Stochastic Analysis

Andreas Eberle

January 28, 2013

Contents

Chapter 1

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 [3] and Bertoin [6].

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 .

1.1 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. [32, Ch.I Thm.30].

Remark (Lévy processes in discrete time are Random Walks). A discrete-time process $(X_n)_{n=0,1,2,\ldots}$ 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 $Y_1, \ldots, Y_n$ such that $X_{s+t} - X_s$ has the same distribution as

$\sum_{i=1}^{n} Y_i$. 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. [3]. The simplest examples of infinitely divisible distributions are normal and Poisson distributions.

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 $k2^{-n} \le T < (k+1)2^{-n}$. 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

$$E\left[f(X_{T_n+t_1} - X_{T_n}, \ldots, X_{T_n+t_m} - X_{T_n}) I_A\right] = E\left[f(X_{t_1}, \ldots, X_{t_m})\right] \; P[A].$$

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

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

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.

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 $\{0, 1, 2, \ldots\}$ 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

$$N_t \quad = \quad \sum_{n=1}^{\infty} I_{\{S_n \leq t\}} \quad = \quad \sharp\,\{n \in \mathbb{N} \,:\, S_n \leq t\} \tag{1.1}$$

where $S_n = T_1 + T_2 + \cdots + T_n$ with independent random variables $T_i \sim \mathrm{Exp}(\lambda)$. 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. [12, Satz 10.12]. Note that by (1.1), the sample paths $t \mapsto N_t(\omega)$ are càdlàg. In general, any Lévy process with

$$X_t - X_s \quad \sim \quad \mathrm{Poisson}\,(\lambda(t-s)) \qquad \text{for any } 0 \leq s \leq t$$

is called a **Poisson process with intensity $\boldsymbol{\lambda}$**, and can be represented as above. The 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.

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 $\lambda > 0$.

For a different point of view on Poisson processes let

$$\mathcal{M}_c^+(S) = \left\{ \sum \delta_{y_i} \;\middle|\; (y_i) \text{ finite or countable sequence in } S \right\}$$

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 $\nu/\nu(S)$:.

$$N = \sum_{j=1}^{K} \delta_{X_j} \quad \text{with } X_j \sim \nu/\nu(s) \text{ i.i.d.}, \quad K \sim \mathrm{Poisson}(\nu(S)) \text{ independent.}$$

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*.

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 $\beta \in \mathbb{R}^{n\times d'}$. For example, linear combinations of independent Brownian motions and Poisson processes are again Lévy 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

$$T_s \quad = \quad \inf\{t \geq 0 \,:\, B_t = s\}$$

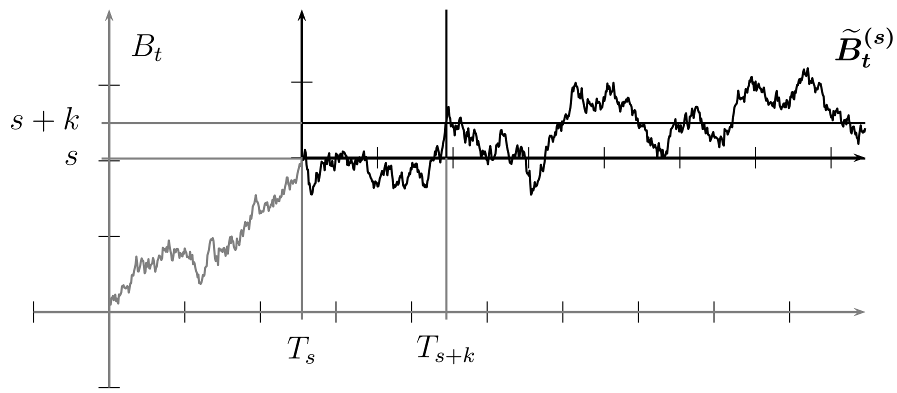
denote the first passage time to a level $s \in \mathbb{R}$. Then $(T_s)_{s\geq 0}$ is an increasing stochastic process that is adapted w.r.t. the filtration $(\mathcal{F}_{T_s})_{s\geq 0}$. For any ω, $s \mapsto T_s(\omega)$ is the generalized right inverse of the Brownian path $t \mapsto B_t(\omega)$. Moreover, by the strong Markov property, the process

$$\widetilde{B}_t^{(s)} \quad := \quad B_{T_s+t} - B_t \qquad , \; t \geq 0,$$

is a Brownian motion independent of $\mathcal{F}_{T_s}$ for any $s \geq 0$, and

$$T_{s+u} \quad = \quad T_s + \widetilde{T}_u^{(s)} \qquad \text{for } s, u \geq 0, \tag{1.2}$$

where $\widetilde{T}_u^{(s)} = \inf\left\{t \geq 0 \,:\, \widetilde{B}_t^{(s)} = u\right\}$ is the first passage time to u for the process $\widetilde{B}^{(s)}$.



By (1.2), 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

$$T_{s+} \;=\; \lim_{\varepsilon\downarrow 0} T_{s+\varepsilon} \;=\; \inf\left\{t \geq 0 \;:\; \widetilde{B}_t^{(s)} > u\right\}$$

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.,

$$k^{-1/\alpha}X_{kt} \quad \sim \quad X_t \qquad \text{for any } k \in (0,\infty) \text{ and } t \geq 0. \tag{1.3}$$

Definition. *Let $\alpha \in (0, 2]$. A Lévy process (X_t) satisfying (1.3) 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 1.2 below. There is a broader class of Lévy processes that is called α-stable in the literature, cf. e.g. [25]. 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 $\sigma \in [0, \infty)$. Other examples of stable processes will be considered below.

Characteristic exponents

From now on we 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 $\varphi_t(p) = E[\exp(ip \cdot X_t)]$:

Theorem 1.1 (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*

$$E[e^{ip \cdot X_t}] \quad = \quad e^{-t\psi(p)} \qquad \text{for any } t \geq 0 \text{ and } p \in \mathbb{R}^d. \tag{1.4}$$

Moreover, if (X_t) has finite first or second moments, then ψ is $\mathcal{C}^1$, $\mathcal{C}^2$ respectively, and

$$E[X_t] \quad = \quad it\nabla\psi(0) \qquad , \qquad \mathrm{Cov}[X_t^k, X_t^l] \quad = \quad t\,\frac{\partial^2 \psi}{\partial p_k \partial p_l}(0) \tag{1.5}$$

for any $k, l = 1, \ldots, d$ and $t \geq 0$.

Proof. Stationarity and independence of the increments implies the identity

$$\begin{aligned} \varphi_{t+s}(p) &= E[\exp(ip \cdot X_{t+s})] = E[\exp(ip \cdot X_s)] \cdot E[\exp(ip \cdot (X_{t+s} - X_s))] \\ &= \varphi_t(p) \cdot \varphi_s(p) \end{aligned} \tag{1.6}$$

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 $t \mapsto \varphi_t(p)$ is right-continuous. Since

$$\varphi_{t-\varepsilon}(p) \quad = \quad E[\exp(ip \cdot (X_t - X_\varepsilon))],$$

the function $t \mapsto \varphi_t(p)$ is also left continuous, and hence continuous. By (1.6) and since $\varphi_0(p) = 1$, we can now conclude that for each $p \in \mathbb{R}^d$, there exists $\psi(p) \in \mathbb{C}$ such that (1.4) 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 (1.5) for the first and second moment now follow by computing the derivatives w.r.t. p at $p = 0$ in (1.4). □

The function ψ is called the **characteristic exponent** of the Lévy process.

Examples. 1) For the Gaussian Lévy processes considered above,

$$\psi(p) \quad = \quad \frac{1}{2}|\sigma^T p|^2 - ib \cdot p \quad = \quad \frac{1}{2} p \cdot ap - ib \cdot p \qquad \text{with } a = \sigma\sigma^T.$$

2) The characteristic exponent of a Poisson process with intensity λ is

$$\psi(p) \quad = \quad \lambda(1 - e^{ip}).$$

3) In the superposition example above,

$$\psi_{\alpha X + \beta X'}(p) \quad = \quad \psi_X(\alpha^T p) + \psi_Y(\beta^T p).$$

Characteristic exponents can be applied to classify all α-stable processes:

Theorem 1.2 (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

$$\psi(p) \quad = \quad \sigma^\alpha |p|^\alpha (1 + i\mu \operatorname{sgn}(p)).$$

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 $\varphi_t(p) = \exp(-t\psi(p))$ 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 $\boldsymbol{\alpha}$-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*. It can be shown by Fourier transformation 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.

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 1.1 we obtain:

Corollary 1.3. *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 $0 \le s \le t$,

$$E[X_t - X_s \mid \mathcal{F}_s] \quad = \quad E[X_t - X_s] \quad = \quad i(t-s)\psi'(0)$$

by independence and stationarity of the increments and by (1.5). 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 1.3 (ii) shows that an integrable Lévy process is a *semimartingale* with martingale part M_t and continuous finite variation part bt. The identity (1.4) can be used to classify all Lévy processes, c.f. e.g. [3]. In particular, we will prove below that by Corollary 1.3, 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 $b \in \mathbb{R}^d$.

From now on, we will focus on discontinuous Lévy processes.

Compound Poisson processes

Compound Poisson processes are pure jump Lévy processes, i.e., the paths are constant apart from a finite number of jumps in finite time. A compound Poisson process is a continuous time Random Walk defined by

$$X_t = \sum_{j=1}^{N_t} \eta_j \qquad , \quad t \ge 0,$$

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. A compound Poisson process 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 1.4. *A compound Poisson process is a Lévy process with characteristic exponent*

$$\psi(p) = \int (1 - e^{ip\cdot y})\,\nu(dy). \tag{1.7}$$

Proof. Let $0 = t_0 < t_1 < \cdots < t_n$. Then the increments

$$X_{t_k} - X_{t_{k-1}} = \sum_{j=N_{t_{k-1}}+1}^{N_{t_k}} \eta_j \quad , \quad k = 1, 2, \ldots, n\,, \tag{1.8}$$

are conditionally independent given the σ-algebra generated by the Poisson process $(N_t)_{t\geq 0}$. Therefore, for $p_1, \ldots, p_n \ \in \mathbb{R}^d$,

$$\begin{aligned} E\big[\exp\big(i\sum_{k=1}^{n} p_k\cdot(X_{t_k} - X_{t_{k-1}}\big) \mid (N_t)\big] &= \prod_{k=1}^{n} E[\exp(ip_k\cdot(X_{t_k} - X_{t_{k-1}}) \mid (N_t)] \\ &= \prod_{k=1}^{n} \varphi(p_k)^{N_{t_k} - N_{t_{k-1}}}, \end{aligned}$$

where φ denotes the characteristic function of the jump sizes η_j. By taking the expectation value on both sides, we see that the increments in (1.8) are independent and stationary, since the same holds for the Poisson process (N_t). Moreover, by a similar computation,

$$\begin{aligned} E[\exp(ip\cdot X_t)] &= E[E[\exp(ip\cdot X_t) \mid (N_s)]] = E[\varphi(p)^{N_t}] \\ &= e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}\varphi(p)^k = e^{\lambda t(\varphi(p)-1)} \end{aligned}$$

for any $p \in \mathbb{R}^d$, which proves (1.7). □

The paths of a compound Poisson process are again 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 1.13 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

$$\psi_\nu(p) = \int (1 - e^{ip\cdot y})\nu(dy) \quad , \quad p \in \mathbb{R}^d. \tag{1.9}$$

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 1.5. *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 $(\Omega, \mathcal{A}, P)$. 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,*

$$X_t \quad = \quad \sum_{j=1}^{N_t} \eta_j \quad . \tag{1.10}$$

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 (1.10) 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. □

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 (1.9).*

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.

Superpositions and subdivisions of Poisson processes

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 lemma:

Theorem 1.6. *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 $q_k = P[C_1 = k]$.

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 1.4, the characteristic function of $\vec{N}_t$ for $t \geq 0$ is given by
$$E\left[\exp\left(ip\cdot\vec{N}_t\right)\right] = \exp\left(\lambda t(\varphi(p) - 1)\right), \quad p \in \mathbb{R}^K,$$

where

$$\varphi(p) = E\left[\exp(ip \cdot \eta_1)\right] = E\left[\exp\left(i \sum_{k \in K} p_k I_{\{k\}}(C_1)\right)\right] = \sum_{k \in K} q_k e^{ip_k}.$$

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 $q_k \lambda$. □

1.2 Poisson point processes

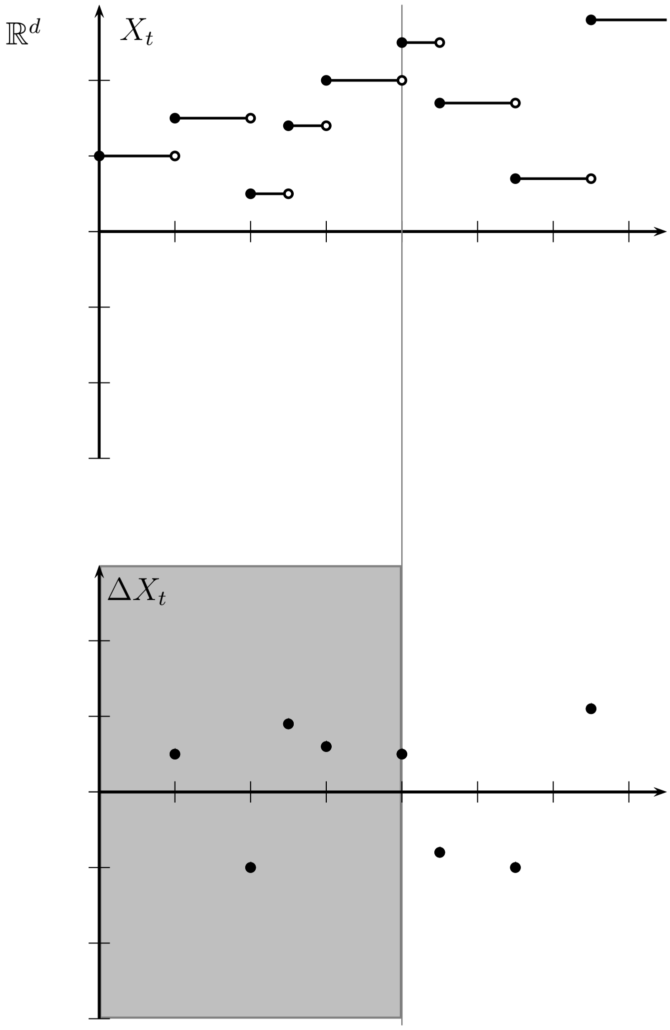
A compensated Poisson process has only finitely many jumps in a finite time interval. General Lévy jump processes may have a countably infinite number of (small) jumps in finite time. In the next section, we will construct such processes from their jumps. As a preparation we will now study Poisson point 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.

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 $N_t : \Omega \to \mathcal{M}_c^+(\mathbb{R}^d \setminus \{0\})$,

$$N_t(dy) = \sum_{\substack{s \leq t \\ \Delta X_s \neq 0}} \delta_{\Delta X_s}(dy), \quad t \geq 0,$$

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

$$N(dt\, dy) = \sum_{\substack{s \leq t \\ \Delta X_s \neq 0}} \delta_{(s, \Delta X_s)}(dt\, dy).$$



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 1.1, 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*

$$\mathbb{E}[N_t(B)] \;=\; t \cdot \nu(B) \qquad \forall\, t \geq 0,\; B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}). \tag{1.11}$$

It is elementary to verify that for any Lévy process, there is a unique measure ν satisfying (1.11). 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

$$\nu(dy) \quad = \quad \big(c_+ I_{(0,\infty)}(y) + c_- I_{(-\infty,0)}(y)\big)\; |y|^{-1-\alpha}\, dy \tag{1.12}$$

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.

Definition and construction of 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 $(\Omega, \mathcal{A}, P)$ is called a* ***Poisson point process of intensity ν*** *if and only if*

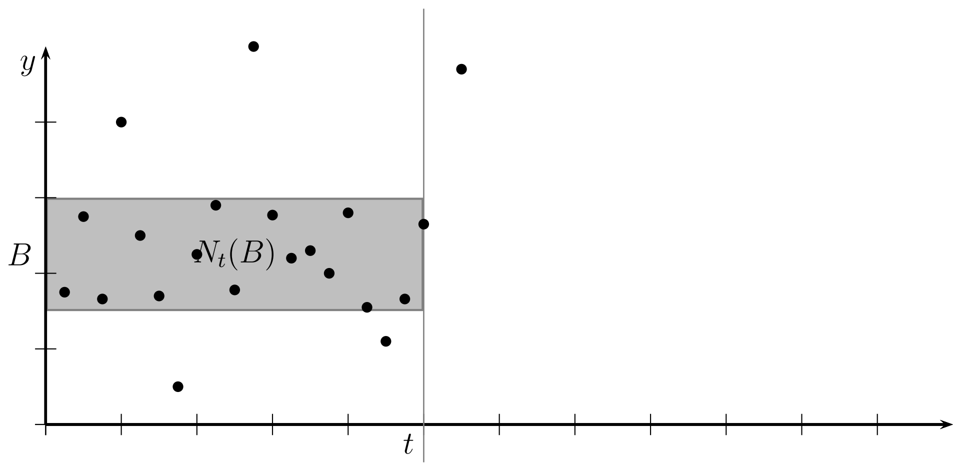
(i) $B \mapsto N_t(B)(\omega)$ *is a counting measure on S for any $t \geq 0$ and $\omega \in \Omega$,*

(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 1.7 (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 1.7, 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 1.5.

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 1.8. *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 1.7 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. $\square$

As a direct consequence of the theorem and the properties of compound Poisson processes derived above, we obtain:

Corollary 1.9 (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 1.9 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 $L^1(\Omega, \mathcal{A}, P)$.

b) Proceeding similarly as in a), prove the identities

$$\begin{aligned} E[N_t^f] \quad &= \quad t\int f\ d\nu && \text{for any } f \in \mathcal{L}^1(\nu), \\ \mathrm{Cov}[N_t^f, N_t^g] \quad &= \quad t\int fg\ d\nu && \text{for any } f, g \in \mathcal{L}^1(\nu) \cap \mathcal{L}^2(\nu), \\ E[\exp(ipN_t^f)] \quad &= \quad \exp(t\int (e^{ipf} - 1)\ d\nu) && \text{for any } f \in \mathcal{L}^1(\nu). \end{aligned}$$

c) Show that the processes considered in Corollary 1.9 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 1.9 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*.

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

$$(G_\bullet N)_t \quad = \quad \int_{(0,t] \times S} G_s(y)\; N(ds\, dy), \tag{1.13}$$

$$(G_\bullet \widetilde{N})_t \quad = \quad \int_{(0,t] \times S} G_s(y)\; \widetilde{N}(ds\, dy) \tag{1.14}$$

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 $G.N$ is defined in an elementary way as

$$(G.N)_t \;=\; \sum_{(s,y) \in \text{supp}(N),\; s \leq t} G_s(y).$$

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. [5].

We denote by $\mathcal{E}$ the vector space consisting of all **elementary predictable processes** G of the form

$$G_t(y)(\omega) \quad = \quad \sum_{i=0}^{n-1} \sum_{k=1}^{m} Z_{i,k}(\omega)\, I_{(t_i, t_{i+1}]}(t)\, I_{B_k}(y) \tag{1.15}$$

with $m, n \in \mathbb{N}$, $0 \leq t_0 < t_1 < \cdots < t_n$, $B_1, \ldots, B_m \in \mathcal{S}$ 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

$$(G_\bullet N)_t \quad = \quad \sum_{i=0}^{n-1} \sum_{k=1}^{m} Z_{i,k} \left(N_{t_{i+1} \wedge t}(B_k) - N_{t_i \wedge t}(B_k) \right), \tag{1.16}$$

Notice that the summands vanish for $t_i \geq t$ 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 1.10 (Elementary properties of stochastic integrals w.r.t. PPP). *Let $G \in \mathcal{E}$. Then the following assertions hold:*

1) For any $t \geq 0$,

$$E\left[(G_\bullet N)_t\right] \;=\; E\left[\int_{(0,t] \times S} G_s(y)\, ds\, \nu(dy)\right].$$

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 $t \geq 0$, $G_\bullet \widetilde{N}$ 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 $0 \leq s \leq t$ 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 $t_i \geq t$, 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 (1.13) and (1.14) are well-defined Lebesgue integrals for every ω. By Lemma 1.10 and monotone convergence, the identity

$$E\left[(G_\bullet N)_t\right] \;=\; E\left[\int_{(0,t]\times S} G_s(y)\, ds\, \nu(dy)\right] \tag{1.17}$$

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 (1.17),

$$E\left[\int_{(0,u]\times S} |G_s(y)|\, N(ds\, dy)\right] \;=\; E\left[\int_{(0,u]\times S} |G_s(y)|\, ds\, \nu(dy)\right] \;<\; \infty.$$

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 (1.17).

Theorem 1.11. *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 (1.17).

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

$$V_t^{(1)}(G_\bullet N) \;\le\; \int_{(0,t]\times S} |G_s(y)|\, N(ds\, dy).$$

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 1.10.

3) We may assume w.l.o.g. $G \ge 0$, 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 $[0,u]$,

$$\begin{aligned} \sum_{r\in\pi} |(G_\bullet N)_{r'} - (G_\bullet N)_r| &= \sum_{r\in\pi} \left| \int_{(r,r']\times S} G_s(y)\, N(ds\, dy) \right| \\ &\le \int_{(0,u]\times S} |G_s(y)|\, N(ds\, dy) \;<\; \infty \qquad \text{a.s.} \end{aligned}$$

□

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 (1.17) 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

$$\mathcal{M}_d^2([0,u]) \quad = \quad \big\{ M \in \mathcal{M}^2([0,u]) \mid t \mapsto M_t(\omega) \text{ càdlàg for any } \omega \in \Omega \big\}$$

denote the space of all square-integrable càdlàg martingales w.r.t. the completed filtration $(\mathcal{F}_t^P)$. 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 $M_d^2([0,u])$ of equivalence classes of indistinguishable martingales in $\mathcal{M}_d^2([0,u])$ 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 1.10, 3), shows that for elementary predictable processes G,

$$||G_\bullet \widetilde{N}||_{M^2([0,u])} \quad = \quad ||G||_{L^2(P\otimes\lambda_{(0,u)}\otimes\nu)}. \tag{1.18}$$

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 $G_\bullet \widetilde{N} \in M_d^2(0,u)$ for any predictable $G \in L^2(P \otimes \lambda_{(0,u)} \otimes \nu)$ in such a way that

$$G^{(k)}_\bullet \widetilde{N} \quad \longrightarrow \quad G_\bullet \widetilde{N} \ \text{ in } \ M^2 \qquad \text{whenever} \quad G^{(k)} \to G \ \text{ in } \ L^2.$$

Theorem 1.12 (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 $M_d^2([0,u])$ such that $G_\bullet \widetilde{N}$ is given by (1.16) for any elementary predictable process G of the form (1.15).*

Proof. As pointed out above, by (1.18), 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 [3] 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 $\nu \circ h^{-1}$.

