals

Suppose that X and Y are semimartingales satisfying $dY = \widetilde{G}\, dX$ for some predictable integrand $\widetilde{G}$, i.e., $Y - Y_0 = \int \widetilde{G}\, dX$. We would like to show that we are allowed to multiply the differential equation formally by another predictable process G, i.e., we would like to prove that $\int G\, dY = \int G\widetilde{G}\, dX$:

$$dY \quad = \quad \widetilde{G}\, dX \qquad \Longrightarrow \qquad G\, dY \quad = \quad G\widetilde{G}\, dX$$

The covariation characterization of stochastic integrals w.r.t. local martingales can be used to prove this rule in a simple way.

Theorem 14.21 ("Associative Law"). *Let $X \in \mathcal{S}$. Then*

$$G_\bullet(\widetilde{G}_\bullet X) \quad = \quad (G\widetilde{G})_\bullet X \tag{14.4.1}$$

holds for any processes $G = H_-$ and $\widetilde{G} = \widetilde{H}_-$ with H and $\widetilde{H}$ càdlàg and adapted.

Remark. The assertion extends with a similar proof to more general predictable integrands.

Proof. We already know that (14.4.1) holds for $X \in \mathrm{FV}$. Therefore, and by bilinearity of the stochastic integral, we may assume $X \in M_{\mathrm{loc}}$. By the Kunita-Watanabe characterization it then suffices to "test" the identity (14.4.1) with local martingales. For $N \in M_{\mathrm{loc}}$, Corollary 14.19 and the associative law for FV processes imply

$$\begin{aligned}
[G_\bullet(\widetilde{G}_\bullet X), N] \quad &= \quad G_\bullet[\widetilde{G}_\bullet X, N] \quad = \quad G_\bullet(\widetilde{G}_\bullet[X, N]) \\
&= \quad (G\widetilde{G})_\bullet[X, N] \quad = \quad [(G\widetilde{G})_\bullet X, N].
\end{aligned}$$

Thus (14.4.1) holds by Corollary 14.19. □

Itô's formula

We are now going to prove a change of variables formula for discontinuous semimartingales. To get an idea how the formula looks like we first briefly consider a semimartingale $X \in \mathcal{S}$ with a finite number of jumps in finite time. Suppose that $0 < T_1 < T_2 < \dots$ are the jump times, and let $T_0 = 0$. Then on each of the intervals $[T_{k-1}, T_k)$, X is continuous. Therefore, by a similar argument as in the proof of Itô's formula for continuous paths (cf. [14, Thm.6.4]), we could expect that

$$\begin{aligned} F(X_t) - F(X_0) \quad &= \quad \sum_k \big(F(X_{t\wedge T_k}) - F(X_{t\wedge T_{k-1}})\big) \\ = \sum_{T_{k-1}<t} \Big(\int\limits_{T_{k-1}}^{t\wedge T_k -} F'(X_{s-})\, dX_s + \frac{1}{2} \int\limits_{T_{k-1}}^{t\wedge T_k -} F''(X_{s-})\, d[X]_s \Big) &+ \sum_{T_k \le t} \big(F(X_{T_k}) - F(X_{T_k -})\big) \\ = \int_0^t F'(X_{s-})\, dX_s^c + \frac{1}{2}\int_0^t F''(X_{s-})\, d[X]_s^c &+ \sum_{s\le t} \big(F(X_s) - F(X_{s-})\big) \end{aligned} \tag{14.4.2}$$

where $X_t^c = X_t - \sum_{s\le t} \Delta X_s$ denotes the continuous part of X. However, this formula does not carry over to the case when the jumps accumulate and the paths are not of finite variation, since then the series may diverge and the continuous part X^c does not exist in general. This problem can be overcome by rewriting (14.4.2) in the equivalent form

$$\begin{aligned} F(X_t) - F(X_0) \quad = \quad & \int_0^t F'(X_{s-})\, dX_s + \frac{1}{2}\int_0^t F''(X_{s-})\, d[X]_s^c \\ & + \sum_{s\le t} \big(F(X_s) - F(X_{s-}) - F'(X_{s-})\,\Delta X_s\big), \end{aligned} \tag{14.4.3}$$

which carries over to general semimartingales.

Theorem 14.22 (Itô's formula for semimartingales). *Suppose that* $X_t = (X_t^1, \dots, X_t^d)$ *with semimartingales* $X^1, \dots, X^d \in \mathcal{S}$. *Then for every function* $F \in C^2(\mathbb{R}^d)$,

$$\begin{aligned} F(X_t) - F(X_0) = & \sum_{i=1}^d \int_{(0,t]} \frac{\partial F}{\partial x^i}(X_{s-})\, dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_{(0,t]} \frac{\partial^2 F}{\partial x^i \partial x^j}(X_{s-})\, d[X^i, X^j]_s^c \\ & + \sum_{s \in (0,t]} \Big(F(X_s) - F(X_{s-}) - \sum_{i=1}^d \frac{\partial F}{\partial x^i}(X_{s-}) \Delta X_s^i \Big) \qquad (14.4.4) \end{aligned}$$

for any $t \geq 0$, *almost surely.*

Remark. The existence of the quadratic variations $[X^i]_t$ implies the almost sure absolute convergence of the series over $s \in (0,t]$ on the right hand side of (14.4.4). Indeed, a Taylor expansion up to order two shows that

$$\begin{aligned} \sum_{s \le t} \Big| F(X_s) - F(X_{s-}) - \sum_{i=1}^d \frac{\partial F}{\partial x^i}(X_{s-}) \Delta X_s^i \Big| \quad & \le \quad C_t \cdot \sum_{s \le t} \sum_i |\Delta X_s^i|^2 \\ & \le C_t \cdot \sum_i [X^i]_t < \infty, \end{aligned}$$

where $C_t = C_t(\omega)$ is an almost surely finite random constant depending only on the maximum of the norm of the second derivative of F on the convex hull of $\{X_s : s \in [0,t]\}$.

It is possible to prove this general version of Itô's formula by a Riemann sum approximation, cf. [36]. Here, following [38], we instead derive the "chain rule" once more from the "product rule":

Proof. To keep the argument transparent, we restrict ourselves to the case $d = 1$. The generalization to higher dimensions is straightforward. We now proceed in three steps:

1. As in the finite variation case (Theorem 14.4), we first prove that the set $\mathcal{A}$ consisting of all functions $F \in C^2(\mathbb{R})$ satisfying (14.4.3) is an algebra, i.e.,

$$F, G \in \mathcal{A} \quad \Longrightarrow \quad FG \in \mathcal{A}.$$

This is a consequence of the integration by parts formula

$$F(X_t)G(X_t) - F(X_0)G(X_0) = \int_0^t F(X_-)\, dG(X) + \int_0^t G(X_-)\, dF(X) \\ + \big[F(X), G(X)\big]^c + \sum_{(0,t]} \Delta F(X) \Delta G(X), \quad (14.4.5)$$

the associative law, which implies

$$\int F(X_-)\, dG(X) \quad = \quad \int F(X_-) G'(X_-)\, dX + \frac{1}{2} \int F(X_-) G''(X_-)\, d[X]^c \\ + \sum F(X_-)\, (\Delta G(X) - G'(X_-)\Delta X), \qquad (14.4.6)$$

the corresponding identity with F and G interchanged, and the formula

$$\begin{aligned} [F(X), G(X)]^c \quad &= \quad \Big[\int F'(X_-)\, dX, \int G'(X_-)\, dX\Big]^c \qquad (14.4.7) \\ &= \quad \Big(\int F'(X_-)G'(X_-)\, d[X]\Big)^c \quad = \quad \int (F'G')(X_-)\, d[X]^c \end{aligned}$$

for the continuous part of the covariation. Both (14.4.6) and (14.4.7) follow from (14.4.4) and the corresponding identity for G. It is straightforward to verify that (14.4.5), (14.4.6) and (14.4.7) imply the change of variable formula (14.4.3) for FG, i.e., $FG \in \mathcal{A}$. Therefore, by induction, the formula (14.4.3) holds for all polynomials F.

2. In the second step, we prove the formula for arbitrary $F \in C^2$ assuming $X = M + A$ with a bounded martingale M and a bounded process $A \in$ FV. In this case, X is uniformly bounded by a finite constant C. Therefore, there exists a sequence (p_n) of polynomials such that $p_n \to F$, $p_n' \to F'$ and $p_n'' \to F''$ uniformly on $[-C, C]$. For $t \geq 0$, we obtain

$$\begin{aligned} F(X_t) - F(X_0) \quad &= \quad \lim_{n\to\infty} \big(p_n(X_t) - p_n(X_0)\big) \\ = \quad &\lim_{n\to\infty} \Big(\int_0^t p_n'(X_{s-})\, dX_s + \frac{1}{2} \int_0^t p_n''(X_{s-})\, d[X]_s^c + \sum_{s\leq t} \int_{X_{s-}}^{X_s} \int_{X_{s-}}^{y} p_n''(z)\, dz\, dy \Big) \\ = \quad &\int_0^t F'(X_{s-})\, dX_s + \frac{1}{2} \int_0^t F''(X_{s-})\, d[X]_s^c + \sum_{s\leq t} \int_{X_{s-}}^{X_s} \int_{X_{s-}}^{y} F''(z)\, dz\, dy \end{aligned}$$

w.r.t. convergence in probability. Here we have used an expression of the jump terms in (14.4.3) by a Taylor expansion. The convergence in probability holds since $X = M+A$,

$$\begin{aligned} & E\Big[\Big|\int_0^t p_n'(X_{s-})\,dM_s - \int_0^t F'(X_{s-})\,dM_s\Big|^2\Big] \\ = \quad & E\Big[\int_0^t (p_n'-F')(X_{s-})^2\,d[M]_s\Big] \quad \leq \quad \sup_{[-C,C]} |p_n'-F'|^2\cdot E\big[[M]_t\big] \end{aligned}$$

by Itô's isometry, and

$$\Big|\sum_{s\leq t}\int_{X_{s-}}^{X_s}\int_{X_{s-}}^{y}(p_n''-F'')(z)\,dz\,dy\Big| \quad \leq \quad \frac{1}{2}\sup_{[-C,C]}|p_n''-F''|\,\sum_{s\leq t}(\Delta X_s)^2.$$

3. Finally, the change of variables formula for general semimartingales $X = M + A$ with $M \in M_{\text{loc}}$ and $A \in$ FV follows by localization. We can find an increasing sequence of stopping times (T_n) such that $\sup T_n = \infty$ a.s., M^{T_n} is a bounded martingale, and the process A^{T_n-} defined by

$$A_t^{T_n-} \quad := \quad \begin{cases} A_t & \text{for } t < T_n \\ A_{T_n-} & \text{for } t \geq T_n \end{cases}$$

is a bounded process in FV for any n. Itô's formula then holds for $X^n := M^{T_n} + A^{T_n-}$ for every n. Since $X^n = X$ on $[0,T_n)$ and $T_n \nearrow \infty$ a.s., this implies Itô's formula for X. □

Note that the second term on the right hand side of Itô's formula (14.4.4) is a continuous finite variation process and the third term is a pure jump finite variation process. Moreover, semimartingale decompositions of X^i, $1 \leq i \leq d$, yield corresponding decompositions of the stochastic integrals on the right hand side of (14.4.4). Therefore, Itô's formula can be applied to derive an explicit semimartingale decomposition of $F(X_t^1,\dots,X_t^d)$ for any C^2 function F. This will now be carried out in concrete examples.

Application to Lévy processes

We first apply Itô's formula to a one-dimensional Lévy process

$$X_t \quad = \quad x + \sigma B_t + bt + \int y\,\widetilde{N}_t(dy) \tag{14.4.8}$$

with $x, \sigma, b \in \mathbb{R}$, a Brownian motion (B_t), and a compensated Poisson point process $\widetilde{N}_t = N_t - t\nu$ with intensity measure ν. We assume that $\int (|y|^2 \wedge |y|)\, \nu(dy) < \infty$. The only restriction to the general case is the assumed integrability of $|y|$ at ∞, which ensures in particular that (X_t) is integrable. The process (X_t) is a semimartingale w.r.t. the filtration $(\mathcal{F}_t^{B,N})$ generated by the Brownian motion and the Poisson point process.