1.3 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.,

$$\nu(|y| > \varepsilon) \quad < \quad \infty \quad \text{for any } \varepsilon > 0, \text{ and} \tag{1.19}$$

$$\int_{|y| \le 1} |y|^2\, \nu(dy) \quad < \quad \infty. \tag{1.20}$$

Note that the condition (1.19) is necessary for the existence of a Lévy process with jump intensity ν. Indeed, if (1.19) 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 (1.20) 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 $N_t(dy)$, $t \geq 0$, 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 (1.19). 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 (1.19) and (1.20), we are able to construct a Lévy process with jump intensity measure ν that is given by

$$\widetilde{X}_t^r \;=\; \int_{|y|>r} y\,N_t(dy) \;+\; \int_{|y|\le r} y\,\widetilde{N}_t(dy). \tag{1.21}$$

for any $r\in(0,\infty)$. Indeed, let

$$X_t^r \;:=\; \int_{|y|>r} y\;N_t(dy) \;=\; \int_{(0,t]\times\mathbb{R}^d} y\,I_{\{|y|>r\}}\;N(ds\,dy), \qquad \text{and} \tag{1.22}$$

$$M_t^{\varepsilon,r} \;:=\; \int_{\varepsilon<|y|\le r} y\,\widetilde{N}_t(dy). \tag{1.23}$$

for $\varepsilon, r\in[0,\infty)$ with $\varepsilon<r$. As a consequence of the Itô isometry for Poisson point processes, we obtain:

Theorem 1.13 (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*

$$\psi_r(p) \quad=\quad \int_{|y|\le r} (1-e^{ip\cdot y}+ip\cdot y)\,\nu(dy) \qquad \forall\,p\in\mathbb{R}^d. \tag{1.24}$$

   Moreover, for any $u\in(0,\infty)$,

$$E\left[\sup_{0\le t\le u}|\,M_t^{\varepsilon,r}-M_t^{0,r}|^2\right] \quad\to\quad 0 \qquad \text{as } \varepsilon\downarrow 0. \tag{1.25}$$

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*

$$\widetilde{\psi}_r(p) \quad=\quad \int\left(1-e^{ip\cdot y}+ip\cdot yI_{\{|y|\le r\}}\right)\nu(dy) \qquad \forall\,p\in\mathbb{R}^d. \tag{1.26}$$

Proof. 1) is a consequence of Theorem 1.8.

2) By (1.20), 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 1.8, $(M_t^{\varepsilon,r})$ is a compensated compound Poisson process with intensity $I_{\{\varepsilon<|y|\le r\}}\,\nu(dy)$ and characteristic exponent

$$\psi_{\varepsilon,r}(p) \;=\; \int_{\varepsilon<|y|\le r} (1 - e^{ip\cdot y} + ip\cdot y)\, \nu(dy).$$

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 $I_{\{|y|\le r\}}\, N_t(dy)$ 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 $0 < \varepsilon < r$,

$$\widetilde{X}_t^\varepsilon \;=\; \widetilde{X}_t^r + bt, \qquad \text{where } b = \int_{\varepsilon<|y|\le r} y\, \nu(dy).$$

A totally uncompensated Lévy process

$$X_t \;=\; \lim_{n\to\infty} \int_{|y|\ge 1/n} y\, N_t(dy)$$

does exist only under additional assumptions on the jump intensity measure:

Corollary 1.14 (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*

$$\psi(p) \quad = \quad \lim_{\varepsilon \downarrow 0} \int_{|y|>\varepsilon} \left(1 - e^{ip\cdot y}\right)\ \nu(dy) \qquad \forall\, p \in \mathbb{R}^d \tag{1.27}$$

such that

$$E\left[\sup_{0\le t\le u} |\, X_t - X_t^\varepsilon|^2\right] \ \to\ 0 \qquad \text{as } \varepsilon \downarrow 0. \tag{1.28}$$

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 (1.28) as $\varepsilon \downarrow 0$. The limit process is again a Lévy process, and, by dominated convergence, the characteristic exponent is given by (1.27). □

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 1.16 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

$$\int (1 \wedge |y|^2)\, \nu(dy) \quad < \quad \infty \qquad . \tag{1.29}$$

Note that (1.29) 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

$$X_t = \sigma B_t + bt + \int_{|y|>1} y\; N_t(dy) + \int_{|y|\leq 1} y\;\; (N_t(dy) - t\nu(dy))\,. \tag{1.30}$$

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

$$\psi(p) \quad = \quad \frac{1}{2}\, p\cdot ap - ib\cdot p + \int_{\mathbb{R}^d\setminus\{0\}} (1 - e^{ip\cdot y} + ip\cdot y\; I_{\{|y|\leq 1\}})\; \nu(dy). \tag{1.31}$$

We have thus proved the first part of the following theorem:

Theorem 1.15 (Lévy-Itô decomposition).

1) *The expression (1.30) defines a Lévy process with characteristic exponent ψ.*
2) *Conversely, any Lévy process (X_t) can be decomposed as in (1.30) with the Poisson point process*

$$N_t \quad = \quad \sum_{\substack{s\leq t\\ \Delta X_s \neq 0}} \Delta X_s \qquad , \quad t \geq 0, \tag{1.32}$$

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 (1.29).*

We will not prove the second part of the theorem here. The principal way to proceed is to define (N_t) via (1.29), and to consider the difference of (X_t) and the integrals w.r.t. (N_t) on the right hand side of (1.30). 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 4.1 below. Carrying out the details of this argument, however, is still a lot of work. A detailed proof is given in [3].

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 1.16 (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 (1.31) 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 1.15.
(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. [16]. □

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

$$E[\exp(-u\Gamma_t)] \;=\; \left(1 + \frac{u}{\lambda}\right)^{-at} \;=\; \exp\left(-t \int_0^\infty (1 - e^{-uxy}) a y^{-1} e^{-\lambda y}\, dy\right) \qquad (1.33)$$

for every $u \geq 0$, cf. e.g. [25, Lemma 1.7]. Since $\Gamma_t \geq 0$, both sides in (1.33) have a unique analytic extension to $\{u \in \mathbb{C} \ : \ \Re(u) \geq 0\}$. Therefore, we can replace u by $-ip$ in (1.33) to conclude that the characteristic exponent of (Γ_t) is

$$\psi(p) \quad = \quad \int_0^\infty (1 - e^{ipy})\, \nu(dy), \qquad \text{where } \nu(dy) = ay^{-1}e^{-\lambda y}\, dy.$$

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

$$X_t \quad = \quad \sigma B_{\Gamma_t} + b\Gamma_t$$

is called a **Variance Gamma process**. It is a Lévy process with characteristic exponent $\psi(p) = \int (1 - e^{ipy})\, \nu(dy)$, where

$$\nu(dy) \quad = \quad c|y|^{-1} \left(e^{-\lambda y} I_{(0,\infty)}(y) + e^{-\mu|y|} I_{(-\infty,0)}(y) \right)\, dy$$

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 [28] 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 [4]. 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 (1.12) that the jump intensity measure of a strictly α-stable process in $\mathbb{R}^1$ is given by

$$\nu(dy) \quad = \quad \big(c_+ I_{(0,\infty)}(y) + c_- I_{(-\infty,0)}(y)\big)\;|y|^{-1-\alpha}\,dy \qquad (1.34)$$

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 (1.34). 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

$$X_t^\varepsilon \quad = \quad \int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} y \; N_t(dy)$$

where (N_t) is a Poisson point process with intensity measure ν. Setting $||X||_u = E[\sup_{t\le a}|X_t|^2]^{1/2}$, an application of Theorem 1.13 yields:

Corollary 1.17 (Construction of $\boldsymbol{\alpha}$-stable processes). *Let ν be the probability measure on $\mathbb{R}\setminus\{0\}$ defined by (1.34) 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 $\psi(p) = z\,|p|^\alpha$, $z = \int\big(1-e^{iy}\big)\,\nu(dy) \;\in \mathbb{C}$.

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 $\psi(p) = \widetilde{z} \cdot |p|^\alpha$, $\widetilde{z} = \int (1 - e^{iy} + iy)\, \nu(dy)$.*

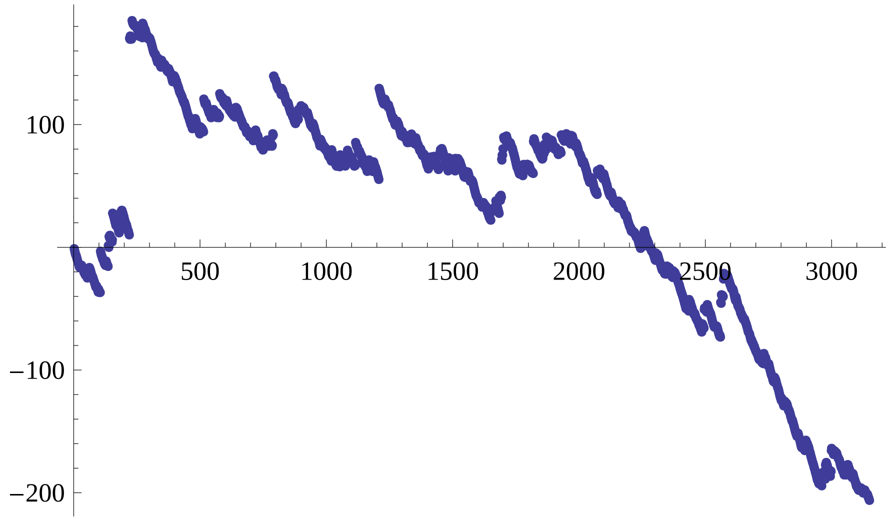
Proof. By Theorem 1.13 it is sufficient to prove convergence of the characteristic exponents

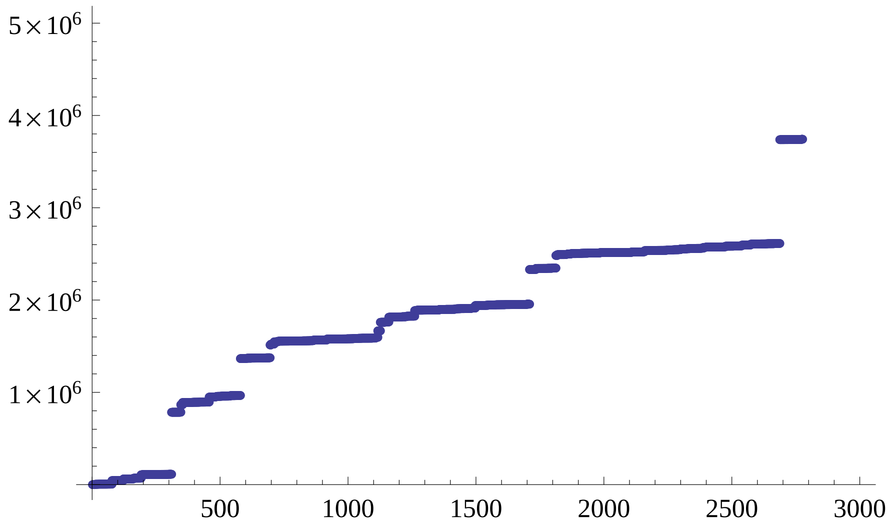
$$\begin{aligned} \psi_\varepsilon(p) &= \int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} \left(1 - e^{ipy}\right) \nu(dy) &= |p|^\alpha \int_{\mathbb{R}\setminus[-\varepsilon p,\varepsilon p]} \left(1 - e^{ix}\right) \nu(dx), \\ \widetilde{\psi}_\varepsilon(p) &= \int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} \left(1 - e^{ipy} + ipy\right) \nu(dy) &= |p|^\alpha \int_{\mathbb{R}\setminus[-\varepsilon p,\varepsilon p]} \left(1 - e^{ix} + ix\right) \nu(dx) \end{aligned}$$

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 $1 - e^{ix} + ix = \mathcal{O}(|x|^2)$. $\square$

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.


100
−100
−200
500
1000
1500
2000
2500
3000


5×10^6
4×10^6
3×10^6
2×10^6
1×10^6
500
1000
1500
2000
2500
3000

Chapter 2

Stochastic integrals and Itô calculus for semimartingales

[Stochastic calculus for semimartingales] 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. [33, Ch.II, §2] and [32, 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

$$X_n = X_0 + M_n + A_n^{\nearrow} - A_n^{\searrow} \tag{2.1}$$

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. [13, 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 $k \in \mathbb{N}$,

$$\begin{aligned} F(X_k) - F(X_{k-1}) \quad &= \quad \int_0^1 F'(X_{k-1} + s\Delta X_k) \, ds \;\; \Delta X_k \\ &= \quad F'(X_{k-1}) \, \Delta X_k \; + \int_0^1 \int_0^s F''(X_{k-1} + r\Delta X_k) \, dr \, ds \;\; (\Delta X_k)^2 \, . \end{aligned}$$

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 2.22 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*

$$X_t \quad = \quad X_0 + M_t + A_t, \qquad\qquad t \geq 0, \tag{2.2}$$

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. [32]). 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 (2.2) is *not unique*, see the example below. Uniqueness holds if, in addition, (A_t) is assumed to be predictable, cf. [5, 32]. Under the extra assumption that (A_t) is continuous, uniqueness is a consequence of Corollary 2.15 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

$$X_t = \int_0^t H_s \cdot dB_s + \int_{(0,t]\times S} G_s(y)\widetilde{N}(ds\, dy) + \int_0^t K_s\, ds + \int_{(0,t]\times S} L_s(y) N(ds\, dy) \quad (2.3)$$

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 2.4 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

$$f(X_t) = \text{local martingale } + \int_0^t (\mathcal{L}f)(X_s)\, ds, \tag{2.4}$$

cf. e.g. [11] or [15]. Indeed, it is possible to define the generator $\mathcal{L}$ of a Markov process through a solution to a martingale problem as in (2.4).

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 2.2, we will consider integrals and calculus w.r.t. finite variation processes in Section 2.1.

2.1 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.,

$$\mu_A\left((s,t]\right) \quad = \quad A_t - A_s \qquad \text{for any } 0 \le s \le t < u. \tag{2.5}$$

The function A has the decomposition

$$A_t \quad = \quad A_t^c + A_t^d \tag{2.6}$$

into the pure jump function

$$A_t^d \quad := \quad \sum_{s \le t} \Delta A_s \tag{2.7}$$

and the continuous function $A_t^c = A_t - A_t^d$. Indeed, the series in (2.7) 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 $V_t^{(1)}(A)$. 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

$$\int_s^u G_r\, dA_r \quad = \quad \int G_r\, I_{(s,t]}(r)\, \mu_A(dr) \qquad \text{for } 0 \le s \le t < u.$$

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

$$V_t^{(1)}(I) \quad = \quad \int_0^t |G_r|\, |\mu_A|(dr) \quad = \quad \int_0^t |G_r|\, dV_r^{(1)}(A).$$

2) I decomposes into the continuous and pure jump parts

$$I_t^c \quad = \quad \int_0^t G_r\, dA_r^c \quad , \qquad I_t^d \quad = \quad \int_0^t G_r\, dA_r^d \quad = \quad \sum_{s \le t} G_s\, \Delta A_s.$$

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 2.1 (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 2.1 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

$$[H, A]_t \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n\\ s<t}} (H_{s'\wedge t} - H_s)(A_{s'\wedge t} - A_s), \tag{2.8}$$

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 2.2. *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 2.3 (Integration by parts, product rule). *Suppose that $H,A : [0,u)\to\mathbb{R}$ are right continuous functions of finite variation. Then*

$$H_tA_t - H_0A_0 \;=\; \int_0^t H_{r-}\, dA_r + \int_0^t A_{r-} dH_r + [H,A]_t \qquad \textit{for any } t\in[0,u). \tag{2.9}$$

In particular, the covariation $[H,A]$ is a càdlàg function of finite variation, and for $a<u$, the approximations in (2.8) converge uniformly on $[0,a]$ w.r.t. any sequence (π_n) such that $\operatorname{mesh}(\pi_n)\to 0$.

In differential notation, (2.9) 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, (2.9) implies

$$H_tA_t - H_0A_0 = \sum_{r\le t} \Delta(HA)_r.$$

Note that this statement is not completely trivial, as it holds even when the jump times of HA form a countable dense subset of $[0,t]$!

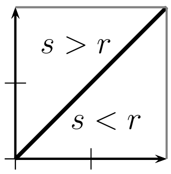
Since the product rule is crucial but easy to prove, we give two proofs of Theorem 2.3:

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 $n \to \infty$, (2.9) follows by Theorem 2.1 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 2.1. □

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 2.4 (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 $t \in [0,u)$,*

$$F(A_t) - F(A_0) = \int_0^t F'(A_{s-})\, dA_s + \sum_{s \le t} \left(F(A_s) - F(A_{s-}) - F'(A_{s-})\Delta A_s \right), \tag{2.10}$$

or, equivalently,

$$F(A_t) - F(A_0) = \int_0^t F'(A_{s-})\, dA_s^c + \sum_{s \le t} \left(F(A_s) - F(A_{s-}) \right). \tag{2.11}$$

If A is continuous then $F(A)$ is also continuous, and (2.10) reduces to the standard chain rule

$$F(A_t) - F(A_0) = \int_0^t F'(A_s)\, dA_s.$$

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

$$F(A_t) - F(A_0) = \sum_{s \le t} \left(F(A_s) - F(A_{s-}) \right).$$

Remark. Note that by Taylor's theorem, the sum in (2.10) 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 2.22 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 (2.10). 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 (2.11),

$$\begin{aligned}(FG)(A_t)-&(FG)(A_0)\\ &= \int_0^t F(A_{s-})\,dG(A)_s + \int_0^t G(A_{s-})\,dF(A)_s + \sum_{s\leq t} \Delta F(A)_s \Delta G(A)_s\\ &= \int_0^t (F(A_{s-})G'(A_{s-}) + G(A_{s-})F'(A_{s-}))\,dA_s^c\\ &\quad + \sum_{s\leq t} (F(A_{s-})\Delta G(A)_s + G(A_{s-})\Delta F(A)_s + \Delta F(A)_s \Delta G(A)_s)\\ &= \int_0^t (FG)'(A_{s-})\,dA_s^c + \sum_{s\leq t} ((FG)(A_s) - (FG)(A_{s-}))\end{aligned}$$

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 $\{A_s \mid s \leq t\}$. Since (2.11) holds for the polynomials p_n, it also holds for F. $\Box$

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

$$\begin{aligned} dZ_t &= Z_{t-}\,dA_t \quad , \qquad Z_0 = 1 \quad , \qquad \text{i.e.,}\\ Z_t &= 1 + \int_0^t Z_{s-}\,dA_s \qquad \text{for any } t \geq 0. \end{aligned} \tag{2.12}$$

If A is continuous then $Z_t = \exp(A_t)$ solves (2.12) 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 (2.12), since

$$Z_t = Z_0 + \sum_{s\leq t} \Delta Z_s = 1 + \sum_{s\leq t} Z_{s-}\Delta A_s = 1 + \int_{(0,t]} Z_{s-}\,dA_s.$$

In general, we obtain:

Theorem 2.5. *The unique càdlàg function solving (2.12) is*

$$Z_t \quad = \quad \exp(A^c_t) \cdot \prod_{s \le t} (1 + \Delta A_s), \tag{2.13}$$

where the product converges for any $t \ge 0$.

Proof. 1) We first show convergence of the product

$$P_t \quad = \quad \prod_{s \le t} (1 + \Delta A_s).$$

Recall that since A is càdlàg, there are only finitely many jumps with $|\Delta A_s| > 1/2$. Therefore, we can decompose

$$P_t \quad = \quad \exp \left(\sum_{\substack{s \le t \\ |\Delta A_s| \le 1/2}} \log(1 + \Delta A_s) \right) \cdot \prod_{\substack{s \le t \\ |\Delta A_s| > 1/2}} (1 + \Delta A_s) \tag{2.14}$$

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 $|x| \le 1/2$. The limit S_t of the series defines a pure jump function with variation $V^{(1)}_t(S) \le \text{const.} \cdot V^{(1)}_t(A)$ for any $t \ge 0$.

2) *Equation for P_t*: The chain and product rule now imply by (2.14) that $t \mapsto P_t$ is also a finite variation pure jump function. Therefore,

$$P_t \quad = \quad P_0 + \sum_{s \le t} \Delta P_s \quad = \quad 1 + \sum_{s \le t} P_{s-} \Delta A_s \quad = \quad 1 + \int_0^t P_{s-} \, dA^d_s, \quad \forall t \ge 0, \tag{2.15}$$

i.e., P is the exponential of the pure jump part $A^d_t = \sum_{s \le t} \Delta A_s$.

3) *Equation for Z_t*: Since $Z_t = \exp(A^c_t) P_t$ and $\exp(A^c)$ is continuous, the product rule and (2.15) imply

$$\begin{aligned} Z_t - 1 \quad &= \quad \int_0^t e^{A^c_s} \, dP_s + \int_0^t P_{s-} \, e^{A^c_s} \, dA^c_s \\ &= \quad \int_0^t e^{A^c_s} P_{s-} \, d(A^d + A^c)_s \quad = \quad \int_0^t Z_{s-} \, dA_s. \end{aligned}$$

4) *Uniqueness*: Suppose that $\widetilde{Z}$ is another càdlàg solution of (2.12), and let $X_t := Z_t - \widetilde{Z}_t$. Then X solves the equation

$$X_t \quad = \quad \int_0^t X_{s-}\, dA_s \quad \forall\, t \geq 0$$

with zero initial condition. Therefore,

$$|X_t| \quad \leq \quad \int_0^t |X_{s-}|\, dV_t \quad \leq \quad M_t V_t \qquad \forall\, t \geq 0,$$

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

$$|X_t| \quad \leq \quad M_t \int_0^t V_{s-}\, dV_s \quad \leq \quad M_t V_t^2/2$$

by the chain rule, and

$$|X_t| \quad \leq \quad \frac{M_t}{n!} \int_0^t V_{s-}^n\, dV_s \quad \leq \quad \frac{M_t}{(n+1)!} V_t^{n+1} \quad \forall\, t \geq 0, n \in \mathbb{N}. \tag{2.16}$$

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 (2.16) 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 (2.12) 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,

$$\begin{aligned}
\mathcal{E}_t^A \mathcal{E}_t^B - 1 &= \int_0^t \mathcal{E}_{s-}^A \, d\mathcal{E}_s^B + \int_0^t \mathcal{E}_{s-}^B \, d\mathcal{E}_s^A + \sum_{s \le t} \Delta \mathcal{E}_s^A \Delta \mathcal{E}_s^B \\
&= \int_0^t \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \, d(A+B)_s + \sum_{s \le t} \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \Delta A_s \Delta B_s \\
&= \int_0^t \mathcal{E}_{s-}^A \mathcal{E}_{s-}^B \, d(A+B+[A,B])_s
\end{aligned}$$

for any $t \ge 0$. This shows that in general, $\mathcal{E}^A \mathcal{E}^B \neq \mathcal{E}^{A+B}$.

Theorem 2.6. *If* $A, B : [0,\infty) \to \mathbb{R}$ *are right continuous with finite variation then*

$$\mathcal{E}^A \mathcal{E}^B = \mathcal{E}^{A+B+[A,B]}.$$

Proof. The left hand side solves the defining equation for the exponential on the right hand side. □

In particular, choosing $B = -A$, we obtain:

$$\frac{1}{\mathcal{E}^A} = \mathcal{E}^{-A+[A]}$$

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

$$dS_t = \sigma S_{t-} \, dN_t + \alpha S_t \, dt \tag{2.17}$$

w.r.t. a Poisson process (N_t) with intensity λ. Geometric Poisson processes are relevant for financial models, cf. e.g. [35]. The equation (2.17) can be interpreted pathwise as the Stieltjes integral equation

$$S_t = S_0 + \sigma \int_0^t S_{r-} \, dN_r + \alpha \int_0^t S_r dr \quad , \quad t \ge 0.$$

Defining $A_t = \sigma N_t + \alpha t$, (2.17) can be rewritten as the exponential equation

$$dS_t \quad = \quad S_{t-}\, dA_t \qquad ,$$

which has the unique solution

$$S_t \quad = \quad S_0 \cdot \mathcal{E}_t^A \quad = \quad S_0 \cdot e^{\alpha t} \prod_{s \le t} (1 + \sigma \Delta N_s) \quad = \quad S_0 \cdot e^{\alpha t} (1+\sigma)^{N_t}.$$

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 4 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 $\mathbb{R}^d \backslash \{0\}$. 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 4, 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

$$\begin{aligned} A_t &= (\lambda - \bar{\lambda})t + \sum_{\substack{s \le t \\ \Delta X_s \ne 0}} (\varrho(\Delta X_s) - 1) \\ &= -(\bar{\lambda} - \lambda)t + \int (\varrho(y) - 1)\; N_t(dy). \end{aligned} \tag{2.18}$$

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

$$A_t = \int (\varrho(y) - 1)\; \widetilde{N}_t(dy) \quad , \quad \text{where } \widetilde{N}_t := N_t - t\nu.$$

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

$$Z_t = 1 + \int_0^t Z_{s-}\; dA_s \tag{2.19}$$

in the equivalent form

$$Z_t = 1 + \int_{(0,t] \times \mathbb{R}^d} Z_{s-}\; (\varrho(y) - 1)\; \widetilde{N}(ds\; dy) \tag{2.20}$$

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 $t \ge 0$ and $B \in \mathcal{B}(\mathbb{R}^d)$. In differential notation, (2.20) is an SDE driven by the compensated Poisson point process $(\widetilde{N}_t)$:

$$dZ_t = \int_{y \in \mathbb{R}^d} Z_{t-}\; (\varrho(y) - 1)\; \widetilde{N}(dt\; dy).$$

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 [34] where more details and applications can be found.

Suppose that (X_t) on $(\Omega, \mathcal{A}, P)$ 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

$$M_t^{[f]} \quad = \quad f(X_t) - \int_0^t (\mathcal{L}f)(X_s)\,ds \quad , \quad t \geq 0, \tag{2.21}$$

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

$$\begin{aligned} E[M_t^{[f]} - M_s^{[f]} \mid \mathcal{F}_s^X] \quad &= \quad E[f(X_t) - f(X_s) - \int_s^t (\mathcal{L}f)(X_r)\,dr \mid \mathcal{F}_s] \\ &= \quad (p_{t-s}f)(X_s) - f(X_s) - \int_s^t (p_{r-s}\mathcal{L}f)(X_s)\,ds = 0 \end{aligned}$$

for any $0 \leq s \leq t$. In particular, choosing $f = I_{\{b\}}$ for $b \in S$, we see that

$$M_t^b \quad = \quad I_{\{b\}}(X_t) - \int_0^t \mathcal{L}(X_s, b)\,ds \tag{2.22}$$

is a martingale, and, in differential notation,

$$dI_{\{b\}}(X_t) \quad = \quad \mathcal{L}(X_t, b)\,dt + dM_t^b. \tag{2.23}$$

Next, we note that by the results in the next section, the stochastic integrals

$$N_t^{a,b} \quad = \quad \int_0^t I_{\{a\}}(X_{s-})\,dM_s^b \quad , \; t \geq 0,$$

are martingales for any $a, b \in S$. Explicitly, for any $a \neq b$,

$$\begin{aligned} N_t^{a,b} \quad = \quad & \sum_{s\leq t} I_{\{a\}}(X_{s-}) \left(I_{S\setminus\{b\}}(X_{s-}) I_{\{b\}}(X_s) - I_{\{b\}}(X_{s-}) I_{S\setminus\{b\}}(X_s) \right) \\ & - \int_0^t I_{\{a\}}(X_s)\,\mathcal{L}(X_s, b)\,ds \quad , \quad \text{i.e.,} \end{aligned}$$

$$N_t^{a,b} \quad = \quad J_t^{a,b} - \mathcal{L}(a,b)\; L_t^a \tag{2.24}$$

where $J_t^{a,b} = |\{s \le t \;:\; X_{s-} = a, X_s = b\}|$ is the number of jumps from a to b until time t, and

$$L_t^a \quad = \quad \int_0^t I_a(X_s)\; ds$$

is the amount of time spent at a before time t (*"local time at a"*). In the form of an SDE,

$$dJ_t^{a,b} \quad = \quad \mathcal{L}(a,b)\; dL_t^a + dN_t^{a,b} \quad \text{for any } a \neq b. \tag{2.25}$$

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 (2.24),

$$N_t^{[g]} \quad = \quad \sum_{s \le t} g(X_{s-}, X_s) - \int_0^t (\mathcal{L}g^T)(X_s, X_s)\; ds \tag{2.26}$$

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

$$\widetilde{\zeta}_t \;=\; \exp\left(-\int_0^t \widetilde{\lambda}(X_s)\, ds \right) \prod_{s \leq t : X_{s-} \neq X_s} \widetilde{\mathcal{L}}(X_{s-}, X_s),$$

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 $t \geq 0$. 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.

2.2 Stochastic integration for semimartingales

Throughout this section we fix a probability space $(\Omega, \mathcal{A}, P)$ 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 $(\mathcal{F}_t^P)$ martingale, and $H = (H_t)_{t \geq 0}$ is a uniformly bounded càdlàg $(\mathcal{F}_t^P)$ adapted process. In particular, the left limit process

$$H_- \quad := \quad (H_{t-})_{t \geq 0}$$

is left continuous with right limits and $(\mathcal{F}_t^P)$ adapted. For a partition π of $\mathbb{R}_+$ we consider the elementary processes

$$H_t^\pi \quad := \quad \sum_{s\in\pi} H_s\, I_{[s,s')}(t), \qquad \text{and} \qquad H_{t-}^\pi \quad = \quad \sum_{s\in\pi} H_s\, I_{(s,s']}(t).$$

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. [26], [17]). 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 $(\mathcal{F}_t^P)$ 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 $(\mathcal{F}_t^P)$ martingales defined up to a finite time a. The following fundamental theorem settles this question completely:

Theorem 2.7 (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*

$$||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \quad < \quad \gamma \tag{2.27}$$

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. [15]), or, alternatively, a lengthy discussion of general semimartingale theory (cf. [34]).

Proof of Theorem 2.7. Let $C \in (0,\infty)$ be a common uniform upper bound for the processes (H_t) and (M_t). To prove the estimate in (2.27), 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

$$I_a^\pi - I_a^{\widetilde{\pi}} \quad = \quad \sum_{s\in\pi}(H_s - H_{\lfloor s \rfloor})(M_{s'} - M_s) \tag{2.28}$$

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

$$\lfloor s \rfloor \quad := \quad \max\{t \in \widetilde{\pi} \;:\; t \leq s\}$$

is the next partition point of the rough partition $\widetilde{\pi}$ below s. Now fix $\varepsilon > 0$. By (2.28), the martingale property for M, and the adaptedness of H, we obtain

$$\begin{aligned} ||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} &= E\big[(I^\pi_a - I^{\widetilde{\pi}}_a)^2\big] \\ &= E\big[\sum_{s\in\pi}(H_s - H_{\lfloor s\rfloor})^2(M_{s'} - M_s)^2\big] \qquad (2.29) \\ &\le \varepsilon^2 E\big[\sum_{s\in\pi}(M_{s'} - M_s)^2\big] + (2C)^2 E\big[\sum_{t\in\widetilde{\pi}}\sum_{\substack{s\in\pi\\ \tau_t(\varepsilon)\le s<\lceil t\rceil}}(M_{s'} - M_s)^2\big] \end{aligned}$$

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 (2.29) are now estimated separately. Since M is a bounded martingale, we can easily control the first summand:

$$E\big[\sum(M_{s'} - M_s)^2\big] = \sum E\big[M^2_{s'} - M^2_s\big] = E\big[M^2_a - M^2_0\big] \le C^2. \qquad (2.30)$$

The second summand is more difficult to handle. Noting that

$$E\big[(M_{s'} - M_s)^2 \mid \mathcal{F}_{\tau_t}\big] \quad = \quad E\big[M^2_{s'} - M^2_s \mid \mathcal{F}_{\tau_t}\big] \qquad \text{on} \quad \{\tau_t \le s\}\,,$$

we can rewrite the expectation value as

$$\begin{aligned} &\sum_{t\in\widetilde{\pi}} E\big[\sum_{\tau_t\le s<\lceil t\rceil} E\big[(M_{s'} - M_s)^2 \mid \mathcal{F}_{\tau_t}\big]\big] \qquad (2.31) \\ &= \sum_{t\in\widetilde{\pi}} E\big[E\big[M^2_{\lceil t\rceil} - M^2_{\tau_t} \mid \mathcal{F}_{\tau_t}\big]\big] = E\big[\sum_{t\in\widetilde{\pi}}(M_{\lceil t\rceil} - M_{\tau_t})^2\big] \quad =: \quad B \end{aligned}$$

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 (2.31). 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) $|u-v| \le \delta \quad \Rightarrow \quad |H_u - H_v| \le \varepsilon$ or $(u,v] \cap D_{\varepsilon/2} \neq \emptyset$,

(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

$$B_1 = E\big[\sum_{t \in \widetilde{\pi}} (M_{\lceil t \rceil} - M_{\tau_t})^2 \,;\, \Delta \le \delta,\ |D_{\varepsilon/2}| \le k\big] \le k\bar{\varepsilon}^2, \tag{2.32}$$

$$\begin{aligned} B_2 &= E\big[\sum_{t \in \widetilde{\pi}} (M_{\lceil t \rceil} - M_{\tau_t})^2 \,;\, \Delta > \delta \text{ or } |D_{\varepsilon/2}| > k\big] \\ &\le E\big[(\sum_{t \in \widetilde{\pi}} (M_{\lceil t \rceil} - M_{\tau_t})^2)^2\big]^{1/2}\, P\big[\Delta > \delta \text{ or } |D_{\varepsilon/2}| > k\big]^{1/2} \qquad (2.33) \\ &\le \sqrt{6}\, C^2 \big(P[\Delta > \delta] + P\big[|D_{\varepsilon/2}| > k\big]\big)^{1/2} \quad . \end{aligned}$$

In the last step we have used the following upper bound for the martingale increments $\eta_t := M_{\lceil t \rceil} - M_{\tau_t}$:

$$\begin{aligned} E\big[\big(\sum_{t \in \widetilde{\pi}} \eta_t^2\big)^2\big] &= E\big[\sum_t \eta_t^4\big] + 2E\big[\sum_t \sum_{u>t} \eta_t^2 \eta_u^2\big] \\ &\le 4C^2 E\big[\sum_t \eta_t^2\big] + 2E\big[\sum_t \eta_t^2 E\big[\sum_{u>t} \eta_u^2 \mid \mathcal{F}_t\big]\big] \\ &\le 6C^2 E\big[\sum_t \eta_t^2\big] \le 6C^2 E\big[M_a^2 - M_0^2\big] \le 6C^4. \end{aligned}$$

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 (2.29), (2.30) and (2.31),

$$||I^\pi - I^{\widetilde{\pi}}||^2_{M^2([0,a])} \le \varepsilon^2 C^2 + 4C^2(B_1 + B_2) \tag{2.34}$$

where B_1 and B_2 are estimated in (2.32) and (2.33). Let $\gamma > 0$ be given. To bound the right hand side of (2.34) 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 $4\sqrt{6}\, C^4 P\big[|D_{\varepsilon/2}| > k\big]^{1/2} < \gamma/4$,
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 $4\sqrt{6}\, C^4 P\big[\Delta > \delta\big]^{1/2} < \gamma/4$.

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*

$$H_{-\bullet}M\big|_{[0,a]} \quad = \quad \lim_{n\to\infty} H^{\pi_n}_{-\bullet}M\big|_{[0,a]} \qquad \text{in } M^2_d([0,a])$$

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 2.8 (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

$$\Delta(H_{-\bullet}M)_t \quad = \quad \lim_{n\to\infty} H_{\lfloor t\rfloor_n}(M_t - M_{t-}) \quad = \quad H_{t-}\Delta M_t \tag{2.35}$$

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, (2.35) 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. [32]. 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 $M_t = \int yI_{\{|y|\leq 1\}}\ (N_t(dy) - t\nu(dy))$ is a strict (local) martingale.

Strict local martingales can be localized by bounded martingales:

Lemma 2.9. *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

$$S_n \quad := \quad T_n \wedge \inf\{t \geq 0\ :\ |M_t| \geq n\} \quad , \quad n \in \mathbb{N},$$

is a non-decreasing sequence of stopping times with $\sup S_n = \infty$, and the stopped processes M^{S_n} are uniformly bounded, since

$$|M_{t\wedge S_n}| \quad \leq \quad n + |\Delta M_{S_n}| \quad = \quad n + |\Delta M^{T_n}_{S_n}| \qquad \text{for any } t \geq 0.$$

□

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 $(\mathcal{F}_t^P)$ martingale, and*

$(H_t)_{t\geq 0}$ is càdlàg and $(\mathcal{F}_t^P)$ adapted. Then the stochastic integral $H_{-\bullet}M$ is the unique equivalence class of local $(\mathcal{F}_t^P)$ martingales satisfying

$$H_{-\bullet}M\big|_{[0,T]} \quad = \quad \widetilde{H}_{-\bullet}\widetilde{M}\big|_{[0,T]} \quad \textit{a.s.}, \tag{2.36}$$

whenever T is an $(\mathcal{F}_t^P)$ stopping time, $\widetilde{H}$ is a bounded càdlàg $(\mathcal{F}_t^P)$ adapted process with $H|_{[0,T)} = \widetilde{H}|_{[0,T)}$ almost surely, and $\widetilde{M}$ is a bounded càdlàg $(\mathcal{F}_t^P)$ 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 (2.36) because of the local dependence of the stochastic integral w.r.t. bounded martingales on H and M (Lemma 2.8, 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 2.9 and Lemma 2.8, 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 2.8, $\Delta(H^{(n)}_{-\bullet}M^{T_n}) = H^{(n)}_{-}\Delta M^{T_n}$ 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 $(\mathcal{F}_t^P)$ semimartingale with decomposition

$$X_t \quad = \quad X_0 + M_t + A_t \quad , \quad t \geq 0,$$

into a strict local $(\mathcal{F}_t^P)$ martingale M and an $(\mathcal{F}_t^P)$ 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 $(\mathcal{F}_t^P)$ 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 2.10. *Suppose that (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$. 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 $(\mathcal{F}_t^P)$ semimartingale and H is $(\mathcal{F}_t^P)$ adapted.*
3) *If X is also a semimartingale w.r.t. a probability measure Q that is absolutely continuous w.r.t. P then each version of the integral $(H_{-\bullet}X)_P$ defined w.r.t. 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 2.11 (Elementary properties of stochastic integrals).

1) ***Semimartingale decomposition**: The integral $H_{-\bullet}X$ is again an $(\mathcal{F}_t^P)$ 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 $(\mathcal{F}_t^P)$ stopping time then*

$$(H_{-\bullet}X)^T \quad = \quad H_{-\bullet}X^T \quad = \quad (H \cdot I_{[0,T)})_{-\bullet}X.$$

2.3 Quadratic variation and covariation

From now on we fix a probability space $(\Omega, \mathcal{A}, P)$ with a filtration $(\mathcal{F}_t)$. The vector space of (equivalence classes of) **strict** local $(\mathcal{F}_t^P)$ martingales and of $(\mathcal{F}_t^P)$ 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 $(\mathcal{F}_t^P)$ 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*

$$\sup_{t \leq a} |X_t^n - X_t| \xrightarrow{P} 0 \qquad \text{as } n \to \infty \qquad \text{for any } a \in \mathbb{R}_+.$$

By Theorem (2.10), 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 $\mathcal{S} \times \mathcal{S} \to \mathrm{FV}$. 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 2.10.

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 $(\mathcal{F}_t^P)$ adapted càdlàg process. Moreover, $(X,Y) \mapsto [X,Y]$ is symmetric and bilinear, and hence the polarization identity

$$[X,Y] \quad = \quad \frac{1}{2}\left([X+Y] - [X] - [Y]\right)$$

holds for any $X, Y \in \mathcal{S}$ where

$$[X] \quad = \quad [X,X]$$

denotes the **quadratic variation** of X. The next corollary shows that $[X,Y]$ deserves the name "covariation":

Corollary 2.12. *For any sequence (π_n) of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$,*

$$[X,Y]_t \quad = \quad \mathrm{ucp} - \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s). \tag{2.37}$$

In particular, the following statements hold almost surely:

1) $[X]$ *is non-decreasing, and* $[X,Y]$ *has finite variation.*

2) $\Delta[X,Y] \quad = \quad \Delta X \Delta Y.$

3) $[X,Y]^T \quad = \quad [X^T, Y] \quad = \quad [X, Y^T] \quad = \quad [X^T, Y^T].$

4) $|[X,Y]| \quad \le \quad [X]^{1/2}[Y]^{1/2}.$

Proof. (2.37) is a direct consequence of Theorem 2.10, and 1) follows from (2.37) and the polarization identity. 2) follows from Theorem 2.11, 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 (2.37) 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

$$[X,Y]_t = [X,Y]_t^c + \sum_{s\le t}\Delta X_s\Delta Y_s \tag{2.38}$$

into a continuous part and a pure jump part.

If Y has finite variation then by Lemma 2.2,

$$[X,Y]_t = \sum_{s\le t}\Delta X_s\Delta Y_s.$$

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,

$$[X,Y]^c = [M,N]^c + [M,B]^c + [A,N]^c + [A,B]^c = [M,N]^c.$$

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 2.13. *Let $M \in \mathcal{M}_{loc}$ and $a \in [0,\infty)$. Then $M \in \mathcal{M}^2_d([0,a])$ if and only if $M_0 \in \mathcal{L}^2$ and $[M]_a \in \mathcal{L}^1$. In this case, $M_t^2 - [M]_t$ $(0 \le t \le a)$ is a martingale, and*

$$||M||^2_{M^2([0,a])} \quad = \quad E\big[M_0^2\big] + E\big[[M]_a\big]. \tag{2.39}$$

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,

$$E\big[M^2_{t\wedge T_n}\big] \quad = \quad E\big[[M]_{t\wedge T_n}\big] \qquad \text{for any } \ t \ge 0 \quad \text{and any } n \in \mathbb{N}. \tag{2.40}$$

Since M^2 is a submartingale, we have

$$E[M_t^2] \quad \le \quad \liminf_{n\to\infty} E[M^2_{t\wedge T_n}] \quad \le \quad E[M_t^2] \tag{2.41}$$

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 (2.41), we obtain

$$E[M_t^2] \quad = \quad E\big[[M]_t\big] \qquad \text{for any } t \ge 0.$$

For $t \le a$, the right-hand side is dominated from above by $E\big[[M]_a\big]$, Therefore, if $[M]_a$ is integrable then M is in $M^2_d([0,a])$ with M^2 norm $E\big[[M]_a\big]$. Moreover, in this case, the sequence $\big(M^2_{t\wedge T_n} - [M]_{t\wedge T_n}\big)_{n\in\mathbb{N}}$ is uniformly integrable for each $t \in [0,a]$, because,

$$\sup_{t\le a} |M_t^2 - [M]_t| \quad \le \quad \sup_{t\le a} |M_t|^2 + [M]_a \quad \in \mathcal{L}^1,$$

Therefore, the martingale property carries over from the stopped processes $M^2_{t\wedge T_n} - [M]_{t\wedge T_n}$ to $M_t^2 - [M]_t$. □

Remark. The assertion of Theorem 2.13 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 (2.40) 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 2.14. *The map $(M,N) \mapsto [M,N]$ is symmetric, bilinear and continuous on $M^2_d([0,a])$ in the sense that*

$$E[\sup_{t\le a}|[M,N]|_t] \quad \le \quad ||M||_{M^2([0,a])}||N||_{M^2([0,a])}.$$

Proof. By the Cauchy-Schwarz inequality for the covariation (Cor. 2.12,4),

$$|[M,N]_t| \quad \le \quad [M]_t^{1/2}[N]_t^{1/2} \quad \le \quad [M]_a^{1/2}[N]_a^{1/2} \quad \forall\, t \le a.$$

Applying the Cauchy-Schwarz inequality w.r.t. the L^2-inner product yields

$$E[\sup_{t\le a}|[M,N]_t|] \quad \le \quad E\big[[M]_a\big]^{1/2}E\big[[N]_a\big]^{1/2} \quad \le \quad ||M||_{M^2([0,a])}||N||_{M^2([0,a])}$$

by Theorem 2.13. □

Corollary 2.15. *Let $M \in \mathcal{M}_{loc}$ and suppose that $[M]_a = 0$ almost surely for some $a \in [0,\infty]$. Then almost surely,*

$$M_t \quad = \quad M_0 \qquad \textit{for any } t \in [0,a].$$

In particular, continuous local martingales of finite variation are almost surely constant.

Proof. By Theorem 2.13, $||M - M_0||_{M^2([0,a])} = E\big[[M]_a\big] = 0$. □

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 $(B_t^H)_{t\ge0}$ satisfying

$$E\big[B_t^H\big] \quad = \quad 0 \qquad \text{and} \qquad \mathrm{Cov}\left[B_s^H, B_t^H\right] \quad = \quad \frac{1}{2}\big(t^{2H} + s^{2H} - |t-s|^{2H}\big)$$

for any $s,t \ge 0$. It has been introduced by Mandelbrot as an example of a self-similar process and is used in various applications, cf. [1]. 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. [17]. 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 $B_t^H = M_t + A_t$ 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 2.15, 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. [32]). 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. [17].

The covariation $[M,N]$ of local martingales can be characterized in an alternative way that is often useful for determining $[M,N]$ explicitly.

Theorem 2.16 (**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) $MN - A \quad \in M_{loc} \quad ,$ *and*

(ii) $\Delta A \quad = \quad \Delta M\, \Delta N \quad , \quad A_0 = 0$ *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 2.15. ☐

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 2.15 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. [32]. For applications it is more important that in many situations the conditional covariance process can be easily determined explicitly, see the example below.

Corollary 2.17. *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 2.13 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 2.16.

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 2.13 now follow similarly as above. ☐

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

$$X_t \quad = \quad \int_{\mathbb{R}^d \setminus \{0\}} y \, \big(N_t(dy) - t \, I_{\{|y| \leq 1\}} \nu(dy) \big) + bt$$

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 $\operatorname{supp}(\nu) \subset \big\{ y \in \mathbb{R}^d : |y| \geq \varepsilon \big\}$ for some $\varepsilon > 0$. Then the components X^k are finite variation processes, and hence

$$[X^k, X^l]_t \quad = \quad \sum_{s \leq t} \Delta X_s^k \Delta X_s^l \quad = \quad \int y^k y^l \, N_t(dy). \tag{2.42}$$

In general, (2.42) 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 2.14,

$$\big[X^k, X^l\big]_t \quad = \quad \text{ucp} - \lim_{\varepsilon \downarrow 0} \big[X^{(\varepsilon),k}, X^{(\varepsilon),l}\big]_t \quad = \quad \sum_{s \leq t} \Delta X_s^k \Delta X_s^l.$$

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

$$\langle B^k, X^l \rangle \quad = \quad [B^k, X^l] \quad = \quad 0 \qquad \text{almost surely for any } k \text{ and } l. \tag{2.43}$$

Indeed, (2.43) 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 (2.43) 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 2.19 below.

Theorem 2.18. *Suppose that X and Y are $(\mathcal{F}_t^P)$ semimartingales, and H is $(\mathcal{F}_t^P)$ adapted and càdlàg. Then*

$$\left[\int H_-\,dX, Y\right] \quad = \quad \int H_-\,d[X,Y] \qquad \textit{almost surely}. \tag{2.44}$$

Proof. 1. We first note that (2.44) 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

$$\left[\int H_-\,dX, Y\right] \;=\; \sum \Delta(H_{-\bullet}X)\Delta Y \;=\; \sum H_-\Delta X\Delta Y \;=\; \int H_-\,d[X,Y]\ .$$

The same holds if $Y \in \mathrm{FV}$.