We now apply Itô's formula to $F(X_t)$ where $F \in C^2(\mathbb{R})$. Setting $C_t = \int y\, \widetilde{N}_t(dy)$ we first note that almost surely,

$$[X]_t \quad = \quad \sigma^2 [B]_t + 2\sigma [B,C]_t + [C]_t \quad = \quad \sigma^2 t + \sum_{s \le t} (\Delta X_s)^2.$$

Therefore, by (14.4.9),

$$\begin{aligned}
& F(X_t) - F(X_0) \\
&= \int_0^t F'(X_-)\, dX + \frac{1}{2} \int_0^t F''(X_-)\, d[X]^c + \sum_{s \le t} \big(F(X) - F(X_-) - F'(X_-)\Delta X \big) \\
&= \int_0^t (\sigma F')(X_{s-})\, dB_s + \int_0^t (bF' + \frac{1}{2}\sigma^2 F'')(X_s)\, ds + \int\limits_{(0,t]\times\mathbb{R}} F'(X_{s-})\, y\, \widetilde{N}(ds\, dy) \\
&\quad + \int\limits_{(0,t]\times\mathbb{R}} \big(F(X_{s-} + y) - F(X_{s-}) - F'(X_{s-}) y \big)\, N(ds\, dy), \qquad (14.4.9)
\end{aligned}$$

where $N(ds\, dy)$ is the Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ corresponding to the Poisson point process, and $\widetilde{N}(ds\, dy) = N(ds\, dy) - ds\, \nu(dy)$. Here, we have used a rule for evaluating a stochastic integral w.r.t. the process $C_t = \int y\, \widetilde{N}_t(dy)$ which is intuitively clear and can be verified by approximating the integrand by elementary processes. Note also that in the second integral on the right hand side we could replace X_{s-} by X_s since almost surely, $\Delta X_s = 0$ for almost all s.

To obtain a semimartingale decomposition from (14.4.9), we note that the stochastic integrals w.r.t. (B_t) and w.r.t. $(\widetilde{N}_t)$ are local martingales. By splitting the last integral on the right hand side of (14.4.9) into an integral w.r.t. $\widetilde{N}(ds\, dy)$ (i.e., a local martingale) and an integral w.r.t. the compensator $ds\, \nu(dy)$, we have proven:

Corollary 14.23 (**Martingale problem for Lévy processes**). *For any $F \in C^2(\mathbb{R})$, the process*

$$M_t^{[F]} = F(X_t) - F(X_0) - \int_0^t (\mathcal{L}F)(X_s)\, ds,$$

$$(\mathcal{L}F)(x) = \frac{1}{2}(\sigma F'')(x) + (bF')(x) + \int \big(F(x+y) - F(x) - F'(x)y\big)\, \nu(dy),$$

is a local martingale vanishing at 0. *For $F \in \mathcal{C}_b^2(\mathbb{R})$, $M^{[F]}$ is a martingale, and*

$$(\mathcal{L}F)(x) = \lim_{t\downarrow 0} \frac{1}{t} E\big[F(X_t) - F(X_0)\big].$$

Proof. $M^{[F]}$ is a local martingale by the considerations above and since $X_s(\omega) = X_{s-}(\omega)$ for almost all (s,ω). For $F \in \mathcal{C}_b^2$, $\mathcal{L}F$ is bounded since $\big|F(x+y) - F(x) - F'(x)y\big| = \mathcal{O}(|y| \wedge |y|^2)$. Hence $M^{[F]}$ is a martingale in this case, and

$$\frac{1}{t} E\big[F(X_t) - F(X_0)\big] = E\Big[\frac{1}{t}\int_0^t (\mathcal{L}F)(X_s)\, ds\Big] \quad \to \quad (\mathcal{L}F)(x)$$

as $t \downarrow 0$ by right continuity of $(\mathcal{L}F)(X_s)$. □

The corollary shows that $\mathcal{L}$ is the infinitesimal generator of the Lévy process. The martingale problem can be used to extend results on the connection between Brownian motion and the Laplace operator to general Lévy processes and their generators. For example, exit distributions are related to boundary value problems (or rather complement value problems as $\mathcal{L}$ is not a local operator), there is a potential theory for generators of Lévy processes, the Feynman-Kac formula and its applications carry over, and so on.

Example (**Fractional powers of the Laplacian**). By Fourier transformation one verifies that the generator of a symmetric α-stable process with characteristic exponent $|p|^\alpha$ is $\mathcal{L} = -(-\Delta)^{\alpha/2}$. The behaviour of symmetric α-stable processes is therefore closely linked to the potential theory of these well-studied pseudo-differential operators.

Exercise (**Exit distributions for compound Poisson processes**). Let $(X_t)_{t\geq 0}$ be a compound Poisson process with $X_0 = 0$ and jump intensity measure $\nu = N(m, 1),\ m > 0$.

i) Determine $\lambda \in \mathbb{R}$ such that $\exp(\lambda X_t)$ is a local martingale.

ii) Prove that for $a < 0$,

$$P[T_a < \infty] \;=\; \lim_{b\to\infty} P[T_a < T_b] \;\leq\; \exp(ma/2).$$

Why is it not as easy as for Brownian motion to compute $P[T_a < T_b]$ exactly?

Burkholder's inequality

As another application of Itô's formula, we prove an important inequality for càdlàg local martingales that is used frequently to derive L^p estimates for semimartingales. For real-valued càdlàg functions $x = (x_t)_{t\geq 0}$ we set

$$x_t^\star \quad := \quad \sup_{s<t} |x_s| \qquad \text{for } \; t > 0, \qquad \text{and} \qquad x_0^\star \;:=\; |x_0|.$$

Theorem 14.24 (Burkholder's inequality). *Let $p \in [2,\infty)$. Then the estimate*

$$E[(M_T^\star)^p]^{1/p} \quad \leq \quad \gamma_p \, E[[M]_T^{p/2}]^{1/p} \tag{14.4.10}$$

holds for any strict local martingale $M \in \mathcal{M}_{loc}$ such that $M_0 = 0$, and for any stopping time $T : \Omega \to [0,\infty]$, where

$$\gamma_p \quad = \quad \left(1 + \tfrac{1}{p-1}\right)^{(p-1)/2} p/\sqrt{2} \quad \leq \quad \sqrt{e/2}\, p.$$

Remark. The estimate does not depend on the underlying filtered probability space, the local martingale M, and the stopping time T. However, the constant γ_p goes to ∞ as $p \to \infty$.

Notice that for $p = 2$, Equation (14.4.10) holds with $\gamma_p = 2$ by Itô's isometry and Doob's L^2 maximal inequality. Burkholder's inequality can thus be used to generalize arguments based on Itô's isometry from an L^2 to an L^p setting.

Proof. 1) We first assume that $T = \infty$ and M is a bounded càdlàg martingale. Then, by the Martingale Convergence Theorem, $M_\infty = \lim_{t\to\infty} M_t$ exists almost surely. Since the function $f(x) = |x|^p$ is C^2 for $p \geq 2$ with $\phi''(x) = p(p-1)|x|^{p-2}$, Itô's formula implies

$$\begin{aligned} |M_\infty|^p \quad &= \quad \int_0^\infty \phi'(M_{s-})\, dM_s + \frac{1}{2}\int_0^\infty \phi''(M_s)\, d[M]^c_s \\ &\qquad + \sum_s \left(\phi(M_s) - \phi(M_{s-}) - \phi'(M_{s-})\Delta M_s, \right), \end{aligned} \tag{14.4.11}$$

where the first term is a martingale since $\phi' \circ M$ is bounded, in the second term

$$\phi''(M_s) \quad \leq \quad p(p-1)(M^\star_\infty)^{p-2},$$

and the summand in the third term can be estimated by

$$\begin{aligned} \phi(M_s) - \phi(M_{s-}) - \phi'(M_{s-})\Delta M_s \quad &\leq \quad \frac{1}{2}\sup(\phi'' \circ M)(\Delta M_s)^2 \\ &\leq \quad \frac{1}{2}p(p-1)(M^\star_\infty)^{p-2}(\Delta M_s)^2. \end{aligned}$$

Hence by taking expectation values on both sides of (14.4.11), we obtain for q satisfying $\frac{1}{p} + \frac{1}{q} = 1$:

$$\begin{aligned} E[(M^\star_\infty)^p] \quad &\leq \quad q^p\, E[|M_\infty|^p] \\ &\leq \quad q^p\, \frac{p(p-1)}{2}\, E\Big[(M^\star_\infty)^{p-2}\Big([M]^c_\infty + \sum(\Delta M)^2\Big)\Big] \\ &\leq \quad q^p\, \frac{p(p-1)}{2}\, E[(M^\star_\infty)^p]^{\frac{p-2}{p}}\, E[[M]^{\frac{p}{2}}_\infty]^{\frac{2}{p}} \end{aligned}$$

by Doob's inequality, Hölder's inequality, and since $[M]^c_\infty + \sum(\Delta M)^2 = [M]_\infty$. The inequality (14.4.10) now follows by noting that $q^p p(p-1) = q^{p-1}p^2$.

2) For $T = \infty$ and a strict local martingale $M \in M_{\mathrm{loc}}$, there exists an increasing sequence (T_n) of stopping times such that M^{T_n} is a bounded martingale for each n. Applying Burkholder's inequality to M^{T_n} yields

$$E[(M^\star_{T_n})^p] \quad = \quad E[(M^{T_n,\star}_\infty)^p] \quad \leq \quad \gamma^p_p\, E[[M^{T_n}]^{p/2}_\infty] \quad = \quad \gamma^p_p\, E[[M]^{p/2}_{T_n}].$$

Burkholder's inequality for M now follows as $n \to \infty$.

3) Finally, the inequality for an arbitrary stopping time T can be derived from that for $T = \infty$ by considering the stopped process M^T. □

For $p \geq 4$, the converse estimate in (12.0.1) can be derived in a similar way:

Exercise. Prove that for a given $p \in [4, \infty)$, there exists a global constant $c_p \in (1, \infty)$ such that the inequalities

$$c_p^{-1} E\left[[M]_\infty^{p/2}\right] \leq E\left[(M_\infty^*)^p\right] \leq c_p E\left[[M]_\infty^{p/2}\right]$$

with $M_t^* = \sup_{s<t} |M_s|$ hold for any continuous local martingale $(M_t)_{t\in[0,\infty)}$.

The following concentration inequality for martingales is often more powerful than Burkholder's inequality:

Exercise. Let M be a continuous local martingale satisfying $M_0 = 0$. Show that

$$P\Big[\sup_{s\leq t} M_s \geq x \,;\, [M]_t \leq c\Big] \quad \leq \quad \exp\Big(-\frac{x^2}{2c}\Big)$$

for any $c, t, x \in [0, \infty)$.

14.5 Stochastic exponentials and change of measure

A change of the underlying probability measure by an exponential martingale can also be carried out for jump processes. In this section, we first introduce exponentials of general semimartingales. After considering absolutely continuous measure transformations for Poisson point processes, we apply the results to Lévy processes, and we prove a general change of measure result for possibly discontinuous semimartingales. Finally, we provide a counterpart to Lévy's characterization of Brownian motion for general Lévy processes.

Exponentials of semimartingales

If X is a continuous semimartingale then by Itô's formula,

$$\mathcal{E}_t^X \quad = \quad \exp\Big(X_t - \frac{1}{2}[X]_t\Big)$$

is the unique solution of the exponential equation

$$d\mathcal{E}^X \quad = \quad \mathcal{E}^X\, dX, \qquad \mathcal{E}_0^X \quad = \quad 1.$$

In particular, $\mathcal{E}^X$ is a local martingale if X is a local martingale. Moreover, if

$$h_n(t,x) \quad = \quad \frac{\partial^n}{\partial\alpha^n}\exp(\alpha x - \alpha^2 t/2)\Big|_{\alpha=0} \tag{14.5.1}$$

denotes the Hermite polynomial of order n and $X_0 = 0$ then

$$H^n_t \quad = \quad h_n\big([X]_t, X_t\big) \tag{14.5.2}$$

solves the SDE

$$dH^n \quad = \quad n\, H^{n-1}\, dX, \qquad H^n_0 = 0,$$

for any $n \in \mathbb{N}$, cf. Section 6.4 in [14]. In particular, H^n is an iterated Itô integral:

$$H^n_t \quad = \quad n! \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} dX_{s_1} dX_{s_2} \cdots dX_{s_n}.$$

The formula for the stochastic exponential can be generalized to the discontinuous case:

Theorem 14.25 (**Doléans-Dade**). *Let $X \in \mathcal{S}$. Then the unique solution of the exponential equation*

$$Z_t \quad = \quad 1 + \int_0^t Z_{s-}\, dX_s, \qquad t \geq 0, \tag{14.5.3}$$

is given by

$$Z_t \quad = \quad \exp\Big(X_t - \frac{1}{2}[X]^c_t\Big) \prod_{s\in(0,t]} (1+\Delta X_s)\exp(-\Delta X_s). \tag{14.5.4}$$

Remarks. 1) In the finite variation case, (14.5.4) can be written as

$$Z_t \quad = \quad \exp\Big(X^c_t - \frac{1}{2}[X]^c_t\Big) \prod_{s\in(0,t]} (1+\Delta X_s).$$

In general, however, neither X^c nor $\prod(1+\Delta X)$ exist.