2. Now we show that (2.44) 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

$$[I^\pi, Y] \quad = \quad \int H_-^\pi\,d[X,Y] \quad \text{almost surely}. \tag{2.45}$$

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

$$\sum_{\substack{r\in\widetilde{\pi}_n\\ r<t}} (I_{r'\wedge t}^\pi - I_r^\pi)(Y_{r'\wedge t} - Y_r) \quad = \quad \sum_{s\in\pi} H_s \sum_{\substack{r\in\widetilde{\pi}_n\\ s\le r<s'\wedge t}} (X_{r'\wedge t} - X_r)(Y_{r'\wedge t} - Y_r).$$

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 (2.45) holds.
Having verified (2.45) 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 2.14 and (2.45),

$$[\int H_- \, dX, Y] \quad = \quad \mathrm{ucp}\text{-}\lim_{n\to\infty} [I^{\pi_n}, Y] \quad = \quad \int H_- \, d[X,Y].$$

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 2.11 and Corollary 2.12,

$$\begin{aligned} [\int H_- \, dX, Y]^T \quad &= \quad [(\int H_- \, dX)^T, Y^T] \quad = \quad [\int (H_- \, I_{[0,T)}) \, dX^T, Y^T] \\ &= \quad \int H_- \, I_{[0,T)} \, d[X^T, Y^T] \quad = \quad (\int H_- \, d[X,Y])^T. \end{aligned}$$

Since this holds for all localizing stopping times as above, (2.45) is satisfied as well.

4. Finally, suppose that X and Y are arbitrary semimartingales. Then $X = M + A$ and $Y = N + B$ with $M, N \in M_{\text{loc}}$ and $A, B \in \mathrm{FV}$. The assertion (2.44) now follows by Steps 1 and 3 and by the bilinearity of stochastic integral and covariation. $\square$

Perhaps the most remarkable consequences of Theorem 2.18 is:

Corollary 2.19 (Kunita-Watanabe characterization of stochastic integrals). *Let $M \in M_{loc}$ and $G = H_-$with H $(\mathcal{F}_t^P)$ 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 2.18, $G_\bullet M$ 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 $(\mathcal{F}_\mathbf{t}^\mathbf{P})$ **predictable** iff the function $(\omega, t) \to G_t(\omega)$ is measurable w.r.t. the σ-algebra $\mathscr{P}$ on $\Omega \times [0, \infty)$ generated by all $(\mathcal{F}_t^P)$ adapted left-continuous processes.

Definition (Stochastic integrals with general predictable integrands).
Let $M \in M_{loc}$, and suppose that G is an $(\mathcal{F}_t^P)$ 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 2.19 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 2.21 below.

The Itô isometry for stochastic integrals w.r.t. martingales

Of course, Theorem 2.18 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 2.20 (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*

$$\left|\left| \int G\, dM \right|\right|^2_{M^2([0,a])} = E\Big[\Big(\int_0^a G\, dM\Big)^2\Big] = E\Big[\int_0^a G^2\, d[M]\Big] \qquad \forall\, a \geq 0, \quad a.s\ .$$

Proof. If $M \in M_{\mathrm{loc}}$ then $G_\bullet M \in M_{\mathrm{loc}}$, and hence $(G_\bullet M)^2 - [G_\bullet M] \in M_{\mathrm{loc}}$. Moreover, by Theorem 2.13,

$$||G_\bullet M||^2_{M^2([0,a])} = E\big[[G_\bullet M]_a\big] = E\big[(G^2_\bullet [M])_a\big].$$

□

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

$$P_{[M]}(d\omega\ dt) = P(d\omega)\ [M](\omega)(dt).$$

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.

2.4 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 $\int G\, dY = \int G\widetilde{G}\, dX$:

$$dY = \widetilde{G}\, dX \quad \Longrightarrow \quad G\, dY = G\widetilde{G}\, dX$$

The covariation characterization of stochastic integrals w.r.t. local martingales can be used to prove this rule in a simple way.

Theorem 2.21 ("Associative Law"). *Let $X \in \mathcal{S}$. Then*

$$G_\bullet(\widetilde{G}_\bullet X) \quad = \quad (G\widetilde{G})_\bullet X \tag{2.46}$$

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 (2.46) 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 (2.46) with local martingales. For $N \in M_{\mathrm{loc}}$, Corollary 2.19 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 (2.46) holds by Corollary 2.19. ☐

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 $0 < T_1 < T_2 < \ldots$ 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. [13, Thm.6.4]), we could expect that

$$\begin{aligned} F(X_t) - F(X_0) \quad &= \quad \sum_k \big(F(X_{t\wedge T_k}) - F(X_{t\wedge T_{k-1}})\big) \\ = \sum_{T_{k-1}<t} \Bigg(\int\limits_{T_{k-1}}^{t\wedge T_k -} F'(X_{s-})\, dX_s + \frac{1}{2} \int\limits_{T_{k-1}}^{t\wedge T_k -} F''(X_{s-})\, d[X]_s \Bigg) &+ \sum_{T_k \le t} \big(F(X_{T_k}) - F(X_{T_k -})\big) \\ = \int_0^t F'(X_{s-})\, dX_s^c + \frac{1}{2}\int_0^t F''(X_{s-})\, d[X]_s^c &+ \sum_{s\le t} \big(F(X_s) - F(X_{s-})\big) \end{aligned} \tag{2.47}$$

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 (2.47) in the equivalent form

$$\begin{aligned} F(X_t)-F(X_0) &= \int_0^t F'(X_{s-})\,dX_s + \frac{1}{2}\int_0^t F''(X_{s-})\,d[X]_s^c \\ &\quad + \sum_{s\le t}\big(F(X_s)-F(X_{s-})-F'(X_{s-})\,\Delta X_s\big), \end{aligned} \tag{2.48}$$

which carries over to general semimartingales.

Theorem 2.22 (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 $F \in C^2(\mathbb{R}^d)$,*

$$\begin{aligned} F(X_t)-F(X_0) =& \sum_{i=1}^d \int_{(0,t]} \frac{\partial F}{\partial x^i}(X_{s-})\,dX_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_{(0,t]} \frac{\partial^2 F}{\partial x^i\partial x^j}(X_{s-})\,d[X^i,X^j]_s^c \\ &+ \sum_{s\in(0,t]}\Big(F(X_s)-F(X_{s-})-\sum_{i=1}^d \frac{\partial F}{\partial x^i}(X_{s-})\Delta X_s^i\Big) \end{aligned} \tag{2.49}$$

for any $t \ge 0$, 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 (2.49). Indeed, a Taylor expansion up to order two shows that

$$\begin{aligned} \sum_{s\le t}\Big|F(X_s)-F(X_{s-})-\sum_{i=1}^d \frac{\partial F}{\partial x^i}(X_{s-})\Delta X_s^i\Big| &\le C_t\cdot\sum_{s\le t}\sum_i |\Delta X_s^i|^2 \\ &\le C_t\cdot\sum_i [X^i]_t < \infty, \end{aligned}$$

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. [32]. Here, following [34], 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 2.4), we first prove that the set $\mathcal{A}$ consisting of all functions $F \in C^2(\mathbb{R})$ satisfying (2.48) 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

$$\begin{aligned} F(X_t)G(X_t) - F(X_0)G(X_0) = &\int_0^t F(X_-)\, dG(X) + \int_0^t G(X_-)\, dF(X) \\ &+ \left[F(X), G(X)\right]^c + \sum_{(0,t]} \Delta F(X) \Delta G(X), \end{aligned} \tag{2.50}$$

the associative law, which implies

$$\begin{aligned} \int F(X_-)\, dG(X) \quad &= \quad \int F(X_-)G'(X_-)\, dX + \frac{1}{2} \int F(X_-)G''(X_-)\, d[X]^c \\ &\qquad + \sum F(X_-)\left(\Delta G(X) - G'(X_-)\Delta X\right), \end{aligned} \tag{2.51}$$

the corresponding identity with F and G interchanged, and the formula

$$\begin{aligned} [F(X), G(X)]^c \quad &= \quad \Big[\int F'(X_-)\, dX, \int G'(X_-)\, dX \Big]^c \qquad (2.52) \\ &= \quad \Big(\int F'(X_-)G'(X_-)\, d[X] \Big)^c \quad = \quad \int (F'G')(X_-)\, d[X]^c \end{aligned}$$

for the continuous part of the covariation. Both (2.51) and (2.52) follow from (2.49) and the corresponding identity for G. It is straightforward to verify that (2.50), (2.51) and (2.52) imply the change of variable formula (2.48) for FG, i.e., $FG \in \mathcal{A}$. Therefore, by induction, the formula (2.48) 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 $A \in \mathrm{FV}$. 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

$$
\begin{aligned}
F(X_t) - F(X_0) &= \lim_{n\to\infty} \big(p_n(X_t) - p_n(X_0)\big) \\
= \quad \lim_{n\to\infty} \Big(\int_0^t p_n'(X_{s-})\, dX_s &+ \frac{1}{2}\int_0^t p_n''(X_{s-})\, d[X]_s^c + \sum_{s\leq t} \int_{X_{s-}}^{X_s} \int_{X_{s-}}^{y} p_n''(z)\, dz\, dy \Big) \\
= \quad \int_0^t F'(X_{s-})\, dX_s &+ \frac{1}{2}\int_0^t F''(X_{s-})\, d[X]_s^c + \sum_{s\leq t} \int_{X_{s-}}^{X_s} \int_{X_{s-}}^{y} F''(z)\, dz\, dy
\end{aligned}
$$

w.r.t. convergence in probability. Here we have used an expression of the jump terms in (2.48) 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 $A \in \mathrm{FV}$ 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 (2.49) 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 (2.49). 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

$$X_t \quad = \quad x+\sigma B_t+bt+\int y\,\widetilde{N}_t(dy) \tag{2.53}$$

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 (2.54),

$$\begin{aligned}
&F(X_t)-F(X_0)\\
&=\int_0^t F'(X_-)\,dX+\frac{1}{2}\int_0^t F''(X_-)\,d[X]^c+\sum_{s\le t}\big(F(X)-F(X_-)-F'(X_-)\Delta X\big)\\
&=\int_0^t(\sigma F')(X_{s-})\,dB_s+\int_0^t(bF'+\frac{1}{2}\sigma^2F'')(X_s)\,ds+\int\limits_{(0,t]\times\mathbb{R}}F'(X_{s-})\,y\,\widetilde{N}(ds\,dy)\\
&\qquad+\int\limits_{(0,t]\times\mathbb{R}}\big(F(X_{s-}+y)-F(X_{s-})-F'(X_{s-})y\big)\,N(ds\,dy),
\end{aligned} \tag{2.54}$$

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 (2.54), 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 (2.54) 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 2.23 (Martingale problem for Lévy processes). *For any $F \in C^2(\mathbb{R})$, the process*

$$\begin{aligned} M_t^{[F]} &= F(X_t) - F(X_0) - \int_0^t (\mathcal{L}F)(X_s)\, ds, \\ (\mathcal{L}F)(x) &= \frac{1}{2}(\sigma F'')(x) + (bF')(x) + \int \big(F(x+y) - F(x) - F'(x)y\big)\, \nu(dy), \end{aligned}$$

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?

Applications to Itô diffusions

Next we consider a solution of a stochastic differential equation

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \qquad X_0 = x_0, \tag{2.55}$$

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 continuous time-dependent vector fields in $\mathbb{R}^n$, 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 (2.55) is a continuous $(\mathcal{F}_t^P)$ semimartingale (X_t) satisfying

$$X_t \quad = \quad x_0 + \int_0^t b(s, X_s)\, ds + \sum_{k=1}^d \int_0^t \sigma_k(s, X_s)\, dB_s^k \qquad \forall\, t \geq 0 \;\; a.s. \tag{2.56}$$

If X is a solution then

$$\begin{aligned}
[X^i, X^j]_t \quad &= \quad \sum_{k,l} \Big[\int \sigma_k^i(s, X)\, dB^k, \int \sigma_l^j(s, X)\, dB^l \Big]_t \\
&= \quad \sum_{k,l} \int_0^t (\sigma_k^i \sigma_l^j)(s, X)\, d[B^k, B^l] \quad = \quad \int_0^t a^{ij}(s, X_s)\, ds
\end{aligned}$$

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

$$(\mathcal{L}F)(t,x) = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(t,x)\frac{\partial^2 F}{\partial x^i \partial x^j}(t,x) + \sum_{i=1}^{d} b^i(t,x)\frac{\partial F}{\partial x^i}(t,x).$$

We have thus derived the *Itô-Doeblin formula*

$$\boxed{F(t,X_t) - F(0,X_0) = \int_0^t (\sigma^T \nabla F)(s,X_s)\cdot dB_s + \int_0^t \big(\frac{\partial F}{\partial t} + \mathcal{L}F\big)(s,X_s)\,ds} \tag{2.57}$$

Again, the formula provides a semimartingale decomposition for $F(t, X_t)$. It establishes a connection between the stochastic differential equation (2.55) 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.

$$\frac{\partial F}{\partial t}(t,x) + (\mathcal{L}F)(t,x) = -g(t,x) \qquad \forall\, t \ge 0,\ x \in U$$

on an open subset $U \subset \mathbb{R}^n$ with boundary values

$$F(t,x) = \varphi(t,x) \qquad \forall\, t \ge 0,\ x \in \partial U.$$

Then by (2.57), the process

$$M_t = F(t,X_t) + \int_0^t g(s,X_s)\,ds$$

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 $T = \inf\{t \ge 0 \,:\, X_t \notin U\}$ is a martingale. Hence, if T is almost surely finite then

$$E[\varphi(T,X_T)] + E\Big[\int_0^T g(s,X_s)\,ds\Big] = F(0,x_0).$$

This can be used, for example, to compute exit distributions (for $g \equiv 0$) and mean exit times (for $\varphi \equiv 0$, $g \equiv 1$) 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

$$\begin{aligned} \frac{\partial u}{\partial s}(s,x) &= (\mathcal{L}u)(s,x) - V(s,x)u(s,x) \qquad \text{for } s \in (0,t],\ x \in \mathbb{R}^n, \\ u(0,x) &= \varphi(x), \end{aligned}$$

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

$$U_k \quad = \quad \{x \in U \;:\; |x| < k \text{ and } \operatorname{dist}(x,U^c) > 1/k\}\,, \quad k \in \mathbb{N}.$$

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 (2.55) 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 (2.56) for $t \le T_k$. By applying Itô's formula to the stopped processes, we obtain:

Corollary 2.24 (Martingale problem for Itô diffusions). *If $X_t : \Omega \to U$ is a solution of (2.55) up to the explosion time ζ, then for any $F \in C^2(\mathbb{R}_+ \times U)$ and $x_0 \in U$, the process*

$$M_t \quad := \quad F(t, X_t) - \int_0^t \Big(\frac{\partial F}{\partial t} + \mathcal{L}F\Big)(s, X_s)\, ds, \quad t < \zeta,$$

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 $F_k \in \mathcal{C}_b^2([0,a] \times U)$, $k \in \mathbb{N}$, $a \geq 0$, 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 $t \leq a$,

$$M_t^{T_k} \quad = \quad M_{t \wedge T_k} \quad = \quad F_k(t, X_{t\wedge T_k}) - \int_0^t \Big(\frac{\partial F_k}{\partial t} + \mathcal{L}F_k\Big)(s, X_{s\wedge T_k})\, ds.$$

By (2.57), the right hand side is a bounded martingale. □

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

$$h_n(t,x) \quad = \quad \frac{\partial^n}{\partial \alpha^n} \exp(\alpha x - \alpha^2 t/2)\Big|_{\alpha=0} \tag{2.58}$$

denotes the Hermite polynomial of order n and $X_0 = 0$ then

$$H_t^n \quad = \quad h_n\big([X]_t, X_t\big) \tag{2.59}$$

solves the SDE

$$dH^n \quad = \quad n\, H^{n-1}\, dX, \qquad H_0^n = 0,$$

for any $n \in \mathbb{N}$, cf. Section 6.4 in [13]. In particular, H^n is an iterated Itô integral:

$$H_t^n \quad = \quad n! \int_0^t \int_0^{s_n} \cdots \int_0^{s_2} dX_{s_1} dX_{s_2} \cdots dX_{s_n}.$$

The formula for the stochastic exponential can be generalized to the discontinuous case:

Theorem 2.25 (Doléans-Dade). *Let $X \in \mathcal{S}$. Then the unique solution of the exponential equation*

$$Z_t \quad = \quad 1 + \int_0^t Z_{s-}\, dX_s, \qquad t \geq 0, \tag{2.60}$$

is given by

$$Z_t \quad = \quad \exp\left(X_t - \frac{1}{2}[X]_t^c\right) \prod_{s \in (0,t]} (1 + \Delta X_s) \exp(-\Delta X_s). \tag{2.61}$$

Remarks. 1) In the finite variation case, (2.61) can be written as

$$Z_t \quad = \quad \exp\left(X_t^c - \frac{1}{2}[X]_t^c\right) \prod_{s \in (0,t]} (1 + \Delta X_s).$$

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 (2.59) . 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 2.25. The proof is partially similar to the one given above for $X \in$ FV, cf. Theorem 2.5. 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

$$S_t \quad = \sum_{\substack{s \le t \\ |\Delta X_s| \le 1/2}} (\log(1+\Delta X_s) - \Delta X_s), \qquad t \ge 0,$$

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

$$G \quad = \quad 1 + \int G_- \, dX + \sum (\Delta G - G_- \, \Delta X) \tag{2.62}$$

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 (2.62),

$$\begin{aligned} Z - 1 \quad &= \quad \int P_- \, dG + \int G_- \, dP + [G,P] \\ &= \quad \int P_- G_- \, dX + \sum (P_- \Delta G - P_- G_- \, \Delta X + G_- \, \Delta P + \Delta G \, \Delta P) \\ &= \quad \int Z_- \, dX + \sum (\Delta Z - Z_- \, \Delta X) \quad = \quad \int Z_- \, dX. \end{aligned}$$

This proves that Z solves the SDE (2.60). Uniqueness of the solution follows from a general uniqueness result for SDE with Lipschitz continuous coefficients, cf. Section 3.1. □

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

$$Z_t \quad = \quad \exp(\sigma X_t + \mu t), \qquad t \ge 0,$$

where σ and μ are real constants. Since $[X]^c \equiv 0$, Itô's formula yields

$$\begin{aligned} Z_t - 1 \quad &= \sigma \int_{(0,t]} Z_- \, dX + \mu \int_{(0,t]} Z_- \, ds + \sum_{(0,t]} Z_-\left(e^{\sigma \Delta X} - 1 - \sigma \Delta X\right) \qquad (2.63) \\ = \sigma \int_{(0,t]\times\mathbb{R}} Z_{s-} \, y \, \widetilde{N}(ds\,dy) &+ \mu \int_{(0,t]} Z_{s-} \, ds + \int_{(0,t]\times\mathbb{R}} Z_{s-}\left(e^{\sigma y} - 1 - \sigma y\right) N(ds\,dy). \end{aligned}$$

If $\int e^{2\sigma y}\,\nu(dy) < \infty$ then (2.63) leads to the semimartingale decomposition

$$dZ_t \quad = \quad Z_{t-}\,dM_t^\sigma + \alpha Z_{t-}\,dt, \qquad Z_0 = 1, \tag{2.64}$$

where

$$M_t^\sigma = \int \left(e^{\sigma y} - 1\right) \widetilde{N}_t(dy)$$

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).

2.5 General predictable integrands and local times

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 [32]. At the end of the section, we apply the results to define the local time process (occupation time density) of a continuous semimartingale.

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 $0 \le s \le t$, 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 $n \in \mathbb{N}$, $0 \le t_0 < t_1 < \cdots < t_n$, 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

$$\begin{aligned} ||G_\bullet M||^2_{M^2(0,\infty)} \quad &= \quad E\left[\left(\int_0^\infty G \, dM \right)^2 \right] \\ &= \quad E\left[\int_0^\infty G^2 \, d[M] \right] \quad = \quad \int_{\Omega \times \mathbb{R}_+} G^2 \, dP_{[M]} \end{aligned} \tag{2.65}$$

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 2.20, 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

$$\begin{aligned} P_{[M]}(C) \quad &= \quad E\left[[M]_t - [M]_s \, ; \, A \right] \quad = \quad E\left[E[[M]_t - [M]_s | \mathcal{F}_s] \, ; \, A \right] \\ &= \quad E\left[E[\langle M \rangle_t - \langle M \rangle_s | \mathcal{F}_s] \, ; \, A \right] \quad = \quad P_{\langle M \rangle}(C). \end{aligned}$$

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

$$P_{[M]} \quad = \quad P_{\langle M \rangle} \quad = \quad \psi''(0)\, P \otimes \lambda_{(0,\infty)}$$

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 [32]:

Fact. *Any predictable local martingale with finite variation paths is almost surely constant.*

The result implies that the *Doob-Meyer semimartingale decomposition*

$$X \quad = \quad M + A \tag{2.66}$$

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

$$||X||^2_{\mathcal{H}^2} \quad = \quad ||M||^2_{M^2} + ||V_\infty^{(1)}(A)||^2_{L^2} \quad = \quad E\left[[M]_\infty + \left(\int_0^\infty |dA|\right)^2\right].$$

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 (2.65), we obtain:

Corollary 2.26 (Itô isometry for semimartingales). *Let $X \in \mathcal{H}^2$ with semimartingale decomposition as above. Then*

$$||G_\bullet X||_{\mathcal{H}^2} \quad = \quad ||G||_X \qquad \textit{for any } G \in \mathcal{E}, \;\; \textit{where}$$
$$||G||^2_X \quad := \quad ||G||^2_{L^2(P_{[M]})} + \left|\left| \int_0^\infty |G|\, |dA| \right|\right|^2_{L^2(P)}.$$

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 $M_d^2(0,\infty)$ and $G_\bullet A$ is a predictable finite variation process. Therefore, and by (2.65),

$$||G_\bullet X||^2_{\mathcal{H}^2} = ||G_\bullet M||^2_{M^2} + ||V_\infty^{(1)}(G_\bullet A)||^2_{L^2} = ||G||^2_{L^2(P_{[M]})} + \left|\left| \int |G|\,|dA| \right|\right|^2_{L^2(P)}.$$

□

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 2.20, 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 2.27. $\overline{\mathcal{E}}^X$ *contains all predictable processes* G *with* $||G||_X < \infty$.

Proof. We only mention the main steps of the proof, cf. [32] 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 $G^n := G \cdot I_{\{G \le n\}}$, $n \in \mathbb{N}$, 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 [32] 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 [32].

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 $M \in \mathcal{M}_c^{\mathrm{loc}}$ and $A \in \mathrm{FV}_c$. 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 [32] 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:

$$\int_0^a |G||H| \, |d[X,Y]| \quad \leq \quad \left(\int_0^a G^2 \, d[X] \right)^{1/2} \left(\int_0^a H^2 \, d[Y] \right)^{1/2}, \tag{2.67}$$

$$E\Big[\int_0^a |G||H| \, |d[X,Y]| \Big] \quad \leq \quad \Big\| \Big(\int_0^a G^2 \, d[X] \Big)^{1/2} \Big\|_{L^p} \Big\| \Big(\int_0^a H^2 \, d[Y] \Big)^{1/2} \Big\|_{L^q}. \tag{2.68}$$

Hint: First consider elementary processes G, H.