2) The analogues to the stochastic polynomials H^n in the discontinuous case do not have an equally simply expression as in (14.5.2) . This is not too surprising: Also for

continuous two-dimensional semimartingales (X_t, Y_t) there is no direct expression for the iterated integral $\int_0^t \int_0^s dX_r\, dY_s = \int_0^t (X_s - X_0)\, dY_s$ and for the Lévy area process

$$A_t \quad = \quad \int_0^t \int_0^s dX_r\, dY_s - \int_0^t \int_0^s dY_r\, dX_s$$

in terms of X,Y and their covariations. If X is a one-dimensional discontinuous semimartingale then X and X_- are different processes that have both to be taken into account when computing iterated integrals of X.

Proof of Theorem 14.25. The proof is partially similar to the one given above for $X \in$ FV, cf. Theorem 14.5. The key observation is that the product

$$P_t \quad = \quad \prod_{s \in (0,t]} (1 + \Delta X_s) \exp(-\Delta X_s)$$

exists and defines a finite variation pure jump process. This follows from the estimate

$$\sum_{\substack{0<s\leq t \\ |\Delta X_s| \leq 1/2}} |\log(1 + \Delta X_s) - \Delta X_s| \quad \leq \quad \text{const.} \cdot \sum_{s \leq t} |\Delta X_s|^2 \quad \leq \quad \text{const.} \cdot [X]_t$$

which implies that

$$S_t \quad = \quad \sum_{\substack{s\leq t \\ |\Delta X_s| \leq 1/2}} (\log(1 + \Delta X_s) - \Delta X_s), \qquad t \geq 0,$$

defines almost surely a finite variation pure jump process. Therefore, (P_t) is also a finite variation pure jump process.
Moreover, the process $G_t = \exp\left(X_t - \frac{1}{2}[X]_t^c\right)$ satisfies

$$G \quad = \quad 1 + \int G_-\, dX + \sum (\Delta G - G_-\, \Delta X) \tag{14.5.5}$$

by Itô's formula. For $Z = G\, P$ we obtain

$$\Delta Z \quad = \quad Z_- \left(e^{\Delta X} (1 + \Delta X) e^{-\Delta X} - 1 \right) \quad = \quad Z_-\, \Delta X,$$

and hence, by integration by parts and (14.5.5),

$$\begin{aligned} Z - 1 \quad &= \quad \int P_-\, dG + \int G_-\, dP + [G, P] \\ &= \quad \int P_- G_-\, dX + \sum (P_- \Delta G - P_- G_-\, \Delta X + G_-\, \Delta P + \Delta G\, \Delta P) \\ &= \quad \int Z_-\, dX + \sum (\Delta Z - Z_-\, \Delta X) \quad = \quad \int Z_-\, dX. \end{aligned}$$

This proves that Z solves the SDE (14.5.3). Uniqueness of the solution follows from a general uniqueness result for SDE with Lipschitz continuous coefficients, cf. Section 12.1. □

Example (Geometric Lévy processes). Consider a Lévy martingale $X_t = \int y\, \widetilde{N}_t(dy)$ where (N_t) is a Poisson point process on $\mathbb{R}$ with intensity measure ν satisfying $\int (|y| \wedge |y|^2)\, \nu(dy) < \infty$, and $\widetilde{N}_t = N_t - t\nu$. We derive an SDE for the semimartingale

$$Z_t \quad = \quad \exp(\sigma X_t + \mu t), \qquad t \geq 0,$$

where σ and μ are real constants. Since $[X]^c \equiv 0$, Itô's formula yields

$$\begin{aligned} Z_t - 1 \quad &= \sigma \int\limits_{(0,t]} Z_-\, dX + \mu \int\limits_{(0,t]} Z_-\, ds + \sum_{(0,t]} Z_- \Big(e^{\sigma \Delta X} - 1 - \sigma \Delta X \Big) \qquad (14.5.6) \\ &= \sigma \int\limits_{(0,t]\times\mathbb{R}} Z_{s-}\, y\, \widetilde{N}(ds\, dy) + \mu \int\limits_{(0,t]} Z_{s-}\, ds + \int_{(0,t]\times\mathbb{R}} Z_{s-} \Big(e^{\sigma y} - 1 - \sigma y \Big)\, N(ds\, dy). \end{aligned}$$

If $\int e^{2\sigma y}\, \nu(dy) < \infty$ then (14.5.6) leads to the semimartingale decomposition

$$dZ_t \quad = \quad Z_{t-}\, dM_t^\sigma + \alpha Z_{t-}\, dt, \qquad Z_0 = 1, \tag{14.5.7}$$

where

$$M_t^\sigma = \int \Big(e^{\sigma y} - 1 \Big)\, \widetilde{N}_t(dy)$$

is a square-integrable martingale, and

$$\alpha \;=\; \mu + \int (e^{\sigma y} - 1 - \sigma y)\, \nu(dy).$$

In particular, we see that although (Z_t) again solves an SDE driven by the compensated process $(\widetilde{N}_t)$, this SDE can not be written as an SDE driven by the Lévy process (X_t).

Change of measure for Poisson point processes

Let $(N_t)_{t\geq 0}$ be a Poisson point process on a σ-finite measure space $(S, \mathcal{S}, \nu)$ that is defined and adapted on a filtered probability space $(\Omega, \mathcal{A}, Q, (\mathcal{F}_t))$. Suppose that $(\omega, t, y) \mapsto$

$H_t(y)(\omega)$ is a predictable process in $\mathcal{L}^2_{\text{loc}}(Q \otimes \lambda \otimes \nu)$. Our goal is to change the underlying measure Q to a new measure P such that w.r.t. P, $(N_t)_{t\geq 0}$ is a point process with intensity of points in the infinitesimal time interval $[t, t+dt]$ given by

$$(1 + H_t(y))\, dt\, \nu(dy).$$

Note that in general, this intensity may depend on ω in a predictable way. Therefore, under the new probability measure P, the process (N_t) is not necessarily a **Poisson** point process. We define a local exponential martingale by

$$Z_t \quad := \quad \mathcal{E}^L_t \qquad \text{where} \qquad L_t \quad := \quad (H_\bullet \widetilde{N})_t. \tag{14.5.8}$$

Lemma 14.26. *Suppose that* $\inf H > -1$*, and let* $G := \log(1+H)$*. Then for* $t \geq 0$*,*

$$\mathcal{E}^L_t \quad = \quad \exp\Big(\int_{(0,t]\times S} G_s(y)\, \widetilde{N}(ds\, dy) - \int_{(0,t]\times S} (H_s(y) - G_s(y))\, ds\, \nu(dy) \Big).$$

Proof. The assumption $\inf H > -1$ implies $\inf \Delta L > -1$. Since, moreover, $[L]^c = 0$, we obtain

$$\begin{aligned} \mathcal{E}^L \quad &= \quad e^{L - [L]^c/2} \prod (1 + \Delta L) e^{-\Delta L} \\ &= \quad \exp\Big(L + \sum (\log(1+\Delta L) - \Delta L) \Big) \\ &= \quad \exp\Big(G_\bullet \widetilde{N} + \int (G - H)\, ds\, \nu(dy) \Big). \end{aligned}$$

Here we have used that

$$\sum (\log(1+\Delta L) - \Delta L) \quad = \quad \int \big(\log(1 + H_s(y)) - H_s(y) \big)\, N(ds\, dy)$$

holds, since $|\log(1 + H_s(y)) - H_s(y)| \leq \text{const.}\ |H_s(y)|^2$ is integrable on finite time intervals. □

The exponential $Z_t = \mathcal{E}^L_t$ is a strictly positive local martingale w.r.t. Q, and hence a supermartingale. As usual, we fix $t_0 \in \mathbb{R}_+$, and we assume:

Assumption. $(Z_t)_{t\leq t_0}$ is a martingale w.r.t. Q, i.e. $E_Q[Z_{t_0}] = 1$.

Then there is a probability measure P on $\mathcal{F}_{t_0}$ such that

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_t} \quad = \quad Z_t \qquad \text{for any } \ t \leq t_0.$$

In the deterministic case $H_t(y)(\omega) = h(y)$, we can prove that w.r.t. P, (N_t) is a Poisson point process with changed intensity measure

$$\mu(dy) \quad = \quad (1 + h(y))\,\nu(dy) :$$

Theorem 14.27 (Change of measure for Poisson point processes). *Let (N_t, Q) be a Poisson point process with intensity measure ν, and let $g := \log(1+h)$ where $h \in \mathcal{L}^2(\nu)$ satisfies $\inf h > -1$. Suppose that the exponential*

$$Z_t \quad = \quad \mathcal{E}_t^{\widetilde{N}(h)} \quad = \quad \exp\Big(\widetilde{N}_t(g) + t\int (g - h)\,d\nu\Big) \tag{14.5.9}$$

is a martingale w.r.t. Q, and assume that $P \ll Q$ on $\mathcal{F}_t$ with relative density $\frac{dP}{dQ}\Big|_{\mathcal{F}_t} = Z_t$ for any $t \geq 0$. Then the process (N_t, P) is a Poisson point process with intensity measure

$$d\mu \quad = \quad (1 + h)\,d\nu.$$

Proof. By the Lévy characterization for Poisson point processes (cf. the exercise below Lemma 11.1) it suffices to show that the process

$$M_t^{[f]} \quad = \quad \exp\big(iN_t(f) + t\psi(f)\big), \qquad \psi(f) \quad = \quad \int \big(1 - e^{if}\big)\,d\mu,$$

is a local martingale w.r.t. P for any elementary function $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$. Furthermore, by Lemma 9.8, $M^{[f]}$ is a local martingale w.r.t. P if and only if $M^{[f]}Z$ is a local martingale w.r.t. Q. The local martingale property for $(M^{[f]}Z, Q)$ can be verified by a computation based on Itô's formula. □

Remark (Extension to general measure transformations). The approach in Theorem 14.27 can be extended to the case where the function $h(y)$ is replaced by a general predictable process $H_t(y)(\omega)$. In that case, one verifies similarly that under a new measure P with local densities given by (14.5.8), the process

$$M_t^{[f]} \quad = \quad \exp\Big(iN_t(f) + \int (1 - e^{if(y)})(1 + H_t(y))\,dy\Big)$$

is a local martingale for any elementary function $f \in \mathcal{L}^1(\nu)$. This property can be used as a definition of a point process with predictable intensity $(1 + H_t(y))\,dt\,\nu(dy)$.

Change of measure for Lévy processes

Since Lévy processes can be constructed from Poisson point processes, a change of measure for Poisson point processes induces a corresponding transformation for Lévy processes. Suppose that ν is a σ-finite measure on $\mathbb{R}^d \setminus \{0\}$ such that

$$\int (|y| \wedge |y|^2)\, \nu(dy) \;<\; \infty, \qquad \text{and let}$$

$$\mu(dy) \quad = \quad (1 + h(y))\, \nu(dy).$$

Recall that if (N_t, Q) is a Poisson point process with intensity measure ν, then

$$X_t \quad = \quad \int y \; \widetilde{N}_t(dy), \qquad \widetilde{N}_t \;=\; N_t - t\nu,$$

is a Lévy martingale with Lévy measure ν w.r.t. Q.

Corollary 14.28 (Girsanov transformation for Lévy processes). *Suppose that $h \in \mathcal{L}^2(\nu)$ satisfies* $\inf h > -1$ *and* $\sup h < \infty$*. If $P \ll Q$ on $\mathcal{F}_t$ with relative density Z_t for any $t \geq 0$, where Z_t is given by* (14.5.9)*, then the process*

$$\overline{X}_t \quad = \quad \int y\; \overline{N}_t\,(dy), \qquad \overline{N}_t \;=\; N_t - t\mu,$$

is a Lévy martingale with Lévy measure μ w.r.t. P, and

$$X_t \;=\; \overline{X}_t \,+\, t \int y\, h(y)\, \nu(dy).$$

Notice that the effect of the measure transformation consists of both the addition of a drift and a change of the intensity measure of the Lévy martingale. This is different to the case of Brownian motion where only a drift is added.

Example (Change of measure for compound Poisson processes). Suppose that (X, Q) is a compound Poisson process with finite jump intensity measure ν, and let

$$N_t^h \quad = \quad \sum_{s \leq t} h(\Delta X_s)$$

with h as above. Then (X,P) is a compound Poisson process with jump intensity measure $d\mu = (1+h)\,d\nu$ provided

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_t} \quad = \quad \mathcal{E}_t^{\widetilde{N}(h)} \quad = \quad e^{-t\int h\,d\nu}\prod_{s\le t}(1+h(\Delta X_s)).$$

Lévy's characterization for Brownian motion has an extension to Lévy processes, too:

Theorem 14.29 (Lévy characterization of Lévy processes). *Let $a\in\mathbb{R}^{d\times d}$, $b\in\mathbb{R}$, and let ν be a σ-finite measure on $\mathbb{R}^d\setminus\{0\}$ satisfying $\int(|y|\wedge|y|^2)\,\nu(dy)<\infty$. If $X_t^1,\dots,X_t^d:\Omega\to\mathbb{R}$ are càdlàg stochastic processes such that*

(i) $M_t^k := X_t^k - b^k t$ *is a local $(\mathcal{F}_t)$ martingale for any $k\in\{1,\dots,d\}$,*

(ii) $[X^k,X^l]_t^c = a^{kl}\,t$ *for any $k,l\in\{1,\dots,d\}$, and*

(iii) $E\Big[\sum_{s\in(r,t]} I_B(\Delta X_s)\Big|\mathcal{F}_r\Big] = \nu(B)\cdot(t-r)$ *almost surely for any $0\le r\le t$ and for any $B\in\mathcal{B}(\mathbb{R}^d\setminus\{0\})$,*

then $X_t=(X_t^1,\dots,X_t^d)$ is a Lévy process with characteristic exponent

$$\psi(p) \quad = \quad \frac{1}{2}p\cdot ap - ip\cdot b + \int(1-e^{ip\cdot y}+ip\cdot y)\,\nu(dy). \tag{14.5.10}$$

For proving the theorem, we assume without proof that a local martingale is a semi-martingale even if it is not strict, and that the stochastic integral of a bounded adapted left-continuous integrand w.r.t. a local martingale is again a local martingale, cf. [36].

Proof of Theorem 14.29. We first remark that (iii) implies

$$E\Big[\sum_{s\in(r,t]} G_s\cdot f(\Delta X_s)\Big|\mathcal{F}_r\Big] \quad = \quad E\Big[\int_r^t\int G_s\cdot f(y)\,\nu(dy)\,ds\Big|\mathcal{F}_r\Big], \quad \text{a.s. for } 0\le r\le t \tag{14.5.11}$$

for any bounded left-continuous adapted process G, and for any measurable function $f:\mathbb{R}^d\setminus\{0\}\to\mathbb{C}$ satisfying $|f(y)|\le \text{const.}\cdot(|y|\wedge|y|^2)$. This can be verified by first considering elementary functions of type $f(y) = \sum c_i \, I_{B_i}(y)$ and $G_s(\omega) = \sum A_i(\omega) \, I_{(t_i,t_{i+1}]}(s)$ with $c_i \in \mathbb{R}$, $B_i \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$, $0 \le t_0 < t_1 < \cdots < t_n$, and A_i bounded and $\mathcal{F}_{t_i}$-measurable.

Now fix $p \in \mathbb{R}^d$, and consider the semimartingale

$$Z_t \quad = \quad \exp(ip \cdot X_t + t\psi(p)) \quad = \quad \exp(ip \cdot M_t + t(\psi(p) + ip \cdot b)).$$

Noting that $[M^k, M^l]_t^c = [X^k, X^l]_t^c = a^{kl} t$ by (ii), Itô's formula yields

$$\begin{aligned} Z_t \quad &= \quad 1 + \int_0^t Z_- \, ip \cdot dM + \int_0^t Z_- \, (\psi(p) + ip \cdot b - \frac{1}{2} \sum_{k,l} p_k p_l a^{kl}) \, dt \qquad (14.5.12) \\ &\quad + \sum_{(0,t]} Z_- \left(e^{ip \cdot \Delta X} - 1 - ip \cdot \Delta X \right). \end{aligned}$$

By (14.5.11) and since $|e^{ip \cdot y} - 1 - ip \cdot y| \le \text{const.} \cdot (|y| \wedge |y|^2)$, the series on the right hand side of (14.5.12) can be decomposed into a martingale and the finite variation process

$$A_t \quad = \quad \int_0^t \int Z_{s-} \, (e^{ip \cdot y} - 1 - ip \cdot y) \, \nu(dy) \, ds$$

Therefore, by (14.5.12) and (14.5.10), (Z_t) is a martingale for any $p \in \mathbb{R}^d$. The assertion now follows by Lemma 11.1. □

An interesting consequence of Theorem 14.29 is that a Brownian motion B and a Lévy process without diffusion part w.r.t. the same filtration are always independent, because $[B^k, X^l] = 0$ for any k, l.

Exercise (Independence of Brownian motion and Lévy processes). Suppose that $B_t : \Omega \to \mathbb{R}^d$ and $X_t : \Omega \to \mathbb{R}^{d'}$ are a Brownian motion and a Lévy process with characteristic exponent $\psi_X(p) = -ip \cdot b + \int (1 - e^{ip \cdot y} + ip \cdot y) \, \nu(dy)$ defined on the same filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$. Assuming that $\int (|y| \wedge |y|^2) \, \nu(dy) < \infty$, prove that (B_t, X_t) is a Lévy process on $\mathbb{R}^{d \times d'}$ with characteristic exponent

$$\psi(p, q) \quad = \quad \frac{1}{2} |p|^2_{\mathbb{R}^d} + \psi_X(q), \qquad p \in \mathbb{R}^d, \ q \in \mathbb{R}^{d'}.$$

Hence conclude that B and X are independent.

Change of measure for general semimartingales

We conclude this section with a general change of measure theorem for possibly discontinuous semimartingales:

Theorem 14.30 (P.A. Meyer). *Suppose that the probability measures P and Q are equivalent on $\mathcal{F}_t$ for any $t \geq 0$ with relative density $\frac{dP}{dQ}\Big|_{\mathcal{F}_t} = Z_t$. If M is a local martingale w.r.t. Q then $M - \int \frac{1}{Z}\, d[Z, M]$ is a local martingale w.r.t. P.*

The theorem shows that w.r.t. P, (M_t) is again a semimartingale, and it yields an explicit semimartingale decomposition for (M, P). For the proof we recall that (Z_t) is a local martingale w.r.t. Q and $(1/Z_t)$ is a local martingale w.r.t. P.

Proof. The process $ZM - [Z, M]$ is a local martingale w.r.t. Q. Hence by Lemmy 9.8, the process $M - \frac{1}{Z}[Z, M]$ is a local martingale w.r.t. P. It remains to show that $\frac{1}{Z}[Z, M]$ differs from $\int \frac{1}{Z}\, d[Z, M]$ by a local P-martingale. This is a consequence of the Itô product rule: Indeed,

$$\frac{1}{Z}[Z, M] \quad = \quad \int [Z, M]_-\, d\frac{1}{Z} + \int \frac{1}{Z_-}\, d[Z, M] + \big[\frac{1}{Z}, [Z, M]\big].$$

The first term on the right-hand side is a local Q-martingale, since $1/Z$ is a Q-martingale. The remaining two terms add up to $\int \frac{1}{Z}\, d[Z, M]$, because

$$\big[\frac{1}{Z}, [Z, M]\big] \quad = \quad \sum \Delta \frac{1}{Z}\, \Delta[Z, M].$$

□

Remark. Note that the process $\int \frac{1}{Z}\, d[Z, M]$ is not predictable in general. For a predictable counterpart to Theorem 14.30 cf. e.g. [36].

14.6 General predictable integrands

So far, we have considered stochastic integrals w.r.t. general semimartingales only for integrands that are left limits of adapted càdlàg processes. This is indeed sufficient for many applications. For some results including in particular convergence theorems for stochastic integrals, martingale representation theorems and the existence of local time, stochastic integrals with more general integrands are important. In this section, we sketch the definition of stochastic integrals w.r.t. not necessarily continuous semimartingales for general predictable integrands. For details of the proofs, we refer to Chapter IV in Protter's book [36].

Throughout this section, we fix a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$. Recall that the **predictable σ-algebra** $\mathcal{P}$ on $\Omega \times (0,\infty)$ is generated by all sets $A \times (s,t]$ with $A \in \mathcal{F}_s$ and $0 \leq s \leq t$, or, equivalently, by all left-continuous $(\mathcal{F}_t)$ adapted processes $(\omega, t) \mapsto G_t(\omega)$. We denote by $\mathcal{E}$ the vector space consisting of all **elementary predictable processes** G of the form

$$G_t(\omega) \quad = \quad \sum_{i=0}^{n-1} Z_i(\omega) I_{(t_i,t_{i+1}]}(t)$$

with $n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n$, and $Z_i : \Omega \to \mathbb{R}$ bounded and $\mathcal{F}_{t_i}$-measurable. For $G \in \mathcal{E}$ and a semimartingale $X \in \mathcal{S}$, the stochastic integral $G_\bullet X$ defined by

$$(G_\bullet X)_t \quad = \quad \int_0^t G_s \, dX_s \quad = \quad \sum_{i=0}^{n-1} Z_i \left(X_{t_{i+1}\wedge t} - X_{t_i \wedge t} \right)$$

is again a semimartingale. Clearly, if A is a finite variation process then $G_\bullet A$ has finite variation as well.

Now suppose that $M \in M_d^2(0,\infty)$ is a square-integrable martingale. Then $G_\bullet M \in M_d^2(0,\infty)$, and the Itô isometry

$$\begin{aligned} ||G_\bullet M||^2_{M^2(0,\infty)} \quad &= \quad E\left[\left(\int_0^\infty G \, dM \right)^2 \right] \\ &= \quad E\left[\int_0^\infty G^2 \, d[M] \right] \quad = \quad \int_{\Omega \times \mathbb{R}_+} G^2 \, dP_{[M]} \end{aligned} \tag{14.6.1}$$

holds, where

$$P_{[M]}(d\omega \, dt) \quad = \quad P(d\omega) \, [M](\omega)(dt)$$

is the **Doléans measure** of the martingale M on $\Omega \times \mathbb{R}_+$. The Itô isometry has been derived in a more general form in Corollary 14.20, but for elementary processes it can easily be verified directly (Excercise!).

In many textbooks, the angle bracket process $\langle M\rangle$ is used instead of $[M]$ to define stochastic integrals. The next remark shows that this is equivalent for predictable integrands:

Remark ($[M]$ vs. $\langle M\rangle$). Let $M \in M_d^2(0,\infty)$. If the angle-bracket process $\langle M\rangle$ exists then **the measures $P_{[M]}$ and $P_{\langle M\rangle}$ coincide on predictable sets.** Indeed, if $C = A \times (s,t]$ with $A \in \mathcal{F}_s$ and $0 \le s \le t$ then

$$\begin{aligned} P_{[M]}(C) &= E\left[[M]_t - [M]_s\,;\,A\right] = E\left[E[[M]_t - [M]_s|\mathcal{F}_s]\,;\,A\right] \\ &= E\left[E[\langle M\rangle_t - \langle M\rangle_s|\mathcal{F}_s]\,;\,A\right] = P_{\langle M\rangle}(C). \end{aligned}$$

Since the collection of these sets C is an $\cap$-stable generator for the predictable σ-algebra, the measures $P_{[M]}$ and $P_{\langle M\rangle}$ coincide on $\mathcal{P}$.