Theorem 2.28 (**Covariation of stochastic integrals**). *For any $X, Y \in \mathcal{S}$ and any predictable process G that is integrable w.r.t. X,*

$$\left[\int G\,dX, Y\right] \quad = \quad \int G\,d[X,Y] \qquad \textit{almost surely.} \tag{2.69}$$

Remark. If X and Y are local martingales, and the angle-bracket process $\langle X, Y\rangle$ exists, then also

$$\left\langle \int G\,dX, Y\right\rangle \quad = \quad \int G\,d\langle X,Y\rangle \qquad \text{almost surely.}$$

Proof of Theorem 2.28. We only sketch the main steps briefly, cf. [32] for details. Firstly, one verifies directly that (2.69) 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

$$\left[\int G^n\,dX, Y\right] \quad = \quad \int G^n\,d[X,Y] \qquad \text{for any } \; n \in \mathbb{N}.$$

As $n \to \infty$, $\int G^n\,dX \to \int G\,dX$ in $\mathcal{H}^2$ by the Itô isometry for semimartingales, and hence

$$\left[\int G^n\,dX, Y\right] \quad \longrightarrow \quad \left[\int G\,dX, Y\right] \qquad \text{u.c.p.}$$

by Corollary 2.14. 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 (2.69) 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. $\square$

An important motivation for the extension of stochastic integrals to general predictable integrands is the validity of a Dominated Convergence Theorem:

Theorem 2.29 (**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*

$$G_t^n(\omega) \quad \longrightarrow \quad G_t(\omega) \qquad \textit{for any } \; t \geq 0, \quad \textit{almost surely},$$

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 [3] for basics, and to Jacod & Shiryaev [22] 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.

Local time

The occupation time of a Borel set $U \subseteq \mathbb{R}$ by a one-dimesional 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 a continuous semimartingale (X_t). 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_t^a 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*

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{(a,\infty)}(X_s)\,dX_s \;+\; \frac{1}{2}L_t^a$$

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

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{[a,\infty)}(X_s)\,dX_s \;+\; \frac{1}{2}\hat{L}_t^a$$

involving the right derivative $I_{[a,\infty)}$ instead of the left derivative $I_{(a,\infty)}$. Note that

$$L_t^a - \hat{L}_t^a \;=\; \int_0^t I_{\{a\}}(X_s)\,dX_s.$$

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 2.30. 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 2.30 (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*

$$L_t^a \;=\; \text{ucp}-\lim_{n\to\infty} \int_0^t \varphi_n(X_s)\,d[X]_s.$$

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:*

$$(X_t - a)^+ - (X_0 - a)^+ \;=\; \int_0^t I_{(a,\infty)}(X_s)\,dX_s \,+\, \frac{1}{2}L_t^a, \tag{2.70}$$

$$(X_t - a)^- - (X_0 - a)^- \;=\; -\int_0^t I_{(-\infty,a]}(X_s)\,dX_s \,+\, \frac{1}{2}L_t^a, \tag{2.71}$$

$$|X_t - a| - |X_0 - a| \;=\; \int_0^t \text{sgn}\,(X_s - a)\,dX_s \,+\, L_t^a, \tag{2.72}$$

where $\text{sgn}(x) := +1$ *for* $x > 0$, *and* $\text{sgn}(x) := -1$ *for* $x \leq 0$.

Remark. Note that we set $\text{sgn}(0) := -1$. This is related to our convention of using left derivatives as $\text{sgn}(x)$ 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 $\widehat{\text{sgn}}(x) := +1$ for $x \geq 0$ and $\text{sgn}(x) := -1$ 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,

$$f_n(X_t) - f_n(X_0) - \int_0^t f_n'(X_s)\, dX_s \;=\; \frac{1}{2}\int_0^t \varphi_n(X_s)\, d[X]_s. \tag{2.73}$$

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 2.29. Moreover,

$$f_n(X_t) - f_n(X_0) \;\to\; (X_t - a)^+ - (X_0 - a)^+.$$

The first assertion now follows from (2.73).

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 $x \notin (a, a + 1/n)$,

$$\int_0^t I_{\{|X_s - a| > \varepsilon\}}\, dL_s^a \;\leq\; \liminf_{n\to\infty} \int_0^t I_{\{|X_s - a| > \varepsilon\}}\, \varphi_n(X_s)\, d[X]_s \;=\; 0$$

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 (2.70) holds by definition of L^a. Moreover, subtracting (2.71) from (2.70) yields

$$(X_t - a) - (X_0 - a) \;=\; \int_0^t dX_s,$$

which is a valid equation. Therefore, the formulae (2.71) and (2.70) are equivalent. Finally, (2.72) follows by adding (2.70) and (2.71). □

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 (2.72), X is a semi-martingale with decomposition

$$X_t \;=\; W_t \;+\; L_t \tag{2.74}$$

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 4.2. We now compute the local time L_t^X of X at 0. By (2.71) and Lemma 2.30, 2),

$$\begin{aligned} \frac{1}{2} L_t^X \;&=\; X_t^- - X_0^- + \int_0^t I_{(-\infty,0]}(X_s)\, dX_s \qquad (2.75)\\ &=\; \int_0^t I_{\{0\}}(B_s)\, dW_s + \int_0^t I_{\{0\}}(B_s)\, dL_s \;=\; \int_0^t dL_s \;=\; L_t \qquad \text{a.s.,} \end{aligned}$$

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 (2.74) and (2.75), 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 (2.74) 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 $k_t = \sup_{s \le t}(-y_s)$.

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 $S_t := \sup_{s \le t} B_s$ have the same law. In particular, $L_t \sim |B_t|$ for any $t \ge 0$.

c) More generally, show that the two-dimensional processes $(|B|, L)$ and $(S - B, S)$ have the same law.

Notice that the maximum process $(S_t)_{t \ge 0}$ is the generalized inverse of the Lévy subordinator $(T_a)_{a \ge 0}$ introduced in Section 1.1. 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

$$f(\lambda x + (1-\lambda)y) \;\le\; \lambda f(x) + (1-\lambda) f(y) \qquad \forall\, \lambda \in [0,1],\ x, y \in \mathbb{R}.$$

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 3.10 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 2.31 (Itô-Tanaka formula, Meyer). *Suppose that X is a continuous semimartingale, and $f : \mathbb{R} \to \mathbb{R}$ is convex. Then*

$$f(X_t) - f(X_0) \;=\; \int_0^t f'_-(X_s)\,dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a\, f''(da). \tag{2.76}$$

Proof. We proceed in several steps:

1) Equation (2.76) holds for linear functions f.

2) By localization, we may assume that $|X_t| < C$ for a finite constant C. Then both sides of (2.76) 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.,

$$\operatorname{supp}(f'') \;\subseteq\; [-C,C].$$

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

$$f'_-(y) = \mu(-\infty,y) \qquad \text{and} \qquad f(x) = \int_{-\infty}^x \mu(-\infty,y)\,dy \tag{2.77}$$

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 (2.76) holds by definition of L^a. More generally, by linearity, (2.76) 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.,

$$f_n(X_t) - f_n(X_0) \;=\; \int_0^t f'_{n-}(X_s)\,dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a\,\mu_n(da) \tag{2.78}$$

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 (2.76) for f now follows from (2.78) as $n \to \infty$. □

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 2.32. *For any measurable function* $V : \mathbb{R} \to [0, \infty)$,

$$\int_{\mathbb{R}} L_t^a\; V(da) \;=\; \int_0^t V(X_s)\; d[X]_s \qquad \forall\, t \geq 0. \tag{2.79}$$

Proof. The assertion 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 (2.79) is the occupation time of the set B by (X_t), measured w.r.t. the quadratic variation $d[X]_t$.

Chapter 3

SDE I: Strong solutions and approximation schemes

In this chapter we study strong solutions of stochastic differential equations and the corresponding stochastic flows. We start with a crucial martingale inequality that is used frequently to derive L^p estimates for semimartingales. For real-valued càdlàg functions $x = (x_t)_{t\geq 0}$ we set

$$x_t^\star \quad := \quad \sup_{s<t} |x_s| \qquad \text{for } \ t > 0, \qquad \text{and} \qquad x_0^\star \ := \ |x_0|.$$

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

$$c_p \cdot E[[M]_\infty^{p/2}] \quad \leq \quad E[(M_\infty^\star)^p] \quad \leq \quad C_p \cdot E[[M]_\infty^{p/2}] \tag{3.1}$$

hold for any continuous local martingale M satisfying $M_0 = 0$, cf. [33]. 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 (3.1) 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 3.2 below.

Here, we only prove an easy special case of the Burkholder-Davis-Gundy inequality that will be sufficient for our purposes. This estimate also holds for càdlàg local martingales:

Theorem 3.1 (Burkholder's inequality). *Let $p \in [2,\infty)$. Then the estimate*

$$E[(M_T^\star)^p]^{1/p} \quad \leq \quad \gamma_p \, E[[M]_T^{p/2}]^{1/p} \tag{3.2}$$

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 $p \to \infty$.

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 $\varphi''(x) = p(p-1)|x|^{p-2}$, Itô's formula implies

$$\begin{aligned}|M_\infty|^p \quad = \quad & \int_0^\infty \varphi'(M_{s-})\, dM_s + \frac{1}{2}\int_0^\infty \varphi''(M_s)\, d[M]_s^c \\ & + \sum_s \left(\varphi(M_s) - \varphi(M_{s-}) - \varphi'(M_{s-})\Delta M_s, \right),\end{aligned} \tag{3.3}$$

where the first term is a martingale since $\varphi' \circ M$ is bounded, in the second term

$$\varphi''(M_s) \quad \leq \quad p(p-1)(M_\infty^\star)^{p-2},$$

and the summand in the third term can be estimated by

$$\begin{aligned}\varphi(M_s) - \varphi(M_{s-}) - \varphi'(M_{s-})\Delta M_s \quad & \leq \quad \frac{1}{2}\sup(\varphi'' \circ M)(\Delta M_s)^2 \\ & \leq \quad \frac{1}{2} p(p-1)(M_\infty^\star)^{p-2}(\Delta M_s)^2.\end{aligned}$$

Hence by taking expectation values on both sides of (3.3), we obtain for q satisfying $\frac{1}{p}+\frac{1}{q}=1$:

$$\begin{aligned}
E[(M_\infty^\star)^p] &\leq q^p\, E[|M_\infty|^p] \\
&\leq q^p\, \frac{p(p-1)}{2}\, E\Big[(M_\infty^\star)^{p-2}\Big([M]_\infty^c + \sum(\Delta M)^2\Big)\Big] \\
&\leq q^p\, \frac{p(p-1)}{2}\, E[(M_\infty^\star)^p]^{\frac{p-2}{p}}\, E[[M]_\infty^{\frac{p}{2}}]^{\frac{2}{p}}
\end{aligned}$$

by Doob's inequality, Hölder's inequality, and since $[M]_\infty^c + \sum(\Delta M)^2 = [M]_\infty$. The inequality (3.2) now follows by noting that $q^p p(p-1) = q^{p-1}p^2$.

2) For $T=\infty$ and a strict local martingale $M \in M_{\text{loc}}$, 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

$$E[(M_{T_n}^\star)^p] \quad = \quad E[(M_\infty^{T_n,\star})^p] \quad \leq \quad \gamma_p^p\, E[[M^{T_n}]_\infty^{p/2}] \quad = \quad \gamma_p^p\, E[[M]_{T_n}^{p/2}].$$

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 (3.1) 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 $c,t,x \in [0,\infty)$.

3.1 Existence and uniqueness of strong solutions

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

$$dX_t \quad = \quad b_t(X)\ dt + \sigma_t(X)\ dB_t + \int_{y \in S} c_{t-}(X, y)\ \widetilde{N}(dt\ dy). \qquad (3.4)$$

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 $\mathcal{B}_t := \sigma(x \mapsto x_s : s \leq t)$, $\mathcal{B}_t \otimes \mathcal{S}$ respectively for any $t \geq 0$. We also assume **local boundedness** of the coefficients, i.e.,

$$\sup_{s<t}\ \sup_{x : x_t^\star < r}\ \sup_{y \in S}\ \left(|b_s(x)| + \|\sigma_s(x)\| + |c_s(x, y)| \right) \ < \ \infty \qquad (3.5)$$

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 $(x_s)_{s \leq t}$ 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 (3.5) implies in particular that the jump sizes are locally bounded. Locally unbounded jumps could be taken into account by extending the SDE (3.4) 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 (3.4) 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,*

$$X_t \;=\; X_0 + \int_0^t b_s(X)\,ds + \int_0^t \sigma_s(X)\,dB_s + \int_{(0,t]\times S} c_{s-}(X,y)\,\widetilde{N}(ds\,dy) \quad (3.6)$$

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 4.1, we will see an example of a solution to an SDE that does not possess this property.

Remark. The stochastic integrals in (3.6) 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 (3.4) 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:*

$$|b_t(x) - b_t(\widetilde{x})| + ||\sigma_t(x) - \sigma_t(\widetilde{x})|| + \|c_t(x, \bullet) - c_t(\widetilde{x}, \bullet)\|^2_{L^2(\nu)} \quad \leq \quad L \cdot \sup_{s \leq t} |x_s - \widetilde{x}_s|$$

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 (3.4) that is crucial for studying existence, uniqueness, and dependence on the initial condition:

Theorem 3.2 (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 (3.4) taking values in U for $t < T$, and let*

$$\varphi(t) \quad := \quad E\left[\sup_{s < t \wedge T} |X_s - \widetilde{X}_s|^p\right].$$

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 $t \leq t_0$,

$$\varphi(t) \quad \leq \quad C \cdot \Big(\varphi(0) + \int_0^t \varphi(s)\, ds\Big), \quad \textit{and} \tag{3.7}$$

$$\varphi(t) \quad \leq \quad C \cdot e^{Ct}\, \varphi(0). \tag{3.8}$$

Proof. We only prove the assertion for $p = 2$. For $p > 2$, the proof can be carried out essentially in a similar way by relying on Burkholder's inequality instead of Itô's isometry.

Clearly, (3.8) follows from (3.7) by Gronwell's lemma. To prove (3.7), 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 $t \leq t_0$,

$$(X - \widetilde{X})^\star_{t \wedge T} \quad \leq \quad \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}, \qquad \text{where} \tag{3.9}$$

$$\begin{aligned}
\mathrm{I} &= |X_0 - \widetilde{X}_0|, \\
\mathrm{II} &= \int_0^{t\wedge T} |b_s(X) - b_s(\widetilde{X})|\, ds, \\
\mathrm{III} &= \sup_{u<t\wedge T} \Big| \int_0^u (\sigma_s(X) - \sigma_s(\widetilde{X}))\, dB_s \Big|, \quad \text{and} \\
\mathrm{IV} &= \sup_{u<t\wedge T} \Big| \int_{(0,u]\times S} (c_{s-}(X,y) - c_{s-}(\widetilde{X},y))\, \widetilde{N}(ds\, dy) \Big|.
\end{aligned}$$

The squared L^2-norms of the first two expressions are bounded by

$$\begin{aligned}
E[I^2] &= \varphi(0), \quad \text{and} \\
E[\mathrm{II}^2] &\leq L^2 t\, E\Big[\int_0^{t\wedge T} (X - \widetilde{X})_s^{\star\, 2}\, ds \Big] \leq L^2 t \int_0^t \varphi(s)\, ds.
\end{aligned}$$

Denoting by M_u and K_u the stochastic integrals in III and IV respectively, Doob's inequality and Itô's isometry imply

$$\begin{aligned}
E[\mathrm{III}^2] &= E[M_{t\wedge T}^{\star\, 2}] \leq 4E[M_{t\wedge T}^2] \\
&= 4E\Big[\int_0^{t\wedge T} ||\sigma_s(X) - \sigma_s(\widetilde{X})||^2\, ds \Big] \leq 4L^2 \int_0^t \varphi(s)\, ds,
\end{aligned}$$

$$\begin{aligned}
E[\mathrm{IV}^2] &= E[K_{t\wedge T}^{\star\, 2}] \leq 4E[K_{t\wedge T}^2] \\
&= 4E\Big[\int_0^{t\wedge T} \int |c_{s-}(X,y) - c_{s-}(\widetilde{X},y)|^2\, \nu(dy)\, ds \Big] \leq 4L^2 \int_0^t \varphi(s)\, ds.
\end{aligned}$$

The assertion now follows since by (3.9),

$$\varphi(t) = E\big[(X - \widetilde{X})_{t\wedge T}^{\star\, 2}\big] \leq 4 \cdot E\big[\mathrm{I}^2 + \mathrm{II}^2 + \mathrm{III}^2 + \mathrm{IV}^2\big].$$

□

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 3.3 (Pathwise uniqueness). *Suppose that Assumption (A1) holds. If (X_t) and $(\widetilde{X}_t)$ are strong solutions of (3.1) with $X_0 = \widetilde{X}_0$ almost surely then*

$$P\big[X_t = \widetilde{X}_t \quad \textit{for any } t\big] = 1.$$

Proof. For any open bounded set $U \subset \mathbb{R}^n$ and $t_0 \in \mathbb{R}_+$, the a priori estimate in Theorem 3.2 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 $x \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n)$,

$$\sup_{t<t_0} \left(|b_t(x)|^2 + ||\sigma_t(x)||^2 + \int |c_t(x,y)|^2 \, \nu(dy) \right) \quad \le \quad C(t_0) \cdot (1 + x^\star_{t_0})^2. \tag{3.10}$$

Theorem 3.4 (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 (3.1) has a strong solution $(X_t)_{t<\zeta}$ with initial condition $X_0 = \xi$ that is defined up to the explosion time*

$$\zeta \quad = \quad \sup T_k, \qquad \textit{where } \; T_k \;\; = \;\; \inf \{ t \ge 0 : |X_t| \ge k \}.$$

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 3.4. 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 (3.1) by a *Picard-Lindelöf iteration*, i.e., for $t \geq 0$ and $n \in \mathbb{Z}_+$ we define inductively

$$\begin{aligned} X_t^0 &:= \xi, \\ X_t^{n+1} &:= \xi + \int_0^t b_s(X^n)\,ds + \int_0^t \sigma_s(X^n)\,dB_s + \int\limits_{(0,t]\times S} c_{s-}(X^n,y)\,\widetilde{N}(ds\,dy). \end{aligned} \tag{3.11}$$

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 (3.1). 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 (3.10), 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 (3.11) 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 (3.1) 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 (3.1) that is defined up to the explosion time $\zeta = \sup T_k$. □

Non-explosion criteria

Theorem 3.4 shows that under a global Lipschitz and linear growth condition on the coefficients, the solution to (3.1) 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

$$dX_t \quad = \quad b(X_t)\,dt + \sigma(X_t)\,dB_t \tag{3.12}$$

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 3.5 (Lyapunov condition for non-explosion). *Suppose that there exists a function $\varphi \in C^2(\mathbb{R}^n)$ such that*

(i) $\varphi(x) \geq 0$ *for any* $x \in \mathbb{R}^n$,

(ii) $\varphi(x) \to \infty$ *as* $|x| \to \infty$, *and*

(iii) $\mathcal{L}\varphi \leq \lambda\varphi$ *for some* $\lambda \in \mathbb{R}_+$.

Then the strong solution of (3.1) with initial value $x_0 \in \mathbb{R}^n$ exists up to $\zeta = \infty$ almost surely.

Proof. We first remark that by (iii), $Z_t := \exp(-\lambda t)\varphi(X_t)$ 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,

$$dZ \;=\; -\lambda e^{-\lambda t}\varphi(X)\,dt \;+\; e^{-\lambda_t}\,d\varphi(X) \;=\; dM \;+\; e^{-\lambda t}(\mathcal{L}\varphi - \lambda\varphi)(X)\,dt$$

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),

$$\begin{aligned} \varphi(x_0) \quad &= \quad E\left[\varphi(X_0)\right] \;\geq\; E\left[\exp(-\lambda(t\wedge T_k))\,\varphi(X_{t\wedge T_k})\right] \\ &\geq \quad E\left[\exp(-\lambda t)\,\varphi(X_{T_k})\,;\,T_k \leq t\right] \\ &\geq \quad \exp(-\lambda t)\;\inf_{|y|=k}\varphi(y)\;P\left[T_k \leq t\right] \end{aligned}$$

for any $k \in \mathbb{N}$. As $k \to \infty$, $\inf_{|y|=k}\varphi(y) \to \infty$ 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 3.6. *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

$$\frac{x}{|x|} \cdot b(x) \quad \leq \quad \mathrm{const.} \cdot |x| \qquad \text{and} \qquad \mathrm{tr}\, a(x) \leq \mathrm{const.} \cdot |x|^2$$

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.

3.2 Stochastic flow and Markov property

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

$$dX_t \quad = \quad b_t(X)\, dt + \sigma_t(X)\, dW_t, \qquad X_0 \;=\; a, \tag{3.13}$$

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

$$|b_t(x) - b_t(\widetilde{x})| + ||\sigma_t(x) - \sigma_t(\widetilde{x})|| \quad \leq \quad L\, (x - \widetilde{x})_t^\star \qquad \forall\, t, x, \widetilde{x} \tag{3.14}$$

for some finite constant $L \in \mathbb{R}_+$, as well as

$$\sup_{s\in[0,t]} \big(|b_s(0)| + ||\sigma_s(0)||\big) \quad < \quad \infty \qquad \forall\, t. \tag{3.15}$$

Then by Itô's existence and uniqueness theorem, there exists a unique global strong solution $(X_t^a)_{t\geq 0}$ of (3.13) 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 Kolmogorov's continuity criterion for stochastic processes:

Theorem 3.7 (Kolmogorov, Čentsov). *Suppose that* $(E, ||\cdot||)$ *is a Banach space,* $C = \prod_{k=1}^{n} I_k$ *is a product of real intervals* $I_1, \dots, I_n \subset \mathbb{R}$*, 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* $\gamma, c, \varepsilon \in \mathbb{R}_+$ *such that*

$$E\big[||X_u - X_v||^\gamma\big] \quad \leq \quad c|u-v|^{n+\varepsilon} \qquad \text{for any } u, v \in C, \tag{3.16}$$

then there exists a modification $(\xi_u)_{u\in C}$ *of* $(X_u)_{u\in C}$ *such that*

$$E\Big[\Big(\sup_{u\neq v} \frac{||\xi_u - \xi_v||}{|u-v|^\alpha}\Big)^\gamma\Big] \quad < \quad \infty \qquad \text{for any } \alpha \in [0, \varepsilon/\gamma). \tag{3.17}$$

In particular, $u \mapsto \xi_u$ *is almost surely* α*-Hölder continuous for any* $\alpha < \varepsilon/\gamma$*.*

For the proof cf. e.g. [33, Ch. I, (2.1)].

Example. Brownian motion satisfies (3.16) 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. [13, Thm. 1.20].

Existence of a continuous flow

We now apply the Kolmogorov-Čentsov continuity criterion to the solution $a \mapsto (X_s^a)$ of the SDE (3.13) as a function of its starting point.