Example (Doléans measures of Lévy martingales). If $M_t = X_t - E[X_t]$ with a square integrable Lévy process $X_t : \Omega \to \mathbb{R}$ then

$$P_{[M]} = P_{\langle M\rangle} = \psi''(0)\, P \otimes \lambda_{(0,\infty)}$$

where ψ is the characteristic exponent of X and $\lambda_{(0,\infty)}$ denotes Lebesgue measure on $\mathbb{R}_+$. Hence the Doléans measure of a general Lévy martingale coincides with the one for Brownian motion up to a multiplicative constant.

Definition of stochastic integrals w.r.t. semimartingales

We denote by $\mathcal{H}^2$ the vector space of all semimartingales vanishing at 0 of the form $X = M + A$ with $M \in M_d^2(0,\infty)$ and $A \in$ FV predictable with total variation $V_\infty^{(1)}(A) = \int_0^\infty |dA_s| \in L^2(P)$. In order to define a norm on the space $\mathcal{H}^2$, we assume without proof the following result, cf. e.g. Chapter III in Protter [36]:

Fact. Any predictable local martingale with finite variation paths is almost surely constant.

The result implies that the *Doob-Meyer semimartingale decomposition*

$$X = M + A \tag{14.6.2}$$

is unique if we assume that M is local martingale and A is a **predictable** finite variation process vanishing at 0. Therefore, we obtain a **well-defined norm** on $\mathcal{H}^2$ by setting

$$||X||^2_{\mathcal{H}^2} \quad = \quad ||M||^2_{M^2} + ||V^{(1)}_\infty(A)||^2_{L^2} \quad = \quad E\left[[M]_\infty + \left(\int_0^\infty |dA| \right)^2 \right].$$

Note that the M^2 norm is the restriction of the $\mathcal{H}^2$ norm to the subspace $M^2(0,\infty) \subset \mathcal{H}^2$. As a consequence of (14.6.1), we obtain:

Corollary 14.31 (Itô isometry for semimartingales). *Let $X \in \mathcal{H}^2$ with semimartingale decomposition as above. Then*

$$||G_\bullet X||_{\mathcal{H}^2} \quad = \quad ||G||_X \qquad \text{for any } \; G \in \mathcal{E}, \; \text{ where}$$

$$||G||^2_X \quad := \quad ||G||^2_{L^2(P_{[M]})} + \left|\left| \int_0^\infty |G| \, |dA| \right|\right|^2_{L^2(P)}.$$

Hence the stochastic integral $\mathcal{J} : \mathcal{E} \to \mathcal{H}^2$, $\mathcal{J}_X(G) = G_\bullet X$, has a unique isometric extension to the closure $\overline{\mathcal{E}}^X$ of $\mathcal{E}$ w.r.t. the norm $|| \cdot ||_X$ in the space of all predictable processes in $L^2(P_{[M]})$.

Proof. The semimartingale decomposition $X = M + A$ implies a corresponding decomposition $G_\bullet X = G_\bullet M + G_\bullet A$ for the stochastic integrals. One can verify that for $G \in \mathcal{E}$, $G_\bullet M$ is in $M^2_d(0,\infty)$ and $G_\bullet A$ is a predictable finite variation process. Therefore, and by (14.6.1),

$$||G_\bullet X||^2_{\mathcal{H}^2} \;=\; ||G_\bullet M||^2_{M^2} + ||V^{(1)}_\infty(G_\bullet A)||^2_{L^2} \;=\; ||G||^2_{L^2(P_{[M]})} + \left|\left| \int |G| \, |dA| \right|\right|^2_{L^2(P)}.$$

□

The Itô isometry yields a definition of the stochastic integral $G_\bullet X$ for $G \in \overline{\mathcal{E}}^X$. For $G = H_-$ with H càdlàg and adapted, this definition is consistent with the definition given above since, by Corollary 14.20, the Itô isometry also holds for the integrals defined above, and the isometric extension is unique. The class $\overline{\mathcal{E}}^X$ of admissible integrands is already quite large:

Lemma 14.32. $\overline{\mathcal{E}}^X$ *contains all predictable processes G with $||G||_X < \infty$.*

Proof. We only mention the main steps of the proof, cf. [36] for details:

1) The approximation of bounded left-continuous processes by elementary predictable processes w.r.t. $||\cdot||_X$ is straightforward by dominated convergence.
2) The approximability of bounded predictable processes by bounded left-continuous processes w.r.t. $||\cdot||_X$ can be shown via the Monotone Class Theorem.
3) For unbounded predictable G with $||G||_X < \infty$, the processes $G^n := G \cdot I_{\{G\leq n\}}$, $n \in \mathbb{N}$, are predictable and bounded with $||G^n - G||_X \to 0$.

□

Localization

Having defined $G_\bullet X$ for $X \in \mathcal{H}^2$ and predictable integrands G with $||G||_X < \infty$, the next step is again a localization. This localization is slightly different than before, because there might be unbounded jumps at the localizing stopping times. To overcome this difficulty, the process is stopped just before the stopping time T, i.e., at $T-$. However, stopping at T_- destroys the martingale property if T is not a predictable stopping time. Therefore, it is essential that we localize semimartingales instead of martingales!

For a semimartingale X and a stopping time T we define the stopped process X^{T-} by

$$X_t^{T-} \quad = \quad \begin{cases} X_t & \text{for } t < T, \\ X_{T-} & \text{for } t \geq T > 0, \\ 0 & \text{for } T = 0. \end{cases}$$

The definition for $T = 0$ is of course rather arbitrary. It will not be relevant below, since we are considering sequences (T_n) of stopping times with $T_n \uparrow \infty$ almost surely. We state the following result from Chapter IV in [36] without proof.

Fact. If X is a semimartingale with $X_0 = 0$ then there exists an increasing sequence (T_n) of stopping times with $\sup T_n = \infty$ such that $X^{T_n-} \in \mathcal{H}^2$ for any $n \in \mathbb{N}$.

Now we are ready to state the definition of stochastic integrals for general predictable integrands w.r.t. general semimartingales X. By setting $G_\bullet X = G_\bullet(X - X_0)$ we may assume $X_0 = 0$.

Definition. *Let X be a semimartingale with $X_0 = 0$. A predictable process G is called* ***integrable w.r.t. X*** *iff there exists an increasing sequence (T_n) of stopping times such that* $\sup T_n = \infty$ *a.s., and for any $n \in \mathbb{N}$, $X^{T_n-} \in \mathcal{H}^2$ and $||G||_{X^{T_n-}} < \infty$.*
If G is integrable w.r.t. X then the ***stochastic integral*** *$G_\bullet X$ is defined by*

$$(G_\bullet X)_t \quad = \quad \int_0^t G_s\, dX_s \quad = \quad \int_0^t G_s\, dX_s^{T_n-} \qquad \textit{for any } \ t \in [0, T_n), \ \ n \in \mathbb{N}.$$

Of course, one has to verify that $G_\bullet X$ is well-defined. This requires in particular a locality property for the stochastic integrals that are used in the localization. We do not carry out the details here, but refer once more to Chapter IV in [36].

Exercise (Sufficient conditions for integrability of predictable processes).

1) Prove that if G is predictable and **locally bounded** in the sense that G^{T_n} is bounded for a sequence (T_n) of stopping times with $T_n \uparrow \infty$, then G is integrable w.r.t. any semimartingale $X \in \mathcal{S}$.

2) Suppose that $X = M + A$ is a continuous semimartingale with $M \in \mathcal{M}_c^{\text{loc}}$ and $A \in \text{FV}_c$. Prove that G is integrable w.r.t. X if G is predictable and

$$\int_0^t G_s^2\, d[M]_s + \int_0^t |G_s|\, |dA_s| \quad < \quad \infty \qquad \text{a.s. for any } \ t \geq 0.$$

Properties of the stochastic integral

Most of the properties of stochastic integrals can be extended easily to general predictable integrands by approximation with elementary processes and localization. The proof of Property (2) below, however, is not trivial. We refer to Chapter IV in [36] for detailed proofs of the following basic properties:

(1) The map $(G, X) \mapsto G_\bullet X$ is bilinear.

(2) $\Delta(G_\bullet X) = G\Delta X$ almost surely.

(3) $(G_\bullet X)^T = (G\, I_{[0,T]})_\bullet X = G_\bullet X^T$.

(4) $(G_\bullet X)^{T-} = G_\bullet X^{T-}$.

(5) $\widetilde{G}_\bullet(G_\bullet X) = (\widetilde{G}G)_\bullet X$.

In all statements, X is a semimartingale, G is a process that is integrable w.r.t. X, T is a stopping time, and $\widetilde{G}$ is a process such that $\widetilde{G}G$ is also integrable w.r.t. X. We state the formula for the covariation of stochastic integrals separately below, because its proof is based on the Kunita-Watanabe inequality, which is of independent interest.

Exercise (Kunita-Watanabe inequality). Let $X, Y \in \mathcal{S}$, and let G, H be measurable processes defined on $\Omega \times (0, \infty)$ (predictability is not required). Prove that for any $a \in [0, \infty]$ and $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$, the following inequalities hold:

$$\int_0^a |G||H|\, |d[X,Y]| \quad \leq \quad \Big(\int_0^a G^2\, d[X]\Big)^{1/2} \Big(\int_0^a H^2\, d[Y]\Big)^{1/2}, \qquad (14.6.3)$$

$$E\Big[\int_0^a |G||H|\, |d[X,Y]|\Big] \quad \leq \quad \Big\|\Big(\int_0^a G^2\, d[X]\Big)^{1/2}\Big\|_{L^p} \Big\|\Big(\int_0^a H^2\, d[Y]\Big)^{1/2}\Big\|_{L^q}. \qquad (14.6.4)$$

Hint: First consider elementary processes G, H.

Theorem 14.33 (Covariation of stochastic integrals). *For any $X, Y \in \mathcal{S}$ and any predictable process G that is integrable w.r.t. X,*

$$\Big[\int G\, dX, Y\Big] \quad = \quad \int G\, d[X,Y] \qquad \textit{almost surely.} \qquad (14.6.5)$$

Remark. If X and Y are local martingales, and the angle-bracket process $\langle X, Y\rangle$ exists, then also

$$\Big\langle \int G\, dX, Y \Big\rangle \quad = \quad \int G\, d\langle X, Y\rangle \qquad \text{almost surely.}$$

Proof of Theorem 14.33. We only sketch the main steps briefly, cf. [36] for details. Firstly, one verifies directly that (14.6.5) holds for $X, Y \in \mathcal{H}^2$ and $G \in \mathcal{E}$. Secondly, for $X, Y \in \mathcal{H}^2$ and a predictable process G with $||G||_X < \infty$ there exists a sequence (G^n) of elementary predictable processes such that $||G^n - G||_X \to 0$, and

$$\Big[\int G^n \, dX, Y\Big] \quad = \quad \int G^n \, d[X, Y] \qquad \text{for any } \; n \in \mathbb{N}.$$

As $n \to \infty$, $\int G^n \, dX \to \int G \, dX$ in $\mathcal{H}^2$ by the Itô isometry for semimartingales, and hence

$$\Big[\int G^n \, dX, Y\Big] \quad \longrightarrow \quad \Big[\int G \, dX, Y\Big] \qquad \text{u.c.p.}$$

by Corollary 14.14. Moreover,

$$\int G^n \, d[X, Y] \quad \longrightarrow \quad \int G \, d[X, Y] \qquad \text{u.c.p.}$$

by the Kunita-Watanabe inequality. Hence (14.6.5) holds for G as well. Finally, by localization, the identity can be extended to general semimartingales X, Y and integrands G that are integrable w.r.t. X. ☐

An important motivation for the extension of stochastic integrals to general predictable integrands is the validity of a Dominated Convergence Theorem:

Theorem 14.34 (Dominated Convergence Theorem for stochastic integrals). *Suppose that X is a semimartingale with decomposition $X = M + A$ as above, and let G^n, $n \in \mathbb{N}$, and G be predictable processes. If*

$$G^n_t(\omega) \quad \longrightarrow \quad G_t(\omega) \qquad \textit{for any } \; t \geq 0, \quad \textit{almost surely},$$

and if there exists a process H that is integrable w.r.t. X such that $|G^n| \leq H$ for any $n \in \mathbb{N}$, then

$$G^n_\bullet X \quad \longrightarrow \quad G_\bullet X \qquad \textit{u.c.p.} \quad \textit{as } \; n \to \infty.$$

If, in addition to the assumptions above, X is in $\mathcal{H}^2$ and $||H||_X < \infty$ then even

$$||G^n_\bullet X - G_\bullet X||_{\mathcal{H}^2} \quad \longrightarrow \quad 0 \qquad \textit{as } \; n \to \infty.$$

Proof. We may assume $G = 0$, otherwise we consider $G^n - G$ instead of G^n. Now suppose first that X is in $\mathcal{H}^2$ and $||H||_X < \infty$. Then

$$||G^n||_X^2 \quad = \quad E\Big[\int_0^\infty |G^n|^2 \, d[M] + \Big(\int_0^\infty |G^n| \, |dA|\Big)^2\Big] \quad \longrightarrow \quad 0$$

as $n \to \infty$ by the Dominated Convergence Theorem for Lebesgue integrals. Hence by the Itô isometry,

$$G^n_\bullet X \quad \longrightarrow \quad 0 \qquad \text{in } \mathcal{H}^2 \quad \text{as } n \to \infty.$$

The general case can now be reduced to this case by localization, where $\mathcal{H}^2$ convergence is replaced by the weaker ucp-convergence. □

We finally remark that basic properties of stochastic integrals carry over to integrals with respect to compensated Poisson point processes. We refer to the monographs by D.Applebaum [5] for basics, and to Jacod & Shiryaev [24] for a detailed study. We only state the following extension of the associative law, which has already been used in the last section:

Exercise (Integration w.r.t. stochastic integrals based on compensated PPP). Suppose that $H : \Omega \times \mathbb{R}_+ \times S \to \mathbb{R}$ is predictable and square-integrable w.r.t. $P \otimes \lambda \otimes \nu$, and $G : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is a bounded predictable process. Show that if

$$X_t \quad = \quad \int_{(0,t]\times S} H_s(y) \, \widetilde{N}(ds \, dy)$$

then

$$\int_0^t G_s \, dX_s \quad = \quad \int_{(0,t]\times S} G_s H_s(y) \, \widetilde{N}(ds \, dy).$$

Hint: Approximate G by elementary processes.

Appendix A

Conditional expectations

A.1 Conditioning on discrete random variables

We first consider conditioning on the value of a random variable $Y : \Omega \to S$ where S is countable. In this case, we can define the *conditional probability measure*

$$P[A \mid Y = z] \quad = \quad \frac{P[A \cap \{Y = z\}]}{P[Y = z]}, \qquad A \in \mathcal{A},$$

and the *conditional expectations*

$$E[X \mid Y = z] \quad = \quad \frac{E[X; Y = z]}{P[Y = z]}, \qquad X \in \mathcal{L}^1(\Omega, \mathcal{A}, P),$$

for any $z \in S$ with $P[Y = z] > 0$ in an elementary way. Note that for $z \in S$ with $P[Y = z] = 0$, the conditional probabilities are not defined.

Conditional expectations as random variables

It will turn out to be convenient to consider the conditional probabilities and expectations not as functions of the outcome z, but as functions of the random variable Y. In this way, the conditional expectations become random variables:

Definition (Conditional expectation given a discrete random variable). *Let $X : \Omega \to \mathbb{R}$ be a random variable such that $E[X^-] < \infty$, and let $Y : \Omega \to S$ be a discrete random variable. The random variable $E[X \mid Y]$ that is P-almost surely uniquely defined by*

$$E[X \mid Y] \quad := \quad g(Y) \quad = \quad \sum_{z \in S} g(z) \cdot I_{\{Y=z\}}$$

with

$$g(z) \quad := \quad \begin{cases} E[X \mid Y = z] & \text{if } P[Y = z] > 0 \\ \text{arbitrary} & \text{if } P[Y = z] = 0 \end{cases}$$

*is called **(a version of the) conditional expectation of** X **given** Y. For an event $A \in \mathcal{A}$, the random variable*

$$P[A \,|\, Y] \quad := \quad E[I_A \,|\, Y]$$

*is called **(a version of the) conditional probability of** A **given** Y.*

The conditional expectation $E[X \,|\, Y]$ and the conditional probability $P[A \,|\, Y]$ are again random variables.They take the values $E[X \mid Y = z]$ and $P[A \mid Y = z]$, respectively, on the sets $\{Y = z\}, z \in S$ with $P[Y = z] > 0$. On each of the null sets $\{Y = z\}, z \in S$ with $P[Y = z] = 0$, an arbitrary constant value is assigned to the conditional expectation. Hence the definition is only almost surely unique.

Characteristic properties of conditional expectations

Let $X : \Omega \to \mathbb{R}$ be a non-negative or integrable random variable on a probability space $(\Omega, \mathcal{A}, P)$. The following alternative characterisation of the conditional expectation of X given Y can be verified in an elementary way:

Theorem A.1. *A real random variable $\overline{X} \geq 0$ (or $\overline{X} \in \mathcal{L}^1$) on $(\Omega, \mathcal{A}, P)$ is a version of the conditional expectation $E[X \mid Y]$ if and only if*

(I) $\overline{X} = g(Y)$ *for a function $g : S \to \mathbb{R}$, and*

(II) $E\left[\overline{X} \cdot f(Y)\right] = E[X \cdot f(Y)]$ *for all non-negative or bounded functions* $f : S \to \mathbb{R}$, *respectively.*

A.2 General conditional expectations

If Y is a real-valued random variable on a probability space $(\Omega, \mathcal{A}, P)$ with continuous distribution function, then $P[Y = z] = 0$ for any $z \in \mathbb{R}$. Therefore, conditional probabilities given $Y = z$ can not be defined in the same way as above. Alternatively, one could try to define conditional probabilities given Y as limits:

$$P[A \,|\, Y = z] \quad = \quad \lim_{h \searrow 0} P[A \,|\, z - h \leq Y \leq z + h]. \tag{A.2.1}$$

In certain cases this is possible but in general, the existence of the limit is not guaranteed.

Instead, the characterization in Theorem A.1 is used to provide a definition of conditional expectations given general random variables Y. The conditional probability of a fixed event A given Y can then be defined almost surely as a special case of a conditional expectation:

$$P[A \,|\, Y] \quad := \quad E[I_A \,|\, Y]. \tag{A.2.2}$$

Note, however, that in general, the exceptional set will depend on the event A !

The factorization lemma

We first prove an important measure theoretic statement.

Theorem A.2 (Factorization lemma). *Suppose that* $(S, \mathcal{S})$ *is a measurable space and* $Y : \Omega \to S$ *is a map. Then a map* $X : \Omega \to \mathbb{R}$ *is measurable w.r.t.* $\sigma(Y)$ *if and only if*

$$X = f(Y) = f \circ Y$$

for a $\mathcal{S}$-measurable function $f : S \to \mathbb{R}$.

$$\begin{tikzcd} (\Omega, \sigma(Y)) \arrow[r, "Y"] \arrow[rr, bend left, "X"] & (S, \mathcal{S}) \arrow[r] & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{tikzcd}$$

Proof. (1). If $X = f \circ Y$ for a measurable function f, then

$$X^{-1}(B) \quad = \quad Y^{-1}(f^{-1}(B)) \in \sigma(Y) \qquad \text{holds for all } B \in \mathcal{B}(\mathbb{R}),$$

as $f^{-1}(B) \in \mathcal{S}$. Therefore, X is $\sigma(Y)$-measurable.

(2). Coversely, we have to show that $\sigma(Y)$-measurability of X implies that X is a measurable function of Y. This is done in several steps:

(a) If $X = I_A$ is an indicator function of an set $A \in \sigma(Y)$, then $A = Y^{-1}(B)$ with $B \in \mathcal{S}$, and thus

$$X(\omega) \quad = \quad I_{Y^{-1}(B)}(\omega) \quad = \quad I_B(Y(\omega)) \qquad \text{for all } \omega \in \Omega.$$

(b) For $X = \sum_{i=1}^n c_i I_{A_i}$ with $A_i \in \sigma(Y)$ and $c_i \in \mathbb{R}$ we have correspondingly

$$X = \sum_{i=1}^n c_i I_{B_i}(Y),$$

where wobei B_i are sets in $\mathcal{S}$ such that $A_i = Y^{-1}(B_i)$.

(c) For an arbitrary non-negative, $\sigma(Y)$-measurable map $X : \Omega \to \mathbb{R}$, there exists a sequence of $\sigma(Y)$-measurable elementary functions such that $X_n \nearrow X$. By (b), $X_n = f_n(Y)$ with $\mathcal{S}$-measurable functions f_n. Hence

$$X \quad = \quad \sup X_n \quad = \quad \sup f_n(Y) \quad = \quad f(Y),$$

where $f = \sup f_n$ is again $\mathcal{S}$-measurable.

(d) For a general $\sigma(Y)$-measurable map $X : \Omega \to \mathbb{R}$, both X^+ and X^- are measurable functions of Y, hence X is a measurable function of Y as well.

□

The factorization lemma can be used to rephrase the *characterizing properties* (I) und (II) of conditional expectations in Theorem A.1 in the following way:

$\overline{X}$ is a version of $E[X \mid Y]$ if and only if

(i) $\overline{X}$ ist $\sigma(Y)$-messbar,

(ii) $E[\overline{X}\ ;\ A] \;=\; E[X\ ;\ A]$ für alle $A \in \sigma(Y)$.

The equivalence of (I) und (i) is a consequence of the factorization lemma, and the equivalence of (II) and (ii) follows by monotone classes, since (ii) states that

$$E[\overline{X} \cdot I_B(Y)] \quad = \quad E[X \cdot I_B(Y)] \qquad \text{holds for all } B \in \mathcal{S}.$$

Conditional expecations given σ-algebras

A remarkable consequence of the characterization of conditional expectations by Conditions (i) and (ii) is that the *conditional expectation $E[X \mid Y]$ depends on the random variable Y only via the σ-algebra $\sigma(Y)$ generated by Y !* If two random variables Y and Z are functions of each other then $\sigma(Y) = \sigma(Z)$, and hence the conditional expectations $E[X \mid Y]$ and $E[X \mid Z]$ coincide (with probability 1). Therefore it is plausible to define directly the conditional expectation given a σ-Algebra. The σ-algebra (e.g. $\sigma(Y)$, or $\sigma(Y_1, \ldots, Y_n)$) then describes the available "information" on which we are conditioning.

The characterization of conditional expectations by (i) and (ii) can be extended immediately to the case of general conditional expectations given a σ-algebra or given arbitrary random variables. To this end let $X : \Omega \to \overline{\mathbb{R}}$ be a non-negative (or integrable) random variable on a probability space $(\Omega, \mathcal{A}, P)$.

Definition (Conditional expectation, general). *(1). Let $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. A non-negative (or integrable) random variable $\overline{X} : \Omega \to \overline{\mathbb{R}}$ is called a **version of the conditional expectation** $E[X \mid \mathcal{F}]$ iff:*

(a) $\overline{X}$ *is* $\mathcal{F}$*-measurable,* *and*

(b) $E[\overline{X}\ ;\ A] = E[X\ ;\ A]$ *for any* $A \in \mathcal{F}$.

(2). For arbitrary random variables $Y, Y_1, Y_2, \ldots, Y_n$ *on* $(\Omega, \mathcal{A}, P)$ *we define*

$$\begin{aligned} E[X \mid Y] &:= E[X \mid \sigma(Y)], \\ E[X \mid Y_1, \ldots Y_n] &:= E[X \mid (Y_1, \ldots, Y_n)] = E[X \mid \sigma(Y_1, \ldots, Y_n)]. \end{aligned}$$

(3). For an event $A \in \mathcal{A}$ *we define*

$$P[A|\mathcal{F}] \quad := \quad E[I_A|\mathcal{F}], \quad \textit{and correspondingly} \quad P[A|Y] \quad = \quad E[I_A|Y].$$

Remark. By monotone classes it can be shown that Condition (b) is equivalent to:

(b') $E[\overline{X} \cdot Z] = E[X \cdot Z]$ for any non-negative (resp. bounded) $\mathcal{F}$-measurable $Z : \Omega \to \mathbb{R}$.