Theorem 3.8 (Flow of an SDE). *Suppose that (3.14) and (3.15) 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 (3.13) 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***

$$\xi_{t+s}^a(\omega) \quad = \quad \xi_s^{\xi_t^a(\omega)}(\Theta_t(\omega)) \qquad \forall\, s, t \geq 0, \;\; a \in \mathbb{R}^n \tag{3.18}$$

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

$$\xi(\omega) \quad := \quad \xi(\omega - \omega(0)) \tag{3.19}$$

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 3.2 there exists a finite constant $c \in \mathbb{R}_+$ such that

$$E[(X^a - X^{\widetilde{a}})_t^{\star\, p}] \quad \le \quad c \cdot e^{ct}\, |a - \widetilde{a}|^p \qquad \text{for any } t \ge 0 \text{ and } a, \widetilde{a} \in \mathbb{R}^n, \tag{3.20}$$

where X^a denotes a version of the strong solution of (3.13) 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 (3.20), there exists a modification ξ of $(X_s^a)_{s \le t, a \in \mathbb{R}^n}$ such that $a \mapsto (\xi_s^a)_{s \le t}$ is almost surely α-Hölder continuous w.r.t. $\|\cdot\|_t$ for any $\alpha < \frac{p-n}{p}$. Clearly, for $t_1 \le t_2$, 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 $t \ge 0$ and $a \in \mathbb{R}^n$. Then μ_0-almost surely, both sides of (3.18) solve the same SDE as a function of s. Indeed,

$$\begin{aligned}
\xi_{t+s}^a \quad &= \quad \xi_t^a + \int_t^{t+s} \widetilde{b}(\xi_u^a)\, du + \int_t^{t+s} \widetilde{\sigma}(\xi_u^a)\, dW_u \\
&= \quad \xi_t^a + \int_0^s \widetilde{b}(\xi_{t+r}^a)\, dr + \int_0^s \widetilde{\sigma}(\xi_{t+r}^a)\, d\left(W_r \circ \Theta_t\right),
\end{aligned}$$

$$\xi_s^{\xi_t^a} \circ \Theta_t \quad = \quad \xi_t^a + \int_0^s \widetilde{b}\big(\xi_r^{\xi_t^a} \circ \Theta_t\big)\, dr + \int_0^s \widetilde{\sigma}(\xi_r^{\xi_t^a} \circ \Theta_t)\, d(W_r \circ \Theta_t)$$

hold μ_0-almost surely for any $s \ge 0$ 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

$$\xi_{t+s}^a \quad = \quad \xi_s^{\xi_t^a} \circ \Theta_t \qquad \text{for any } s \ge 0, \qquad \text{almost surely.}$$

Continuity of ξ then shows that the cocycle property (3.18) holds with probability one for all s, t and a simultaneously. $\Box$

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 5.1 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 3.8 extends to solutions of SDE of type (3.4) 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. [32]. 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 (3.13) on the canonical setup. From this we can obtain strong solutions on other setups:

Exercise. Show that the unique strong solution of (3.13) 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 (3.13) is a direct consequence of the cocycle property:

Corollary 3.9. *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 $0 \le s \le t$,

$$\Theta_t(\omega) \quad = \quad \omega(t) + \big(\omega(t+\cdot) - \omega(t)\big),$$

and hence, by the cocycle property and by (3.19),

$$f(\xi^a_{s+t}(\omega)) \quad = \quad f\big(\xi_s^{\xi_t^a(\omega)}\big(\omega(t+\cdot) - \omega(t)\big)\big)$$

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

$$\frac{1}{2}L_t^a \;=\; (X_0 - a)^+ - (X_t - a)^+ \;-\; \int_0^t I_{(a,\infty)}(X_s)\, dM_s \;-\; \int_0^t I_{(a,\infty)}(X_s)\, dA_s \quad (3.21)$$

almost surely for any $a \in \mathbb{R}$.

Theorem 3.10 (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*

$$L_t^a - L_t^{a-} \;=\; 2\int_0^t I_{\{X_s=a\}}\, dA_s. \quad (3.22)$$

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 $V^{(1)}_\infty(A)$. The map $(a,t) \mapsto (X_t - a)^+$ is jointly continuous in t and a. Moreover, by dominated convergence,

$$Z^a_t \;:=\; \int_0^t I_{(a,\infty)}(X_s)\,dA_s$$

is continuous in t and càdlàg in a with

$$Z^a_t - Z^{a-}_t \;=\; -\int_0^t I_{\{a\}}(X_s)\,dA_s.$$

Therefore it is sufficient to prove that

$$Y^a_t \;:=\; \int_0^t I_{(a,\infty)}(X_s)\,dM_s$$

has a version such that the map $a \mapsto (Y^a_s)_{s\le t}$ 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,

$$\begin{aligned} E\left[\left(Y^a - Y^b\right)^{\star\,p}_t\right] \;&=\; E\left[\sup_{s<t}\left|\int_0^s I_{(a,b]}(X)\,dM\right|^p\right] \\ &\le\; C_1(p)\,E\left[\left|\int_0^t I_{(a,b]}(X)\,d[M]\right|^{p/2}\right] \end{aligned} \tag{3.23}$$

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.,

$$\begin{aligned} \int_0^t I_{(a,b]}(X)\,d[M] \;&=\; \int_0^t I_{(a,b]}(X)\,d[X] \\ &=\; -2\left(f(X_t) - f(X_0) - \int_0^t f'(X)\,dX\right) \\ &\le\; (b-a)^2 \,+\, 2\left|\int_0^t f'(X)\,dM\right| \,+\, |b-a|\,V^{(1)}_t(A) \end{aligned}$$

Here we have used in the last step that $|f'| \leq |b-a|$ and $|f| \leq (b-a)^2/2$. Combining this estimate with 3.23 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. $\square$

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 (3.22).

3.3 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. [32] 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*

$$\int_0^t X_s \circ dY_s \quad := \quad \int_0^t X_s\, dY_s + \frac{1}{2}[X,Y]_t \qquad \textit{for any } t \geq 0.$$

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 3.11. *If (π_n) is a sequence of partitions of $\mathbb{R}_+$ with $\mathrm{mesh}(\pi_n) \to 0$ then*

$$\int_0^t X_s \circ dY_s \quad = \quad \lim_{n\to\infty} \sum_{\substack{s\in\pi_n \\ s<t}} \frac{X_s + X_{s'\wedge t}}{2}\,(Y_{s'\wedge t} - Y_s) \qquad \textit{in the } \mathrm{ucp} \textit{ sense}.$$

Proof. This follows since $\int_0^t X\, dY = \mathrm{ucp}\text{-}\lim \sum_{s<t} X_s\, (Y_{s'\wedge t} - Y_s)$ and $[X,Y]_t = \mathrm{ucp}\text{-}\lim \sum_{s<t} (X_{s'\wedge t} - X_s)(Y_{s'\wedge t} - Y_s)$ by the results above. □

Itô-Stratonovich formula

For Stratonovich integrals w.r.t. continuous semimartingales, the classical chain rule holds:

Theorem 3.12. *Let $X = (X^1, \ldots, X^d)$ with continuous semimartingales X^i. Then for any function $F \in C^2(\mathbb{R}^d)$,*

$$F(X_t) - F(X_0) \quad = \quad \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x^i}(X_s) \circ dX_s^i \qquad \forall\, t \geq 0. \tag{3.24}$$

Proof. To simplify the proof we assume $F \in C^3$. Under this condition, (3.24) is just a reformulation of the Itô rule

$$F(X_t) - F(X_0) \quad = \quad \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x^i}(X_s)\, dX_s^i + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s)\, d[X^i, X^j]_s. \tag{3.25}$$

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 (3.24) and the Itô integral in (3.25) 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. [32], where also a generalization to càdlàg semimartingales is considered.

The product rule for Stratonovich integrals is a special case of the chain rule:

Corollary 3.13. *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

$$\circ\, dX_t \quad = \quad b(X_t)\,dt + \sum_{k=1}^{d} \sigma_k(X_t) \circ dB_t^k, \qquad X_0 \; = \; x_0, \tag{3.26}$$

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 (3.26) is equivalent to the Itô SDE

$$dX_t \quad = \quad \widetilde{b}(X_t)\,dt + \sum_{k=1}^{d} \sigma_k(X_t)\,dB_t^k, \qquad X_0 \; = \; x_0, \tag{3.27}$$

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 (3.26).

Theorem 3.14 (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 (3.26) 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 (3.26),

$$\begin{aligned} F(X_t)-F(X_0) &= \int_0^t \nabla F(X)\cdot \circ dX \\ &= \int_0^t (b\cdot\nabla F)(X)\,dt+\sum_k\int_0^t(\sigma_k\cdot\nabla F)(X)\circ dB^k. \end{aligned} \tag{3.28}$$

By applying this formula to $\sigma_k\cdot\nabla F$, we see that

$$(\sigma_k\cdot\nabla F)(X_t) \quad = \quad A_t+\sum_l\int\sigma_l\cdot\nabla(\sigma_k\cdot\nabla F)(X)\,dB^l$$

with a continuous finite variation process (A_t). Hence

$$\begin{aligned} \int_0^t(\sigma_k\cdot\nabla F)(X)\circ dB^k &= \int_0^t(\sigma_k\cdot\nabla F)(X)\,dB^k+[(\sigma_k\cdot\nabla F)(X),B^k]_t \\ &= \text{local martingale }+\int_0^t\sigma_k\cdot\nabla(\sigma_k\cdot\nabla F)(X)dt. \end{aligned} \tag{3.29}$$

The assertion now follows by (3.28) and (3.29). □

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

$$\mathcal{L} \quad = \quad \frac{1}{2}\sum_k\sigma_k^2+b$$

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. [34], [18] and [20] 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

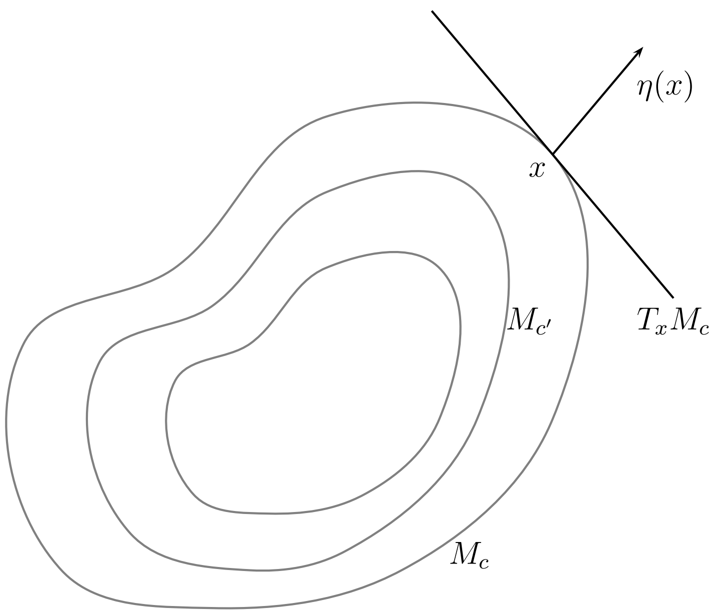
$$\mathfrak{n}(x) \quad = \quad \frac{\nabla f(x)}{|\nabla f(x)|} \quad \in S^n$$

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 $k \in \{1, \ldots, n+1\}$, 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*

$$\circ\, dX_t \quad = \quad P(X_t) \circ dB_t \quad = \quad \sum_{k=1}^{n+1} P_k(X_t) \circ dB_t^k, \qquad X_0 \;=\; x_0, \tag{3.30}$$

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 (3.30) 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 3.15. *If X is a solution of (3.30) 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. $\square$

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 3.14, 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 (3.30) 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

$$\circ\, dX_t \quad = \quad b(X_t)\, dt \;+\; \sigma(X_t) \circ dB_t, \qquad X_0 \;=\; a, \tag{3.31}$$

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 (3.31) is equivalent to the Itô SDE

$$dX_t \quad = \quad \big(b + \frac{1}{2}\sigma \cdot \nabla\sigma\big)(X_t)\, dt + \sigma(X_t)\, dB_t, \qquad X_0 = a. \tag{3.32}$$

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

$$F'(s) \quad = \quad \sigma(F(s)). \tag{3.33}$$

Hence a solution of (3.31) 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 (3.33) 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 (3.31) in the general case, we try the "variation of constants" ansatz

$$X_t \quad = \quad F(B_t, C_t) \tag{3.34}$$

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 (3.34), we obtain

$$\begin{aligned} \circ dX_t &= \frac{\partial F}{\partial s}(B_t, C_t) \circ dB_t \;+\; \frac{\partial F}{\partial x}(B_t, C_t) \circ dC_t \\ &= \sigma(X_t) \circ dB_t \;+\; \frac{\partial F}{\partial x}(B_t, C_t) \circ dC_t \end{aligned}$$

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 (3.31) provided (C_t) is almost surely absolutely continuous with derivative

$$\frac{d}{dt} C_t \quad = \quad \frac{\partial F}{\partial x}(B_t, C_t)^{-1}\, b(F(B_t, C_t)). \tag{3.35}$$

For every given ω, the equation (3.35) 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 3.16 (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 (3.35) 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 (3.31), (3.32) respectively.*

We refer to [23] for a detailed proof.

Exercise (Computing explicit solutions). Solve the following Itô stochastic differential equations explicitly:

$$dX_t \;=\; \frac{1}{2} X_t\, dt + \sqrt{1 + X_t^2}\, dB_t, \qquad X_0 \;=\; 0, \tag{3.36}$$

$$dX_t \;=\; X_t(1 + X_t^2)\, dt + (1 + X_t^2)\, dB_t, \qquad X_0 \;=\; 1. \tag{3.37}$$

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

$$\frac{dC_t(\omega)}{dt} \;=\; f(t, Z_t(\omega)\cdot C_t(\omega))/Z_t(\omega)\;; \qquad C_0 = x. \tag{3.38}$$

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 (3.38) 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

$$B^{(k)} \quad := \quad B \star \varphi_{1/k}, \qquad \varphi_\varepsilon(t) \quad = \quad (2\pi\varepsilon)^{-1/2} \exp\big(-\frac{t^2}{2\varepsilon}\big).$$

Other smooth approximations could be used as well, cf. [23] and [21]. Let $X^{(k)}$ denote the unique solution to the ordinary differential equation

$$\frac{d}{dt} X_t^{(k)} \quad = \quad b(X_t^{(k)}) + \sigma(X_t^{(k)}) \, \frac{d}{dt} B_t^{(k)}, \qquad X_0^{(k)} \ = \ a \tag{3.39}$$

with coefficients $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$.

Theorem 3.17 (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 (3.35) 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. [23]. The proof in the more interesting general case is much more involved, cf. e.g. Ikeda & Watanabe [21, Ch. VI, Thm. 7.2].

3.4 Stochastic Taylor expansions and numerical methods

The goal of this section is to analyse the convergence order of numerical schemes for Itô stochastic differential equations of type

$$dX_t \;=\; b(X_t)\,dt \;+\; \sum_{k=1}^{d} \sigma_k(X_t)\,dB_t \tag{3.40}$$

in $\mathbb{R}^N$, $N \in \mathbb{N}$. We will assume throughout this section 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 (3.40) satisfies

$$X_{t+h} \;=\; X_t \;+\; \int_t^{t+h} b(X_s)\,ds \;+\; \sum_{k=1}^{d} \int_t^{t+h} \sigma_k(X_s)\,dB_s^k \tag{3.41}$$

for any $t, h \geq 0$. By approximating $b(X_s)$ and $\sigma_k(X_s)$ in (3.41) 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 (3.40), 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

$$f(X_{t+h}) \;=\; f(X_t) \;+\; \sum_{k=0}^{d} \int_t^{t+h} (\mathcal{L}_k f)(X_s)\,dB_s^k \tag{3.42}$$

for any $t, h \geq 0$,where $B_t^0 = t$, $a = \sigma\sigma^T$,

$$\mathcal{L}_0 f \;=\; \frac{1}{2}\sum_{i,j=1}^{N} a^{ij}\frac{\partial^2 f}{\partial x^i \partial x^j} \;+\; b \cdot \nabla f, \qquad \text{and} \tag{3.43}$$

$$\mathcal{L}_k f \;=\; \sigma_k \cdot \nabla f, \qquad \text{for } k = 1, \ldots, d. \tag{3.44}$$

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 **??** below. The next lemma yields L^2 bounds on the remainder terms:

Lemma 3.18. *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*

$$E\left[\left(\int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{n-1}} G_{s_n} \, dB_{s_n}^{k_n} \cdots dB_{s_2}^{k_2} \, dB_{s_1}^{k_1}\right)^2\right] \; \leq \; \frac{h^{n+m(k)}}{n!} \sup_{s \in [t,t+h]} E\left[G_s^2\right]$$

for any $n \in \mathbb{N}$ and $k = (k_1, \ldots, k_n) \in \{0, 1, \ldots, d\}^n$, where

$$m(k) \; := \; |\{1 \leq i \leq n : k_i = 0\}|$$

denotes the number of integrations w.r.t. dt.

Proof. By Itô's isometry and the Cauchy-Schwarz inequality,

$$\begin{aligned} E\left[\left(\int_t^{t+h} G_s \, dB_s^k\right)^2\right] \; &\leq \; \int_t^{t+h} E\left[G_s^2\right] \, ds \qquad \text{for any } k \neq 0, \text{ and} \\ E\left[\left(\int_t^{t+h} G_s \, ds\right)^2\right] \; &\leq \; h \int_t^{t+h} E\left[G_s^2\right] \, ds. \end{aligned}$$

By iteratively applying these estimates we see that the second moment of the iterated integral in the assertion is bounded from above by

$$h^{m(k)} \int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{n-1}} E[G_{s_n}^2] \, ds_n \cdots ds_2 \, ds_1.$$

□

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,

$$E[f(B_{t+h})] - E[f(B_t)] \;=\; \int_t^{t+h} E[\frac{1}{2}\Delta f(B_s)]\, ds \;\le\; \frac{h}{2} \sup \Delta f$$

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 3.18 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 (3.40) where $b, \sigma_k : \mathbb{R}^N \to \mathbb{R}^N$, $k = 1, \dots, d$, 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.

Corollary 3.19 (Itô-Taylor expansion with remainder of order α). *Suppose that $\alpha = k/2$ for some $k \in \mathbb{N}$. If X is a solution of (3.40) with coefficients $b, \sigma_1, \dots, \sigma_d \in C_b^{\lfloor 2\alpha \rfloor}(\mathbb{R}^N, \mathbb{R}^N)$ then the following expansions hold for any $f \in C_b^{\lfloor 2\alpha+1 \rfloor}(\mathbb{R}^N)$:*

$$f(X_{t+h}) = \sum_{n<2\alpha} \sum_{k:n+m(k)<2\alpha} \left(\mathcal{L}_{k_n}\mathcal{L}_{k_{n-1}} \cdots \mathcal{L}_{k_1} f\right)(X_t) \times \tag{3.45}$$
$$\times \int_t^{t+h} \int_t^{s_1} \cdots \int_t^{s_{n-1}} G_{s_n}\, dB_{s_n}^{k_n} \cdots dB_{s_2}^{k_2}\, dB_{s_1}^{k_1} + \mathcal{O}(h^\alpha),$$

$$E\left[f(X_{t+h})\right] = \sum_{n<\alpha} E\left[\left(\mathcal{L}_0^n f\right)(X_t)\right] \frac{h^n}{n!} + O(h^\alpha). \tag{3.46}$$

Proof. Iteration of the Itô-Doeblin formula (3.42) shows that (3.45) 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 $k = (k_1, \dots, k_n)$ satisfying $n + m(k) > 2\alpha$ and $n - 1 + m(k_1, \dots, k_{n-1}) < 2\alpha$. By Lemma 3.18, 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 (3.46) 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 (3.45) 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. (2.58). 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

$$I_h^{kl} \;=\; \int_0^h \int_0^s dB_r^k\, dB_s^l \;=\; B_s^k\, dB_s^l$$

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. [17, 26].

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

$$f_{A_t}(x) = \frac{1}{2t\cosh(\frac{\pi x}{2t})}\,.$$

Numerical schemes for SDE

Let X be a solution of the SDE (3.40), (3.41) respectively. 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

$$\begin{aligned}X_{t+h}-X_t &= b(X_t)\,h + \sum_{k=1}^d \sigma_k(X_t)\,(B^k_{t+h}-B^h_t) \\ &\quad + \sum_{k,l=1}^d (\sigma_l\cdot\nabla\sigma_k)\,(X_t)\int_t^{t+h}\int_t^s dB^l_r\,dB^k_s + \mathcal{O}\left(h^{3/2}\right).\end{aligned} \tag{3.47}$$

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** $\mathbf{h}$

$$X^h_{t+h}-X^h_t = b(X^h_t)\,h + \sum_{k=1}^d \sigma_k(X^h_t)\,(B^k_{t+h}-B^h_t) \qquad (t=0,h,2h,3h,\ldots)$$

- **Milstein scheme with step size** $\mathbf{h}$

$$X^h_{t+h}-X^h_t = b(X^h_t)\,h + \sum_{k=1}^d \sigma_k(X^h_t)\,(B^k_{t+h}-B^h_t) + \sum_{k,l=1}^d (\sigma_l\cdot\nabla\sigma_k)\,(X^h_t)\int_t^{t+h}\int_t^s dB^l_r\,dB^k_s$$

The Euler and Milstein scheme provide approximations to the solution of the SDE (3.40) that are defined for integer multiples t of the step size h. To analyse the 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

$$dX_s^h = b(X_{\lfloor s \rfloor_h}^h)\, ds + \sum_k \sigma_k(X_{\lfloor s \rfloor_h}^h)\, dB_s^k, \tag{3.48}$$

$$dX_s^h = b(X_{\lfloor s \rfloor_h}^h)\, ds + \sum_{k,l} \left(\sigma_k(X_{\lfloor s \rfloor_h}^h) + (\sigma_l \nabla \sigma_k)(X_{\lfloor s \rfloor_h}^h) \int_{\lfloor s \rfloor_h}^s dB_r^l \right) dB_s^k \tag{3.49}$$

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 (3.48) and (3.49) respectively at $t = kh$ with $k \in \mathbb{Z}$.

Strong convergence order

Fix $a \in \mathbb{R}^N$, let X be a solution of (3.40) with initial condition $X_0 = a$, and let X^h be a corresponding Euler or Milstein approximation satisfying (3.48, (3.49) respectively with initial condition $X_0^h = a$.

Theorem 3.20 (Strong order for Euler and Milstein scheme). *Let $t \in [0, \infty)$.*

1) *Suppose that the coefficients b and σ_k are Lipschitz continuous. Then the Euler-Maruyama approximation on the time interval $[0, t]$ has strong L^2 order $1/2$ in the following sense:*

$$\sup_{s \leq t} \left| X_s^h - X_s \right| = \mathcal{O}(h^{1/2}).$$

2) *If, moreover, the coefficients b and σ_k are C^3 with bounded derivatives then the Milstein approximation on the time interval $[0, t]$ has strong L^2 order 1, i.e.,*

$$\left| X_t^h - X_t \right| = \mathcal{O}(h).$$

The assumptions on the coefficients in the theorem are not optimal and can be weakened. 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 [**?**]XXX

Proof. For notational simplicity, we only prove the theorem in the one-dimensional case. The proof in higher dimensions is analogous.

1) By (3.48) and since $X_0^h = X_0$, the difference of the Euler approximation and the solution of the SDE satisfies the equation

$$X_t^h - X_t \;=\; \int_0^t \big(b(X^h_{\lfloor s\rfloor_h}) - b(X_s)\big)\; ds \;+\; \int_0^t \big(\sigma(X^h_{\lfloor s\rfloor_h}) - \sigma(X_s)\big)\; dB_s.$$

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,

$$\begin{aligned}
\bar{\varepsilon}_t^h \;&\le\; 2t\int_0^t E\left[\left|b(X^h_{\lfloor s\rfloor_h}) - b(X_s)\right|^2\right]\, ds \;+\; 8\int_0^t E\left[\left|\sigma(X^h_{\lfloor s\rfloor_h}) - \sigma(X_s)\right|^2\right]\, ds \\
&\le\; (2t+8)\cdot L^2\cdot \int_0^t E\left[\left|X^h_{\lfloor s\rfloor_h} - X_s\right|^2\right]\, ds \qquad (3.50)\\
&\le\; (4t+16)\cdot L^2\cdot \left(\int_0^t \bar{\varepsilon}_s^h\, ds \;+\; C_t\, h\right),
\end{aligned}$$

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,

$$E\left[\left|X^h_{\lfloor s\rfloor_h} - X_s\right|^2\right] \;\le\; 2\,E\left[\left|X^h_{\lfloor s\rfloor_h} - X_s^h\right|^2\right] \;+\; 2\,E\left[\left|X_s^h - X_s\right|^2\right],$$

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 Corollary 3.19. By (3.50) and Gronwall's inequality, we conclude that

$$\bar{\varepsilon}_t^h \;\le\; (4t+16)L^2 C_t\cdot \exp\big((4t+16)L^2 t\big)\cdot h,$$

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. By (3.49), the difference of the Milstein approximation and the solution of the SDE satisfies the equation

$$\begin{aligned} X_t - X_t^h &= \int_0^t \left(b(X_s) - b(X^h_{\lfloor s\rfloor_h})\right) ds \\ &\quad + \int_0^t \left(\sigma(X_s) - \sigma(X^h_{\lfloor s\rfloor_h})\right) - (\sigma\sigma')(X^h_{\lfloor s\rfloor_h})(B_s - B_{\lfloor s\rfloor_h})\, dB_s \\ &= I_t + II_t + III_t + IV_t \end{aligned} \tag{3.51}$$

where

$$\begin{aligned} I_t &= \int_0^t \left(b(X_s) - b(X_s^h)\right) ds\,, \\ II_t &= \int_0^t \left(\sigma(X_s) - \sigma(X_s^h)\right) dB_s\,, \\ III_t &= \int_0^t \left(b(X_s^h) - b(X^h_{\lfloor s\rfloor_h})\right) ds\,, \\ IV_t &= \int_0^t \left(\sigma(X_s^h) - \sigma(X^h_{\lfloor s\rfloor_h}) - (\sigma\sigma')(X^h_{\lfloor s\rfloor_h})(B_s - B_{\lfloor s\rfloor_h})\right) dB_s\,. \end{aligned}$$

Here terms I and II describe the continuation of the error made in previous time, whereas III and IV correspond to an additional error by the time discretization on the interval $[\lfloor s\rfloor_h, s]$. We are now going to bound the second moments of I, II, III and IV separately. Similarly as in the Euler case, we obtain by the Cauchy-Schwarz inequality and by Doob's L^2 inequality:

$$E\left[I_t^2\right] \;\leq\; t\int_0^t E\left[\left|b(X_s) - b(X_s^h)\right|^2\right] ds \;\leq\; L^2 t\int_0^t \varepsilon_s^h\, ds, \quad \text{and} \tag{3.52}$$

$$E\left[II_t^2\right] \;\leq\; 4t\int_0^t E\left[\left|\sigma(X_s) - \sigma(X_s^h)\right|^2\right] ds \;\leq\; 4L^2 t\int_0^t \varepsilon_s^h\, ds, \tag{3.53}$$

where

$$\varepsilon_s^h \;:=\; E\left[\left|X_s^h - X_s\right|^2\right].$$

We are now going to prove that III and IV are of strong L^2 order $\mathcal{O}(h)$ uniformly on finite time intervals. The proof can then be completed similarly as in the Euler case.

To bound IV we note that by (3.49) and Itô's formula,

$$\begin{aligned}
&\sigma(X^h_s) - \sigma(X^h_{\lfloor s\rfloor_h}) - \int_{\lfloor s\rfloor_h}^s \sigma(X^h_{\lfloor s\rfloor_h})\sigma'(X^h_r)\,dB_r \\
= & \int_{\lfloor s\rfloor_h}^s \left\{ b(X^h_{\lfloor s\rfloor_h})\sigma'(X^h_r) + \frac{1}{2}\left(\sigma(X^h_{\lfloor s\rfloor_h}) + (\sigma\sigma')(X^h_{\lfloor s\rfloor_h})(B_r - B_{\lfloor s\rfloor_h})\right)^2 \sigma''(X^h_r)\right\}\,dr.
\end{aligned}$$

Since all the coefficients and their derivatives up to order 3 are assumed to be bounded, and, in particular, σ' is Lipschitz continuous, we can conclude that

$$\begin{aligned}
&\left|\sigma(X^h_s) - \sigma(X^h_{\lfloor s\rfloor_h}) - (\sigma\sigma')(X^h_{\lfloor s\rfloor_h})\,(B_s - B_{\lfloor s\rfloor_h})\right| \\
\leq\;& C_1 \cdot (1 + \left|B_s - B_{\lfloor s\rfloor_h}\right|^2 + C_2 \cdot \int_{\lfloor s\rfloor_h}^s 1 + \left|B_r - B_{\lfloor r\rfloor_h}\right|^2\,dr
\end{aligned}$$

with constants C_1 and C_2 that do not depend on s. Hence

$$E\left[IV_t^2\right] \;\leq$$

LETZTER SCHRITT STIMMT NOCH NICHT GANZ !

□

Chapter 4

SDE II: Transformations and weak solutions

Let $U \subseteq \mathbb{R}^n$ be an open set. We consider a stochastic differential equation of the form

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{4.1}$$

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 (4.1) 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 (4.1) w.r.t. *some* Brownian motion (B_t) that is usually *defined through (4.1)*.

Definition. *A* ***weak solution*** *of the stochastic differential equation (4.1) 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* $(\mathcal{F}_t)$ *Brownian motion* $B_t : \Omega \to \mathbb{R}^d$ *w.r.t.* P*,*

(ii) a continuous $(\mathcal{F}_t)$ adapted stochastic process $(X_t)_{t<S}$ where S is an (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.*

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 process $(X_t)_{t\geq 0}$ 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 Section4.1. 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 (B_t) adapted stochastic processes defined on $C(\mathbb{R}_+, \mathbb{R}^n)$, as well as SDE driven by Poisson point processes, cf Chapter 3.

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

$(X_t) \to (\varphi(X_t))$, 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 this chapter we study these transformations as well as relations between them. For identifying the transformed processes, the Lévy characterizations in Section 4.1 play a crucial rôle.

4.1 Lévy characterizations

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 4.1. *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 1.3, 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 $\exp(-(t-s)\psi)$. $\square$

Exercise (Lévy 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

$$M_t^f \quad = \quad \exp(iN_t^f + t\psi(f)), \qquad t \geq 0, \quad \psi(f) \quad = \quad \int \left(1 - e^{if}\right) d\nu,$$

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 4.1, an $\mathbb{R}^d$-valued process (X_t) is a Brownian motion if and only if the processes $\exp\left(ip \cdot X_t + t|p|^2/2\right)$ 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 4.2 (P. Lévy 1948). *Suppose that $M^1, \ldots, M^d$ are continuous local $(\mathcal{F}_t)$ martingales with*

$$[M^k, M^l]_t \quad = \quad \delta_{kl} t \qquad P\text{-a.s. for any } t \geq 0.$$

Then $M = (M^1, \ldots, M^d)$ 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,

$$\begin{aligned} d\Phi_t &= ip\, \Phi_t \cdot dM_t - \frac{1}{2} \sum_{k,l=1}^{d} \Phi_t\, p_k p_l\, d[M^k, M^l]_t \\ &= ip\, \Phi_t \cdot dM_t - \frac{1}{2} \Phi_t\, |p|^2\, dt. \end{aligned}$$

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 4.1, 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

$$dX_t = O_t\, dB_t, \qquad X_0 = x_0, \tag{4.2}$$

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

$$O_t(\omega)\, O_t(\omega)^T = I_n \qquad \text{for any } t \geq 0, \quad \text{almost surely}. \tag{4.3}$$

Indeed, by (4.2) and (4.3), the components

$$X_t^i = x_0^i + \sum_{k=1}^{d} \int_0^t O_s^{ik}\, dB_s^k$$

are continuous local martingales with covariations

$$[X^i, X^j] = \sum_{k,l} \int O^{ik}\, O^{jl}\, d[B^k, B^l] = \int \sum_k O^{ik}\, O^{jk}\, dt = \delta_{ij}\, dt.$$

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

$$dR_t \quad = \quad dW_t \, + \, \frac{d-1}{2R_t}\, dt \tag{4.4}$$

up to the first hitting time of 0. The equation (4.4) 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 [33] and other works by M. Yor.

Exercise. a) Let $(X_t)_{0 \leq t < \zeta}$ 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

$$dX_t \quad = \quad \mathrm{sgn}(X_t)\, dB_t \tag{4.5}$$

where (B_t) is a Brownian motion and $\mathrm{sgn}(x) := \begin{cases} +1 & \text{for } x \geq 0, \\ -1 & \text{for } x < 0 \end{cases}$. Note the unusual convention $\mathrm{sgn}(0) = 1$ that is used below. We prove the following statements:

1) X is a weak solution of (4.5) on $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ if and only if X is an $(\mathcal{F}_t)$ Brownian motion.

2) If X is a weak solution w.r.t. a setup $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$ then for any $t \geq 0$, the process $(B_s)_{s \leq t}$ is measurable w.r.t. the σ-algebra generated by $\mathcal{F}_0$ and $\mathcal{F}^{|X|,P}$.

3) There is no strong solution to (4.5).

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

$$B_t \quad := \quad \int_0^t \mathrm{sgn}(X_s)\, dX_s$$

is a Brownian motion as well, and, by the "associative law",

$$\int_0^t \mathrm{sgn}(X_s)\, dB_s \quad = \quad \int_0^t \mathrm{sgn}(X_s)^2\, dX_s \quad = \quad X_t - X_0,$$

i.e., X is a weak solution to (4.5).

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

$$\begin{aligned} B_t - B_0 &= \int_0^t \operatorname{sgn}(X_s)\, dX_s \quad = \quad \lim_{\varepsilon \downarrow 0} \int_0^t \varphi_\varepsilon''(X_s)\, dX_s \\ &= \lim_{\varepsilon \downarrow 0} \big(\varphi_\varepsilon(X_t) - \varphi_\varepsilon(X_0) - \frac{1}{2}\int_0^t \varphi_\varepsilon''(X_s)\, ds\big) \\ &= \lim_{\varepsilon \downarrow 0} \big(\varphi_\varepsilon(|X_t|) - \varphi_\varepsilon(|X_0|) - \frac{1}{2}\int_0^t \varphi_\varepsilon''(|X_s|)\, ds\big) \end{aligned}$$

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).

Lévy characterization of Lévy processes

Lévy's characterization has a natural extension to discontinuous martingales.

Theorem 4.3. *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 $X_t^1, \ldots, X_t^d : \Omega \to \mathbb{R}$ 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 $k \in \{1, \ldots, d\}$,

(ii) $[X^k, X^l]_t^c = a^{kl}\, t$ for any $k, l \in \{1, \ldots, d\}$, 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 $0 \leq r \leq t$ and for any $B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$,

then $X_t = (X_t^1, \ldots, X_t^d)$ is a Lévy process with characteristic exponent

$$\psi(p) \quad = \quad \frac{1}{2} p\cdot ap - ip\cdot b + \int (1 - e^{ip\cdot y} + ip\cdot y)\, \nu(dy). \tag{4.6}$$

For proving the theorem, we assume without proof that a local martingale is a semimartingale 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. e.g. [32].

Proof of Theorem 4.3. We first remark that (iii) implies

$$E\Big[\sum_{s\in(r,t]} G_s\cdot f(\Delta X_s)\Big|\mathcal{F}_r\Big] \quad = \quad E\Big[\int_r^t\int G_s\cdot f(y)\,\nu(dy)\,ds\Big|\mathcal{F}_r\Big], \quad \text{a.s. for } 0\le r\le t \tag{4.7}$$

for any bounded left-continuous adapted process G, and for any measurable function $f:\mathbb{R}^d\setminus\{0\}\to\mathbb{C}$ satisfying $|f(y)|\le\mathrm{const.}\cdot(|y|\wedge|y|^2)$. 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 $c_i\in\mathbb{R}$, $B_i\in\mathcal{B}(\mathbb{R}^d\setminus\{0\})$, $0\le t_0<t_1<\dots<t_n$, 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

$$\begin{aligned} Z_t \quad &= \quad 1+\int_0^t Z_-\,ip\cdot dM+\int_0^t Z_-\,(\psi(p)+ip\cdot b-\frac{1}{2}\sum_{k,l}p_kp_la^{kl})\,dt \\ &\quad +\sum_{(0,t]} Z_-\,\Big(e^{ip\cdot\Delta X}-1-ip\cdot\Delta X\Big). \end{aligned} \tag{4.8}$$

By (4.7) and since $|e^{ip\cdot y}-1-ip\cdot y|\le\mathrm{const.}\cdot(|y|\wedge|y|^2)$, the series on the right hand side of (4.8) 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 (4.8) and (4.6), (Z_t) is a martingale for any $p\in\mathbb{R}^d$. The assertion now follows again by Lemma 4.1. □

An interesting consequence of Theorem 4.3 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.

Lévy characterization of weak solutions

Lévy's characterization of Brownian motion can be extended to solutions of stochastic differential equations of type

$$dX_t \quad = \quad b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t \tag{4.9}$$

driven by a d-dimensional Brownian motion (B_t). As a consequence, one can show that a process is a weak solution of (4.9) if and only if it solves the corresponding martingale problem. 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, i.e., bounded on $[0, t] \times K$ for any $t \geq 0$ and any compact set $K \subset \mathbb{R}^d$. Let

$$\mathcal{L} \quad = \quad \frac{1}{2} \sum_{i,j=1}^{d} a^{ij}(t, x) \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^{d} b^i(t, x) \frac{\partial}{\partial x^i} \tag{4.10}$$

denote the corresponding generator where $a(t, x) = \sigma(t, x)\sigma(t, x)^T$ is a symmetric $d \times d$ matrix for any t and x.

Theorem 4.4 (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 (4.9) on the setup $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$.*

(ii) *The processes $M^i_t := X^i_t - X^i_0 - \int_0^t b^i(s, X_s)\, ds$, $1 \leq i \leq d$, are continuous local $(\mathcal{F}^P_t)$ martingales with covariations*

$$[M^i, M^j]_t \quad = \quad \int_0^t a^{ij}(s, X_s)\, ds \qquad P\text{-a.s. for any } t \geq 0. \tag{4.11}$$

(iii) The processes $M_t^{[f]} := f(X_t) - f(X_0) - \int_0^t (\mathcal{L}f)(s,X_s)\,ds$, $f \in C^2(\mathbb{R}^d)$, *are continuous local* $(\mathcal{F}_t^P)$ *martingales.*

(iv) The processes $\hat{M}_t^{[f]} := f(t,X_t) - f(0,X_0) - \int_0^t \left(\frac{\partial f}{\partial t} + \mathcal{L}f\right)(s,X_s)\,ds$, $f \in C^2(\mathbb{R}_+ \times \mathbb{R}^d)$, *are continuous local* $(\mathcal{F}_t^P)$ *martingales.*

Proof. (i)⇒(iv) is a consequence of the Itô-Doeblin formula, cf. equation (2.57) 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

$$M_t^i \quad = \quad X_t^i - X_t^0 - \int_0^t b^i(s,X_s)\,ds \quad = \quad M_t^{[f]} \tag{4.12}$$

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

$$X_t^i X_t^j - X_0^i X_0^j \;=\; M_t^{[f]} + \int_0^t \left(a^{ij}(s,X_s) + X_s^i\, b^j(s,X_s) + X_s^j\, b^i(s,X_s)\right) ds. \tag{4.13}$$

On the other hand, by the product rule and (4.12),

$$\begin{aligned} X_t^i X_t^j - X_0^i X_0^j \quad &= \quad \int_0^t X_s^i\, dX_s^j + \int_0^t X_s^j\, dX_s^i + [X^i, X^j]_t \\ &= \quad N_t + \int_0^t \left(X_s^i\, b^j(s,X_s) + X_s^j\, b^i(s,X_s)\right) ds + [X^i, X^j]_t \end{aligned} \tag{4.14}$$

with a continuous local martingale N. Comparing (4.13) and (4.14) 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

$$\begin{aligned} [B^k, B^l]_t &= \int_0^t \sum_{i,j} \left(\sigma_{ki}^{-1}\, \sigma_{lj}^{-1}\right)(s, X_s)\, d[M^i, M^j] \\ &= \int_0^t \left(\sigma^{-1} a (\sigma^{-1})^T\right)_{kl} (s, X_s)\, ds \quad = \quad \delta_{kl}\, t \end{aligned}$$

for any $k, l = 1, \ldots, d$ by (4.11). □

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 [34] 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 [36] and by Ethier and Kurtz [15] for much more on the martingale problem and its applications to Markov processes.

4.2 Random time change

Random time change is already central to the work of Doeblin from 1940 that has been discovered only recently [10]. Independently, Dambis and Dubins-Schwarz have developed a theory of random time changes for semimartingales in the 1960s [23], [33]. 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.,

$$[M]_{T_A} \quad = \quad a \qquad \text{on } \{T_a < \infty\}, \quad \text{and,}$$

$$T_{[M]_t} \quad = \quad \inf\{u : [M]_u > [M]_t\} \quad = \quad \sup\{u : [M]_u = [M]_t\}$$

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 4.5 (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*

$$M_t \quad = \quad B_{[M]_t} \qquad \text{for any } t \geq 0, \quad \text{almost surely}. \tag{4.15}$$

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 4.5 ensures $T_a < \infty$ almost surely. If the assumption is violated then M can still be represented in the form (4.15) 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. [33].

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 4.5. 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 4.5 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 [13].

Doeblin [10] 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

$$dY_t \quad = \quad \sigma(Y_t)\, dB_t \tag{4.16}$$

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 4.5 and the remark below,

$$Y_t \quad = \quad X_{A_t} \qquad \text{with} \quad A_t = [Y]_t = \int_0^t \sigma(Y_r)^2 \, dr \tag{4.17}$$

and a Brownian motion X. Note that A depends on Y, so at first glace (4.17) seems not to be useful for solving the SDE (4.16). 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 (4.16) 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 4.6. *Suppose that (X_a) on $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ is a weak solution of an SDE of the form*

$$dX_a \quad = \quad \sigma(X_a) \, dB_a + b(X_a) \, da \tag{4.18}$$

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,

$$T_a \quad := \quad \int_0^a \varrho(X_u) \, du \quad < \quad \infty \qquad \forall a \in (0, \infty), \quad \textit{and} \quad T_\infty = \infty. \tag{4.19}$$

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

$$dY_t \quad = \quad \left(\frac{\sigma}{\sqrt{\varrho}} \right) (Y_t) \, dB_t + \left(\frac{b}{\varrho} \right) (Y_t) \, dt. \tag{4.20}$$

We only give a sketch of the proof of the theorem:

Proof of 4.6. (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 (4.20). □

In particular, the theorem shows that if X is a Brownian motion and condition (4.19) 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

$$dY_t \quad = \quad |Y_t|^\alpha\,dB_t, \qquad Y_0 \;= 0, \tag{4.21}$$

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 (4.19) holds, and therefore the process $Y_t = X_{A_t}$, $A = T^{-1}$, is a non-trivial weak solution of (4.21). 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 (4.19) 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

$$dX_t \quad = \quad \sigma(X_t)\, dB_t + b(X_t)\, dt, \qquad X_0 = x_0, \tag{4.22}$$

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 (4.22) 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 $\frac{1}{2}\sigma^2 s'' + bs' = 0$. 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

$$dY_t \quad = \quad \widetilde{\sigma}(Y_t)\, dB_t, \qquad Y_0 \quad = \quad s(x_0), \tag{4.23}$$

where $\widetilde{\sigma} := (\sigma s') \circ s^{-1}$. The SDE (4.23) 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 4.7. *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 (4.22) 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 (4.23) 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 [34]. Here we only note that 4.7 immediately implies existence and uniqueness of a maximal weak solution of (4.22):

Corollary 4.8. *Under the regularity and non-degeneracy conditions on σ and b imposed above there exists a weak solution X of (4.22) 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.,

$$R_t \quad = \quad \exp\big(\widetilde{B}_{T^{-1}(t)}\big) \qquad \text{with} \quad T_a \quad = \quad \int_0^a \exp\big(2\widetilde{B}_u\big)\,du$$

and a one-dimensional Brownian motion $\widetilde{B}$.

4.3 Girsanov transformation

In Section 3.3-3.6 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

$$dX_t \quad = \quad b(t, X_t)\,dt + dY_t \tag{4.24}$$

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 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

$$X = b(X) + Y, \qquad Y \sim N(0, I_d) \text{ w.r.t. } P, \tag{4.25}$$

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 $X^i, \ldots, X^{i-1}$ of X. The predictability ensures in particular that the transformation defined by (4.25) is invertible, with $Y^1 = X^1 - b^1$, $Y^2 = X^2 - b^2(X^1)$, $Y^3 = X^3 - b^3(X^1, X^2)$, ... ,$Y^n = X^n - b^n(X^1, \ldots, X^{n-1})$.

A random variable (X, P) is a "weak" solution of the equation (4.25) 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

$$\begin{aligned} f_X^P(x) &= f_Y^P\big(x - b(x)\big) \left| \det \frac{\partial(x - b(x))}{\partial x} \right| \\ &= (2\pi)^{-d/2} e^{-|x-b(x)|^2/2} \\ &= e^{x \cdot b(x) - |b(x)|^2/2} \, \varphi^d(x), \end{aligned}$$

where $\varphi^d(x)$ denotes the standard normal density in $\mathbb{R}^d$. Therefore we can conclude:

(X, P) is a weak solution of (4.25) 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

$$\frac{dP}{dQ} = \exp\big(X \cdot b(X) - |b(X)|^2/2\big). \tag{4.26}$$

In particular, we see that the distribution μ_b of a weak solution of (4.25) is uniquely determined, and μ_b satisfies

$$\mu_b = P \circ X^{-1} \ll Q \circ X^{-1} = N(0, I_d) = \mu_0$$

with relative density

$$\boxed{\frac{d\mu_b}{d\mu_0}(X) = e^{X \cdot b(X) - |b(x)|^2/2}.}$$

The example can be extended to Gaussian measures on Hilbert spaces and to more general transformations, leading to the Cameron-Martin Theorem (cf. Theorem 4.17 below) and Ramer's generalization [2]. Here, we study the more concrete situation where Y and X are replaced by a Brownian motion and a solution of the SDE (4.24) respectively. 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 $\mathcal{F}_t$, $t \le t_0$, 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.

Lemma 4.9. *1) For any $0 \le s \le t \le t_0$, and for any $\mathcal{F}_t$-measurable random variable $X : \Omega \to [0, \infty]$,*

$$E_P[X|\mathcal{F}_s] \;=\; \frac{E_Q[XZ_t|\mathcal{F}_s]}{E_Q[Z_t|\mathcal{F}_s]} \;=\; \frac{E_Q[XZ_t|\mathcal{F}_s]}{Z_s} \qquad P\text{-a.s. and } Q\text{-a.s.} \tag{4.27}$$

2) Suppose that $(M_t)_{t\le t_0}$ is an $(\mathcal{F}_t)$ adapted càdlàg 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 (4.27) 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).

(ii) By symmetry, it is enough to prove the implication "$\Leftarrow$". 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.,

$$E_P[M_{t\wedge T_n}\ ;\ A] \ =\ E_P[M_{s\wedge T_n}\ ;\ A] \quad \text{for any } A \in \mathcal{F}_s,\ \ 0 \le s \le t \le t_0. \tag{4.28}$$

To verify (4.28), we first note that

$$E_P[M_{t\wedge T_n}\ ;\ A \cap \{T_n \le s\}] \quad = \quad E_P[M_{s\wedge T_n}\ ;\ A \cap \{T_n \le s\}] \tag{4.29}$$

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,

$$\begin{aligned} E_P[M_{t\wedge T_n}\ ;\ A \cap \{T_n > s\}] \ &=\ E_Q[M_{t\wedge T_n}\, Z_{t\wedge T_n}\ ;\ A \cap \{T_n > s\}] \\ =\ E_Q[M_{s\wedge T_n}\, Z_{s\wedge T_n}\ ;\ A \cap \{T_n > s\}]] \ &=\ E_P[M_{s\wedge T_n}\ ;\ A \cap \{T_n > s\}] \end{aligned} \tag{4.30}$$

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 $Z_{t\wedge T_n}$. (4.28) follows from (4.29) and (4.30). □

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. Note that $(L_t)_{t\le t_0}$ 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):

$$Z_t \quad = \quad \mathcal{E}_t^L.$$

Translated Brownian motions

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

$$L_t \quad = \quad \int_0^t G_s \cdot dX_s, \quad t \geq 0,$$

with $G \in \mathcal{L}^2_{a,\mathrm{loc}}\big(\mathbb{R}_+, \mathbb{R}^d\big)$. Then $[L]_t = \int_0^t |G_s|^2 \, ds$, and hence

$$Z_t \quad = \quad \exp\Big(\int_0^t G_s \cdot dX_s - \frac{1}{2}\int_0^t |G_s|^2 \, ds \Big) \tag{4.31}$$

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). 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.