Theorem A.3 (Existence and uniqueness of conditional expectations). *Let* $X \geq 0$ *or* $X \in \mathcal{L}^1$*, and let* $\mathcal{F} \subseteq \mathcal{A}$ *be a* σ*-algebra. Then:*

(1). There exists a version of the conditional expectation $E[X \mid \mathcal{F}]$.

(2). Any two versions coincide P*-almost surely.*

Proof. Existence can be shown as a consequence of the Radon-Nikodym theorem. In Theorem A.8 below, we give a different proof of existence that only uses elementary methods.

For proving uniqueness let $\overline{X}$ and $\widetilde{X}$ be two versions of $E[X \mid \mathcal{F}]$. Then both $\overline{X}$ and $\widetilde{X}$ are $\mathcal{F}$-measurable, and

$$E[\overline{X}\ ;\ A] \quad = \quad E[\widetilde{X}\ ;\ A] \qquad \text{for any } A \in \mathcal{F}.$$

Therefore, $\overline{X} = \widetilde{X}$ P-almost surely. □

Properties of conditional expectations

Starting form the definition, we now derive several basic properties of conditional expectations that are used frequently:

Theorem A.4. *Let X, Y and X_n $(n \in \mathbb{N})$ be non-negative or integrable random variables on $(\Omega, \mathcal{A}, P)$, and let $\mathcal{F}, \mathcal{G} \subseteq \mathcal{A}$ be σ-algebras.*
The following assertions hold:

(1). *Linearity: $E[\lambda X + \mu Y \mid \mathcal{F}] = \lambda\, E[X \mid \mathcal{F}] + \mu\, E[Y \mid \mathcal{F}]$ P-almost surely for any $\lambda, \mu \in \mathbb{R}$.*

(2). *Monotonicity: If $X \geq 0$ P-almost surely, then $E[X \mid \mathcal{F}] \geq 0$ P-almost surely.*

(3). *If $X = Y$ P-almost surely then $E[X \mid \mathcal{F}] = E[Y \mid \mathcal{F}]$ P-almost surely.*

(4). *Monotone Convergence: If (X_n) is increasing with $X_1 \geq 0$, then*

$$E[\sup X_n \mid \mathcal{F}] \quad = \quad \sup E[X_n \mid \mathcal{F}] \qquad P\text{-almost surely.}$$

(5). *Tower Property: If $\mathcal{G} \subseteq \mathcal{F}$ then*

$$E[E[X \mid \mathcal{F}] \mid \mathcal{G}] \quad = \quad E[X \mid \mathcal{G}] \qquad P\text{-almost surely.}$$

In particular,

$$E[E[X \mid Y, Z] \mid Y] \quad = \quad E[X|Y] \qquad P\text{-almost surely.}$$

(6). *Taking out what is known: Let Y be $\mathcal{F}$-measurable such that $Y \cdot X \in \mathcal{L}^1$ or ≥ 0. Then*

$$E[Y \cdot X \mid \mathcal{F}] \quad = \quad Y \cdot E[X \mid \mathcal{F}] \qquad P\text{-almost surely.}$$

(7). *Independence: If X is independent of $\mathcal{F}$ then $E[X \mid \mathcal{F}] = E[X]$ P-almost surely.*

(8). *Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be measurable spaces. If $Y : \Omega \to S$ is $\mathcal{F}$-measurable and $X : \Omega \to T$ is independent of $\mathcal{F}$, then for any product-measurable function $f : S \times T \to [0, \infty)$ we have*

$$E[f(X, Y) \mid \mathcal{F}](\omega) \quad = \quad E[f(X, Y(\omega))] \qquad \text{für } P\text{-fast alle } \omega.$$

Proof. (1). Aus der Linearität des Erwartungswertes folgt, dass $\lambda E[X\,|\,\mathcal{F}]+\mu E[Y\,|\,\mathcal{F}]$ eine Version der bedingten Erwartung $E[\lambda X + \mu Y \mid \mathcal{F}]$ ist.

(2). Sei $\overline{X}$ eine Version von $E[X \mid \mathcal{F}]$. Aus $X \geq 0$ P-fast sicher folgt wegen $\{\overline{X} < 0\} \in \mathcal{F}$:

$$E[\overline{X}\,;\,\overline{X} < 0] \quad = \quad E[X\,;\,\overline{X} < 0] \quad \geq \quad 0,$$

und damit $\overline{X} \geq 0$ P-fast sicher.

(3). Dies folgt unmittelbar aus (1) und (2).

(4). Ist $X_n \geq 0$ und monoton wachsend, dann ist $\sup E[X_n \mid \mathcal{F}]$ eine nichtnegative $\mathcal{F}$-messbare Zufallsvariable (mit Werten in $[0,\infty]$), und nach dem "'klassischen "' Satz von der monotonen Konvergenz gilt:

$$E[\sup E[X_n\,|\,\mathcal{F}]\cdot Z] \;=\; \sup E[E[X_n\,|\,\mathcal{F}]\cdot Z] \;=\; \sup E[X_n\cdot Z] \;=\; E[\sup X_n\cdot Z]$$

für jede nichtnegative $\mathcal{F}$-messbare Zufallsvariable Z. Also ist $\sup E[X_n\,|\,\mathcal{F}]$ eine Version der bedingten Erwartung von $\sup X_n$ gegeben $\mathcal{F}$.

(5). Wir zeigen, dass jede Version von $E[X\,|\,\mathcal{G}]$ auch eine Version von $E[E[X\,|\,\mathcal{F}]\,|\,\mathcal{G}]$ ist, also die Eigenschaften (i) und (ii) aus der Definition der bedingten Erwartung erfüllt:

(i) $E[X \mid \mathcal{G}]$ ist nach Definition $\mathcal{G}$-messbar.

(ii) Für $A \in \mathcal{G}$ gilt auch $A \in \mathcal{F}$, und somit $E[E[X \mid \mathcal{G}]\,;\,A] = E[X\,;\,A] = E[E[X \mid \mathcal{F}]\,;\,A]$.

(6) und (7). Auf ähnliche Weise verifiziert man, dass die Zufallsvariablen, die auf der rechten Seite der Gleichungen in (6) und (7) stehen, die definierenden Eigenschaften der bedingten Erwartungen auf der linken Seite erfüllen (Übung).

(8). Dies folgt aus (6) und (7) in drei Schritten:

(a) Gilt $f(x,y) = g(x)\cdot h(y)$ mit messbaren Funktionen $g,h \geq 0$, dann folgt nach (6) und (7) P-fast sicher:

$$\begin{aligned} E[f(X,Y) \mid \mathcal{F}] &= E[g(X)\cdot h(Y) \mid \mathcal{F}] \quad = \quad h(Y)\cdot E[g(X)|\mathcal{F}] \\ &= h(Y)\cdot E[g(X)], \end{aligned}$$

und somit

$$E[f(X,Y)|\mathcal{F}](\omega) = E[g(X)\cdot h(Y(\omega))] = E[f(X,Y(\omega))] \qquad \text{für } P\text{-fast alle } \omega.$$

(b) Um die Behauptung für Indikatorfunktionen $f(x,y) = I_B(x,y)$ von produktmessbaren Mengen B zu zeigen, betrachten wir das Mengensystem

$$\mathcal{D} \quad = \quad \{B \in \mathcal{S} \otimes \mathcal{T} \mid \text{Behauptung gilt für } f = I_B\}.$$

$\mathcal{D}$ ist ein Dynkinsystem, das nach (a) alle Produkte $B = B_1 \times B_2$ mit $B_1 \in \mathcal{S}$ und $B_2 \in \mathcal{T}$ enthält. Also gilt auch

$$\mathcal{D} \quad \supseteq \quad \sigma(\{B_1 \times B_2 \mid B_1 \in \mathcal{S}, B_2 \in \mathcal{T}\}) \quad = \quad \mathcal{S} \otimes \mathcal{T}.$$

(c) Für beliebige produktmessbare Funktionen $f : S \times T \to \mathbb{R}_+$ folgt die Behauptung nun durch maßtheoretische Induktion.

□

Remark (Convergence theorems for conditional expectations). The Monotone Convergence Theorem (Property (4)) implies versions of Fatou's Lemma and of the Dominated Convergence Theorem for conditional expectations. The proofs are similar to the unconditioned case.

The last property in Theorem A.4 is often very useful. For independent random variables X and Y it implies

$$E[f(X,Y) \mid Y](\omega) \quad = \quad E[f(X,Y(\omega))] \qquad \text{für } P\text{-fast alle } \omega, \tag{A.2.3}$$

We stress that independence of X and Y ist essential for (A.2.3) to hold true. The application of (A.2.3) without independence is a common mistake in computations with conditional expectations.

A.3 Conditional expectation as best L^2-approximation

In this section we show that the conditional expectation of a square integrable random variable X given a σ-algebra $\mathcal{F}$ can be characterized alternatively as the best approximation of X in the subspace of $\mathcal{F}$-measurable, square integrable random variables, or, equivalently, as the orthogonal projection of X onto this subspace. Besides obvious applications to non-linear predictions, this point of view is also the basis for a simple existence proof of conditional expectations

Jensen's inequality

Jensen's inequality is valid for conditional expectations as well. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ an integrable random variable, and $\mathcal{F} \subseteq \mathcal{A}$ a σ-algebra.

Theorem A.5 (Jensen). *If $u : \mathbb{R} \to \mathbb{R}$ is a convex function with $u(X) \in \mathcal{L}^1$ or $u \geq 0$, then*

$$E[u(X) \mid \mathcal{F}] \quad \geq \quad u(E[X \mid \mathcal{F}]) \qquad P\text{-almost surely.}$$

Proof. Jede konvexe Funktion u lässt sich als Supremum von abzählbar vielen affinen Funktionen darstellen, d.h. es gibt $a_n, b_n \in \mathbb{R}$ mit

$$u(x) \quad = \quad \sup_{n \in \mathbb{N}} (a_n x + b_n) \qquad \text{für alle } x \in \mathbb{R}.$$

Zum Beweis betrachtet man die Stützgeraden an allen Stellen einer abzählbaren dichten Teilmenge von $\mathbb{R}$, siehe z.B. [Williams: Probability with martingales, 6.6]. Wegen der Monotonie und Linearität der bedingten Erwartung folgt

$$E[u(X) \mid \mathcal{F}] \quad \geq \quad E[a_n X + b_n \mid \mathcal{F}] \quad = \quad a_n \cdot E[X \mid \mathcal{F}] + b_n$$

P-fast sicher für alle $n \in \mathbb{N}$, also auch

$$E[u(X) \mid \mathcal{F}] \quad \geq \quad \sup_{n \in \mathbb{N}} (a_n \cdot E[X \mid \mathcal{F}] + b_n) \qquad P\text{-fast sicher.}$$

□

Corollary A.6 (**L^p-contractivity**). *The map $X \mapsto E[X \mid \mathcal{F}]$ is a contraction on $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ for every $p \geq 1$, i.e.,*

$$E\left[|E[X \mid \mathcal{F}]|^p\right] \quad \leq \quad E[|X|^p] \qquad \text{for any } X \in \mathcal{L}^1(\Omega, \mathcal{A}, P).$$

Proof. Nach der Jensenschen Ungleichung gilt:

$$|E[X \mid \mathcal{F}]|^p \quad \leq \quad E[|X|^p \mid \mathcal{F}] \qquad P\text{-fast sicher.}$$

Die Behauptung folgt durch Bilden des Erwartungswertes. □

The proof of the corollary shows in particular that for a random variable $X \in \mathcal{L}^p$, the conditional expectation $E[X \mid \mathcal{F}]$ is contained in $\mathcal{L}^p$ as well. We now restrict ourselves to the case $p = 2$.

Conditional expectation as best L^2-prediction value

The space $L^2(\Omega, \mathcal{A}, P) = \mathcal{L}^2(\Omega, \mathcal{A}, P)/\sim$ of equivalence classes of square integrable random variables is a Hilbert space with inner product $(X, Y)_{L^2} = E[XY]$. If $\mathcal{F} \subseteq \mathcal{A}$ is a sub-σ-algebra then $L^2(\Omega, \mathcal{F}, P)$ is a **closed subspace** of $L^2(\Omega, \mathcal{A}, P)$, because limits of $\mathcal{F}$-measurable random variables are $\mathcal{F}$-measurable as well. For $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$, each version of the conditional expectation $E[X \mid \mathcal{F}]$ is contained in the subspace $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ by Jensen's inequality. Furthermore, the conditional expectation respects equivalence classes, see Theorem A.3. Therefore, $X \mapsto E[X \mid \mathcal{F}]$ induces a linear map from the Hilbert space $L^2(\Omega, \mathcal{A}, P)$ of equivalence classes onto the subspace $L^2(\Omega, \mathcal{F}, P)$.

Theorem A.7 (Characterization of the conditional expectation as best L^2 approximation and as orthogonal projection). *For $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ the following statements are all equivalent:*

(1). Y is a version of the conditional expectation $E[X \mid \mathcal{F}]$.

*(2). Y is a "**best approximation**" of X in the subspace $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, i.e.,*

$$E[(X-Y)^2] \quad \leq \quad E[(X-Z)^2] \qquad \text{for any } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P).$$

*(3). Y is a version of the **orthogonal projection** of X onto the subspace $L^2(\Omega, \mathcal{F}, P) \subseteq L^2(\Omega, \mathcal{A}, P)$, i.e.,*

$$E[(X-Y)\cdot Z] \quad = \quad 0 \qquad \text{for any } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P).$$

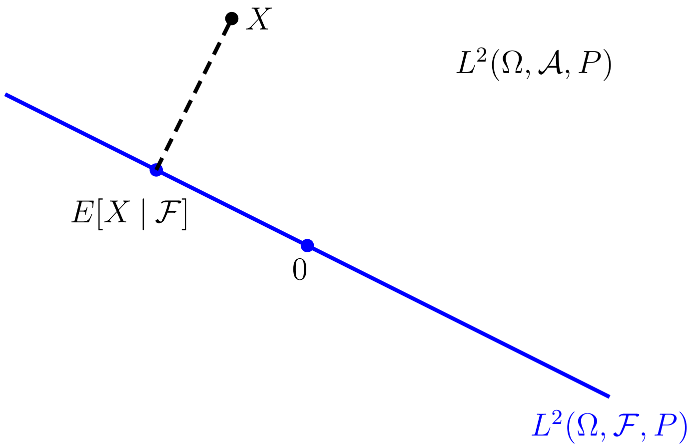


Figure A.1: $X \mapsto E[X \mid \mathcal{F}]$ as orthogonal projection onto the subspace $L^2(\Omega, \mathcal{F}, P)$.

Proof. **(1) $\iff$ (3):** Für $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ gilt:

$$\begin{aligned}
& Y \text{ ist eine Version von } E[X \mid \mathcal{F}] \\
\iff \quad & E[Y \cdot I_A] \;=\; E[X \cdot I_A] \qquad \text{für alle } A \in \mathcal{F} \\
\iff \quad & E[Y \cdot Z] \;=\; E[X \cdot Z] \qquad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P) \\
\iff \quad & E[(X-Y)\cdot Z] \;=\; 0 \qquad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)
\end{aligned}$$

Hierbei zeigt man die zweite Äquivalenz mit den üblichen Fortsetzungsverfahren (maßtheoretische Induktion).

(3) ⇒ (2): Sei Y eine Version der orthogonalen Projektion von X auf $L^2(\Omega, \mathcal{F}, P)$. Dann gilt für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$:

$$\begin{aligned} E[(X-Z)^2] &= E[((X-Y)+(Y-Z))^2] \\ &= E[(X-Y)^2] + E[(Y-Z)^2] + 2E[(X-Y)\underbrace{(Y-Z)}_{\in \mathcal{L}^2(\Omega,\mathcal{F},P)}] \\ &\geq E[(X-Y)^2] \end{aligned}$$

Hierbei haben wir im letzten Schritt verwendet, dass $Y-Z$ im Unterraum $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ enthalten, also orthogonal zu $X - Y$ ist.

(2) ⇒ (3): Ist umgekehrt Y eine beste Approximation von X in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ und $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, dann gilt

$$\begin{aligned} E[(X-Y)^2] &\leq E[(X-Y+tZ)^2] \\ &= E[(X-Y)^2] + 2tE[(X-Y)Z] + t^2E[Z^2] \end{aligned}$$

für alle $t \in \mathbb{R}$, also $E[(X-Y)\cdot Z] = 0$.

□

The equivalence of (2) and (3) is a well-known functional analytic statement: the best approximation of a vector in a closed subspace of a Hilbert space is the orthogonal projection of the vector onto this subspace. The geometric intuition behind this fact is indicated in Figure A.1.

Theorem A.7 is a justification for the interpretation of the conditional expectation as a predicion value. For example, by the factorization lemma, $E[X \mid Y]$ is the best L^2-prediction for X among all functions of type $g(Y)$, $g : \mathbb{R} \to \mathbb{R}$ measurable.

Existence of conditional expectations

By the characterization of the conditional expectation as the best L^2-approximation, the existence of conditional expectations of square integrable random variables is an immediate consequence of the existence of the best approximation of a vector in a closed subspace of a Hilbert space. By monotone approximation, the existence of conditional expectations of general non-negative random variables then follows easily.

Theorem A.8 (Existence of conditional expectations). *For every random variable $X \geq 0$ or $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, and every σ-algebra $\mathcal{F} \subseteq \mathcal{A}$, there exists a version of the conditional expectation $E[X \mid \mathcal{F}]$.*

Proof. (1). Wir betrachten zunächst den Fall $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Wie eben bemerkt, ist der Raum $L^2(\Omega, \mathcal{F}, P)$ ein abgeschlossener Unterraum des Hilbertraums $L^2(\Omega, \mathcal{A}, P)$. Sei $d = \inf\{\|Z - X\|_{L^2} \mid Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)\}$ der Abstand von X zu diesem Unterraum. Um zu zeigen, dass eine beste Approximation von X in $L^2(\Omega, \mathcal{F}, P)$ existiert, wählen wir eine Folge (X_n) aus diesem Unterraum mit $\|X_n - X\|_{L^2} \to d$. Mithilfe der Parallelogramm-Identität folgt für $n, m \in \mathbb{N}$:

$$\begin{aligned}
\|X_n - X_m\|_{L^2}^2 &= \|(X_n - X) - (X_m - X)\|_{L^2}^2 \\
&= 2 \cdot \|X_n - X\|_{L^2}^2 + 2 \cdot \|X_m - X\|_{L^2}^2 - \|(X_n - X) + (X_m - X)\|_{L^2}^2 \\
&= 2 \cdot \underbrace{\|X_n - X\|_{L^2}^2}_{\to d^2} + 2 \cdot \underbrace{\|X_m - X\|_{L^2}^2}_{\to d^2} - 4 \underbrace{\left\| \frac{X_n + X_m}{2} - X \right\|_{L^2}^2}_{\leq d^2},
\end{aligned}$$

und damit

$$\limsup_{n,m\to\infty} \|X_n - X_m\|_{L^2}^2 \quad \leq \quad 0.$$

Also ist die Minimalfolge (X_n) eine CauchyLfolge in dem vollständigen Raum $L^2(\Omega, \mathcal{F}, P)$, d.h. es existiert ein $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ mit

$$\|X_n - Y\|_{L^2} \quad \to \quad 0.$$

Für Y gilt

$$\|Y - X\|_{L^2} \quad = \quad \|\lim_{n\to\infty} X_n - X\|_{L^2} \quad \leq \quad \liminf_{n\to\infty} \|X_n - X\|_{L^2} \quad \leq \quad d,$$

d.h. Y ist die gesuchte Bestapproximation, und damit eine Version der bedingten Erwartung $E[X \mid \mathcal{F}]$.

(2). Für eine beliebige nichtnegative Zufallsvariable X auf $(\Omega, \mathcal{A}, P)$ existiert eine monoton wachsende Folge (X_n) nichtnegativer quadratintegrierbarer Zufallsvariablen mit $X = \sup X_n$. Man verifiziert leicht, dass $\sup\limits_n E[X_n \mid \mathcal{F}]$ eine Version von $E[X \mid \mathcal{F}]$ ist.

(3). Entsprechend verifiziert man, dass für allgemeine $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ durch $E[X|\mathcal{F}] = E[X^+ \mid \mathcal{F}] - E[X^- \mid \mathcal{F}]$ eine Version der bedingten Erwartung gegeben ist.

$\square$

Bibliography

[1]

[2]

[3] *Sur l'équation de Kolmogoroff, par W. Doeblin*. Éditions Elsevier, Paris, 2000. C. R. Acad. Sci. Paris Sér. I Math. **3**31 (2000), Special Issue.

[4] Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.

[5] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge UP, Cambridge, 2004.

[6] Ole E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.*, 2(1):41–68, 1998.

[7] Richard F. Bass. *Stochastic Processes*. Cambridge University Press, Cambridge, 2011.

[8] Jean Bertoin. *Lévy Processes*. Cambridge UP, Cambridge, 1998.

[9] Jean-Michel Bismut. *Large deviations and the Malliavin calculus*. Birkhaeuser, Boston, 1994.

[10] Amir Dembo and Ofur Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998, Berlin.

[11] Frank den Hollander. *Large Deviations*. American Mathematical Society, Providence, 2008.

[12] Andreas Eberle. Markov processes. `http://www.uni-bonn.de/~eberle/MarkovProcesses/MPSkript.pdf`, 2009.

[13] Andreas Eberle. Wahrscheinlichkeitstheorie und Stochastische Prozesse. `http://www.uni-bonn.de/~eberle/StoPr/WthSkriptneu.pdf`, 2010.

[14] Andreas Eberle. Introduction to Stochastic Analysis. `http://www.uni-bonn.de/~eberle/StoAn1011/StoAnSkriptneu.pdf`, 2011.

[15] David Elworthy. *On the Geometry of Diffusion Operators and Stochastic Flows*. Springer, 1999, Berlin.

[16] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes*. Wiley-Interscience, Hoboken, NJ, 2005.

[17] William Feller. *Introduction to Probability Theory and Its Applications*. Wiley, New York, 1957.

[18] Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[19] Peter K. Friz and Nicolas Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge UP, Cambridge, 2010.

[20] W. Hackenbroch and A. Thalmaier. *Stochastic Analysis*. Teubner, 1991.

[21] Martin Hairer. On malliavin's proof of hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[22] Elton Hsu. *Stochastic Analysis on Manifolds*. Oxford University Press, Oxford, 2002.

[23] Nobuyuki Ikeda and Takehiko Ed. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam, 1981.

[24] Jean Jacod and Albert Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2003.

[25] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Springer, New York, 2010.

[26] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[27] Hiroshi Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1997.

[28] Andreas Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin, 2006.

[29] Terry J. Lyons. *Differential Equations Driven by Rough Paths, Ecole d'Eté de Probabilités de Saint-Flour XXXIV*. Springer, 2004.

[30] Zhi-Ming Ma and Michael Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Springer, Berlin, 1992.

[31] Dilip B. Madan and Eugene Seneta. The variance gamma model for share market returns. *Journal of Business*, 63(4):511–524, 1990.

[32] Paul Malliavin. *Stochastic Analysis*. Springer, Berlin, 2002.

[33] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer-Verlag, Berlin, 2004.

[34] David Nualart. *The Malliavin Calculus and Related Topics*. Springer, Berlin, 2006.

[35] Giuseppe Da Prato. *Introduction to Stochastic Analysis and Malliavin Calculus*. Edizioni della Normale, Pisa, 2009.

[36] Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, 2005.

[37] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer, Berlin, 2005.

[38] L. C. G. Rogers and David Williams. *Diffusions, Markov processes and martingales, Vol. 2: Ito calculus*. Cambridge UP, Camebridge, 2008.

[39] Steven E. Shreve. *Stochastic calculus for finance. II*. Springer Finance. Springer-Verlag, New York, 2004. Continuous-time models.

[40] Daniel W. Stroock and S.R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, Berlin, 2005.

[41] Shinzo Watanabe. *Lectures on stochastic differential equations and Malliavin calculus*. Springer, Berlin, 1984.