If the assumption holds then we can consider a probability measure P on $\mathcal{A}$ with

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_{t_0}} \quad = \quad Z_{t_0} \qquad Q\text{-a.s.} \tag{4.32}$$

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 4.10 (Maruyama 1954, Girsanov 1960). *Suppose that X is a d-dimensional Brownian motion w.r.t. Q and $(Z_t)_{t \leq t_0}$ is defined by (4.31) with $G \in \mathcal{L}^2_{a,loc}(\mathbb{R}_+, \mathbb{R}^d)$. If $E_Q[Z_{t_0}] = 1$ then the process*

$$B_t \quad := \quad X_t - \int_0^t G_s \, ds, \qquad t \leq t_0,$$

is a Brownian motion w.r.t. any probability measure P on $\mathcal{A}$ satisfying (4.32).

Proof. By Lévy's characterization, it suffices to show that $(B_t)_{t\leq t_0}$ is an $\mathbb{R}^d$-valued P-martingale with $[B^i, B^j]_t = \delta_{ij} t$ P-almost surely for any $i, j \in \{1, \dots, d\}$. Furthermore, by Lemma 4.9, and since P and Q are mutually absolutely continuous on $\mathcal{F}_{t_0}$, this holds true provided $(B_t Z_t)_{t\leq t_0}$ is a Q-martingale 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 $1 \leq i \leq d$,

$$\begin{aligned} d(B^i\, Z) &= B^i\, dZ + Z\, dB^i + d[B^i, Z] \\ &= B^i Z G \cdot dX + Z\, dX^i - Z\, G^i dt + Z G^i\, dt, \end{aligned} \tag{4.33}$$

where we have used that

$$d[B^i, Z] \quad = \quad ZG \cdot d[B^i, X] \quad = \quad ZG^i\, dt \qquad Q\text{- a.s.}$$

The right-hand side of (4.33) 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 (4.31) is a Q-martingale, then X satisfies

$$dX_t \quad = \quad G_t\, dt + dB_t.$$

with a P-Brownian motion B. It generalizes the Cameron-Martin Theorem to non-deterministic adapted translation

$$X_t(\omega) \quad \longrightarrow \quad X_t(\omega) - H_t(\omega), \qquad H_t \;=\; \int_0^t G_s\, ds,$$

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 4.10.

2) The assumption $E_a[Z_{t_0}] = 1$ ensuring that $(Z_t)_{t\leq t_0}$ is a Q-martingale is not always satisfied $-$ a sufficient condition is given in Theorem 4.12 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.

First applications to SDE

The Girsanov transformation can be used to construct weak solutions of stochastic differential equations. For example, consider an SDE

$$dX_t \quad = \quad b(t, X_t)\, dt + dB_t, \qquad X_0 \;=\; o, \qquad B \;\sim\; \mathrm{BM}(\mathbb{R}^d), \tag{4.34}$$

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. Let $X_t(x) = x_t$ denote the canonical process on

$$\Omega \quad := \quad \left\{ x \in C\big([0,\infty), \mathbb{R}^d\big) \;:\; x_0 \;=\; o \right\},$$

and let μ^0 denote Wiener measure on $(\Omega, \mathcal{F}^X_\infty)$. Then (X, μ^0) is a Brownian motion. By changing measure, we will transform it into a weak solution of (4.34).

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. μ^0.

We will see later that the assumption is always satisfied if b is bounded, or, more generally, growing at most linearly in x. If (A) holds then by the Kolmogorov extension theorem, there exists a probability measure μ^b on $\mathcal{F}^X_\infty$ such that μ^b is mutually absolutely continuous w.r.t. μ^0 on each of the σ-algebras $\mathcal{F}^X_t$, $t \in [0, \infty)$, with relative densities

$$\left.\frac{d\mu^b}{d\mu^0}\right|_{\mathcal{F}^X_t} \quad = \quad Z_t \qquad \mu^0\text{-a.s.}$$

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. μ^b. Thus we have shown:

Corollary 4.11. *Suppose that (A) holds. Then:*

1) *The process (X, μ^b) is a weak solution of (4.34).*

2) *For any $t \in [0,\infty)$, the distribution $\mu^b \circ X^{-1}$ is absolutely continuous w.r.t. Wiener measure μ^0 on $\mathcal{F}_t^X$ with relative density Z_t.*

The second assertion holds since $\mu^b \circ X^{-1} = \mu^b$. It yields a rigorous *path integral representation* for the solution (X, μ^b) of the SDE (4.34): If μ_t^b denotes the distribution of $(X_s)_{s \le t}$ on $C_0\big([0,t], \mathbb{R}^d\big)$ w.r.t. μ^b then

$$\mu_t^b(dx) \quad = \quad \exp\left(\int_0^t b(s,x_s)\cdot dx_s - \frac{1}{2}\int_0^t |b(s,x_s)|^2\, ds\right)\, \mu_t^0(dx). \tag{4.35}$$

By combining (4.35) with the **heuristic** path integral representation

$$\text{“}\ \mu_t^0(dx) \quad = \quad \frac{1}{\infty}\exp\left(-\frac{1}{2}\int_0^t |x_s'|^2\, ds\right)\, \delta_0(dx_0) \prod_{0<s\le t} dx_s\ \text{”}$$

Wiener measure, we obtain the non-rigorous but very intuitive representation

$$\text{“}\ \mu_t^b(dx) \quad = \quad \frac{1}{\infty}\exp\left(-\frac{1}{2}\int_0^t |x_s' - b(s,x_s)|^2\, ds\right)\, \delta_0(dx_0) \prod_{0<s\le t} dx_s\ \text{”} \tag{4.36}$$

of μ^b. Hence intuitively, the “likely” paths w.r.t. μ^b should be those for which the *action functional*

$$I(x) \quad = \quad \frac{1}{2}\int_0^t \left|x_s' - b(s,x_s)\right|^2\, ds$$

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 (4.34) equal to zero. Of course, these arguments do not hold rigorously, because $I(x) = \infty$ for μ_t^0- and μ_t^b- almost every x. Nevertheless, they provide an extremely valuable guideline to conclusions that can then be verified rigorously, e.g. via (4.35).

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 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

$$\begin{aligned} H_0 : \quad & X_t &=& \quad B_t, \\ H_1 : \quad & dX_t &=& \quad b(X_t)\,dt + dB_t, \end{aligned}$$

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 $d\mu_t^b/d\mu_t^0$ which by (4.35) and Itô's formula is given by

$$\begin{aligned} \frac{d\mu_t^b}{d\mu_t^0} &= \exp\left(\int_0^t b(X_s)\cdot dX_s - \frac{1}{2}\int_0 |b(X_s)|^2\,ds\right) \\ &= \exp\left(H(X_0) - H(X_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(X_s)\,ds\right) \end{aligned} \tag{4.37}$$

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

$$H(x_0) - H(x_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(x_s)\,ds \quad > \quad \log c. \tag{4.38}$$

Note that the integration by parts in (4.37) shows that the estimation procedure is quite stable, because the $\log$ likelihood ratio in (4.38) is continuous w.r.t. the supremum norm on $C_o([0,t],\mathbb{R}^d)$.

Novikov's condition

We finally derive a sufficient condition due to Novikov for ensuring that the exponential

$$Z_t \quad = \quad \exp\left(L_t - 1/2\,[L]_t\right)$$

of a continuous local $(\mathcal{F}_t)$ martingale is a martingale. Recall that Z is always a non-negative local martingale, and hence a supermartingale w.r.t. $(\mathcal{F}_t)$.

Theorem 4.12 (Novikov 1971). *Let $t_0 \in \mathbb{R}_+$. If $E[\exp\left(\frac{p}{2}[L]_{t_0}\right)] < \infty$ for some $p > 1$ then $(Z_t)_{t\le t_0}$ is an $(\mathcal{F}_t)$ martingale.*

Remark ($p = 1$). The Novikov criterion also applies if the condition above is satisfied for $p = 1$. Since the proof in this case is slightly more difficult, we only prove the restricted form of Novikov's criterion stated above.

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

$$
\begin{aligned}
& E[Z_{t\wedge T_n} \,;\, Z_{t\wedge T_n} \geq c] \\
= \quad & E\big[\exp\big(L_{t\wedge T_n} - \frac{p}{2}[L]_{t\wedge T_n}\big)\ \exp\big(\frac{p-1}{2}[L]_{t\wedge T_n}\big)\,;\, Z_{t\wedge T_n} \geq c\big] \qquad (4.39) \\
\leq \quad & E\big[\exp\big(pL_{t\wedge T_n} - \frac{p^2}{2}[L]_{t\wedge T_n}\big)\big]^{1/p} \cdot E\big[\exp\big(q\cdot\frac{p-1}{2}[L]_{t\wedge T_n}\big)\,;\, Z_{t\wedge T_n} \geq c\big]^{1/q} \\
\leq \quad & E\big[\exp\big(\frac{p}{2}[L]_t\big)\,;\, Z_{t\wedge T_n} \geq c\big]^{1/q}
\end{aligned}
$$

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 (4.39) 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 $\{Z_{t\wedge T_n} \mid n \in \mathbb{N}\}$ is uniformly integrable. $\square$

Example. 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 the quadratic variation 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 $t \in [0, \infty)$.

A more powerful application of Novikov's criterion is considered in the beginning of Section 3.4.

4.4 Drift transformations for Itô diffusions

We now consider an SDE

$$dX_t \quad = \quad b(X_t)\, dt + dB_t, \qquad X_0 \;=\; o, \quad B \;\sim\; \mathrm{BM}(\mathbb{R}^d) \qquad (4.40)$$

with initial condition $o \in \mathbb{R}^d$ and $b \in \mathcal{L}^\infty_{\text{loc}}(\mathbb{R}^d, \mathbb{R}^d)$. 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^0 denote Wiener measure on $(\Omega, \mathcal{F}^X_\infty)$ where $\Omega = C_0([0,\infty), \mathbb{R}^d)$ and $X_t(x) = x_t$ is the canonical process on Ω. Similarly as above, we assume:

Assumption (A). The exponential $Z_t = \exp\left(\int_0^t b(X_s)\cdot dX_s - \frac{1}{2}\int_0^t |b(X_s)|^2\, ds\right)$ is a martingale w.r.t. μ^0.

We note that by Novikov's criterion, the assumption always holds if

$$|b(x)| \quad \leq \quad c\cdot(1+|x|) \qquad \text{for some finite constant} \quad c > 0: \tag{4.41}$$

Exercise (Martingale property for exponentials).

a) Prove that a non-negative supermartingale (Z_t) satisfying $E[Z_t] = 1$ for any $t \geq 0$ is a martingale.

b) Now consider

$$Z_t \;=\; \exp\left(\int_0^t b(X_s)\cdot dX_s - \frac{1}{2}\int_0^t |b(X_s)|^2\, ds\right),$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a continuous vector field, and (X_t) is a Brownian motion w.r.t. the probability measure P.

(i) Show that (Z_t) is a supermartingale.

(ii) Prove that (Z_t) is a martingale if (4.41) holds.

Hint: Prove first that $E[\exp\int_0^\varepsilon |b(X_s)|^2\, ds] < \infty$ *for* $\varepsilon > 0$ *sufficiently small, and conclude that* $E[Z_\varepsilon] = 1$. *Then show by induction that* $E[Z_{k\varepsilon}] = 1$ *for any* $k \in \mathbb{N}$.

The Markov property

Recall that if (A) holds then there exists a (unique) probability measure μ^b on $(\Omega, \mathcal{F}^X_\infty)$ such that

$$\mu^b[A] \quad = \quad E^0[Z_t\,;\, A] \qquad \text{for any } \ t \geq 0 \ \text{ and } \ A \in \mathcal{F}^X_t.$$

Here E^0 denotes expectation w.r.t. μ^0. By Girsanov's Theorem, the process (X, μ^b) is a weak solution of (4.40). Moreover, we can easily verify that (X, μ^b) is a Markov process:

Theorem 4.13 (Markov property). *If (A) holds then (X, μ^b) is a time-homogeneous Markov process with transition function*

$$p_t^b(o, C) \quad = \quad \mu^b[X_t \in C] \quad = \quad E^0[Z_t \,;\, X_t \in C] \qquad \forall\, C \in \mathcal{B}(\mathbb{R}^d).$$

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,

$$\begin{aligned} E^b[f(X_t)|\mathcal{F}_s^X] &= E^0[f(X_t)Z_t|\mathcal{F}_s^X]/Z_s \\ &= E^0\Big[f(X_t)\exp\Big(\int_s^t b(X_r)\cdot dX_r - \frac{1}{2}\int_s^t |b(X_r)|^2\, dr\Big)\,\Big|\mathcal{F}_s^X\Big] \\ &= E^0_{X_s}[f(X_{t-s})Z_{t-s}] \quad = \quad (p^b_{t-s}f)(X_s) \end{aligned}$$

μ^0- and μ^b-almost surely where E^0_x denotes the expectation w.r.t. Wiener measure with start at x. □

Remark. 1) If b is time-dependent then one verifies in the same way that (X^b, μ) 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 $\Omega = C_0([0, 1], \mathbb{R}^d)$. Note that $\mathcal{F}_1^X$ is the Borel σ-algebra on the Banach space Ω. Our goal is to condition the diffusion process (X^b_μ) on a given terminal value $X_1 = y$, $y \in \mathbb{R}^d$. More precisely, we will construct a **regular version $\boldsymbol{y \mapsto \mu_y^b}$ of the conditional distribution $\boldsymbol{\mu^b[\cdot|X_1 = y]}$** in the following sense:

(i) *For any $y \in \mathbb{R}^d$, μ_y^b is a probability measure on $\mathcal{B}(\Omega)$ such that $\mu_y^b[X_1 = y] = 1$.*

(ii) Disintegration: *For any $A \in \mathcal{B}(\Omega)$, the function $y \mapsto \mu_y^b[A]$ is measurable, and*

$$\mu^b[A] \quad = \quad \int_{\mathbb{R}^d} \mu_y^b[A] p_1^b(o, dy).$$

(iii) The map $y \mapsto \mu_y^b$ is continuous w.r.t. weak convergence of probability measures.

Example (Brownian bridge). For $b = 0$, a regular version $y \mapsto \mu_y^0$ of the conditional distribution $\mu^0[\,\cdot\,|X_1 = y]$ w.r.t. Wiener measure can be obtained by linearly transforming the paths of Brownian motion, cf. Theorem 8.11 in [13]: Under μ^0, the process $X_t^y := X_t - tX_1 + ty$, $0 \le t \le 1$, is independent of X_1 with terminal value y, and the law μ_y^0 of $(X_t^y)_{t\in[0,1]}$ w.r.t. μ^0 is a regular version of $\mu^o[\,\cdot\,|X_1 = y]$. The measure μ_y^0 is called **"pinned Wiener measure"**.

The construction of a bridge process described in the example only applies for Brownian motion, which is a Gaussian process. 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 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

$$Z_t \quad = \quad \exp\left(H(X_0) - H(X_t) + \frac{1}{2}\int_0^t (\Delta H - |\nabla H|^2)(X_s)\, ds \right),$$

cf. (4.37). Note that the expression on the right-hand side is defined μ_y^0-almost surely for any y. Therefore, (Z_t) can be used for changing the measure w.r.t. the Brownian bridge.

Theorem 4.14 (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

$$p_1^b(o, y) \quad = \quad p_1^0(o, y) \cdot E_y^0[Z_1].$$

2) A regular version of $\mu^b[\,\cdot\,|X_1 = y]$ *is given by*

$$\mu_y^b(dx) \quad = \quad \frac{p_1^0(o,y)}{p_1^b(o,y)}\,\frac{\exp H(o)}{\exp H(y)}\,\exp\left(\frac{1}{2}\int_0^1(\Delta H - |\nabla H|^2)(x_s)\,ds\right)\,\mu_y^0(dx).$$

The theorem yields the existence and a formula for the heat kernel $p_1(o,y)$, as well as a path integral representation for the bridge measure μ_y^b:

$$\mu_y^b(dx) \quad \propto \quad \exp\left(\frac{1}{2}\int_0^1(\Delta H - |\nabla H|^2)(x_s)\,ds\right)\,\mu_y^0(dx). \tag{4.42}$$

Proof of 4.14. Let $F : \Omega \to \mathbb{R}_+$ and $g : \mathbb{R} \to \mathbb{R}_+$ be measurable functions. By the disintegration of Wiener measure into pinned Wiener measures,

$$E^b[F\cdot g(X_1)] \quad = \quad E^0[Fg(X_1)Z_1] \quad = \quad \int E_y^0[FZ_1]\,g(y)\,p_1^0(o,y)\,dy.$$

Choosing $F \equiv 1$, we obtain

$$\int g(y)\,p_1^b(o,dy) \quad = \quad \int g(y)\,E_y^0[Z_1]\,p_1^0(x,y)\,dy$$

for any non-negative measurable function g, which implies 1).

Now, by choosing $g \equiv 1$, we obtain

$$E^b[F] \quad = \quad \int E_y^0[FZ_1]\,p_1^0(o,y)\,dy \quad = \quad \int \frac{E_y^0[FZ_1]}{E_y^0[Z_1]}\,p_1^b(o,dy) \tag{4.43}$$

$$\quad = \quad \int E_y^b[F]\,p_1^b(0,dy) \tag{4.44}$$

by 1). This proves 2), because $X_1 = y$ μ_y^b-almost surely, and $y \mapsto \mu_y^b$ 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. One can proceed similarly as above after making sense of this stochastic integral for $\mu_<^0$-almost every x.

Example (**Reversibility in the gradient case**). The representation (4.42) 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 $\mu_y^b \circ R^{-1}$ of the bridge measure from o to y coincides with the bridge measure from y to o. Indeed, this property holds for the Brownian bridge, and the relative density in (4.42) is invariant under time reversal.

Drift transformations for general SDE

We now consider a drift transformation for an SDE of the form

$$dX_t \quad = \quad \sigma(t, X_t)\, dW_t + b(t, X_t)\, dt \tag{4.45}$$

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

$$Z_t \quad = \quad \exp\left(\int_0^t \beta(s, X_s) \cdot dW_s - \frac{1}{2}\int_0^t |\beta(s, X_s)|^2\, ds\right)$$

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 4.15. *Suppose that (X, Q) is a weak solution of (4.45). If $(Z_t)_{t\le t_0}$ is a Q-martingale and $P \ll Q$ on $\mathcal{F}_{t_0}$ with relative density Z_{t_0} then $((X_t)_{t\le t_0}, P)$ is a weak solution of*

$$dX_t \quad = \quad \sigma(t, X_t)\, dB_t + (b + \sigma\beta)(t, X_t)\, dt, \qquad B \ \sim\ \mathrm{BM}(\mathbb{R}^d). \tag{4.46}$$

Proof. By (4.45), the equation (4.46) holds with

$$B_t \quad = \quad W_t - \int_0^t \beta(s, X_s)\, ds.$$

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

$$\mathcal{L} \quad = \quad \frac{1}{2}\sum_{i,j} a^{ij}\frac{\partial^2}{\partial x^i \partial x^j} + b\cdot\nabla, \qquad a \ =\ \sigma\sigma^T,$$

then (X, P) is a solution of the martingale problem for

$$\widetilde{\mathcal{L}} \quad = \quad \mathcal{L} + \beta\cdot\sigma^T\nabla.$$

This *"Girsanov transformation for martingale problems"* carries over to diffusion processes with more general state spaces than $\mathbb{R}^n$.

Doob's h-transform and diffusion bridges

In the case of Itô diffusions, the h-transform for Markov processes is a special case of the drift transform studied above. Suppose that $h \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$ is a strictly positive space-time harmonic function for the generator $\mathcal{L}$ of the Itô diffusion (X, Q) (defined by (4.41)) with $h(0, o) = 1$:

$$\frac{\partial h}{\partial t} + \mathcal{L}h \quad = \quad 0, \qquad h(0, o) \;=\; 1. \tag{4.47}$$

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). For this purpose we write Z_t in exponential form. By the Itô-Doeblin formula and (4.47),

$$dZ_t \quad = \quad (\sigma^T \nabla h)(t, X_t) \cdot dW_t.$$

Hence $Z_t = \mathcal{E}_t^L = \exp(L_t - \frac{1}{2}[L]_t)$ 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 is a martingale, and $P \ll Q$ with $\frac{dP}{dQ}\big|_{\mathcal{F}_t} = Z_t$ then (X, P) solves the SDE (4.45) with

$$\beta \quad = \quad \sigma^T \nabla \log h.$$

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.

A more interesting 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 (X, μ^b) on $C_0([0, 1], \mathbb{R}^d)$ satisfying

$$dX_t \quad = \quad dB_t + b(X_t)\, dt, \qquad X_0 \;=\; o,$$

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

$$\frac{\partial}{\partial t} p_t^b(\cdot, y) \quad = \quad \mathcal{L}^b p_t^b(\cdot, y) \qquad \text{for any } \ t > 0,$$

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 $t \leq 1 - \varepsilon$ for any $\varepsilon > 0$, $h(t, X_t)$ is a martingale for $t < 1$. Now suppose that $\mu_y^b \ll \mu^b$ on $\mathcal{F}_t$ with relative density $h(t, X_t)$ for any $t < 1$. Then the marginal distributions of the process $(X_t)_{t<1}$ under μ^b, μ_y^b respectively are

$$\begin{aligned}(X_{t_1}, \ldots, X_{t_k}) \ &\sim \ p_{t_1}^b(o, x_1) p_{t_2 - t_1}^b(x_1, x_2) \cdots p_{t_k - t_{k-1}}^b(x_{k-1}, x_k)\, dx^k \qquad \text{w.r.t. } \mu^b, \\ &\sim \ \frac{p_{t_1}^b(o, x_1) p_{t_2 - t_1}^b(x_1, x_2) \cdots p_{t_k - t_{k-1}}^b(x_{k-1}, x_k) p_{1-t_k}^b(x_k, y)}{p_1^b(o, y)}\, dx^k \qquad \text{w.r.t. } \mu_y^b.\end{aligned}$$

This shows that $y \to \mu_y^b$ coincides with the regular version of the conditional distribution of μ^b given X_1, i.e., μ_y^b is the bridge measure from o to y. Hence, by Corollary 4.15, we have shown:

Theorem 4.16 (SDE for diffusion bridges). *The diffusion bridge (X, μ_y^b) is a weak solution of the SDE*

$$dX_t \quad = \quad dB_t \ + \ b(X_t)\, dt \ + \ (\nabla \log p_{1-t}^b(\cdot, y))(X_t) \ dt, \qquad t \ < \ 1. \tag{4.48}$$

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 $\mathcal{F}_1$, the measures μ^b and μ_y^b are singular.

Remark (Generalized diffusion bridges). Theorem 4.16 carries over to bridges of diffusion processes with non-constant diffusion coefficients σ. In this case, the SDE (4.48) is replaced by

$$dX_t \quad = \quad \sigma(X_t)\, dB_t + b(X_t)\, dt + \left(\sigma\sigma^T \nabla \log p_{1-t}(\cdot, y)\right)(X_t)\, dt. \tag{4.49}$$

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.

4.5 Large deviations on path spaces

In this section, we apply Girsanov's Theorem to study random perturbations of a dynamical system of type

$$dX_t^\varepsilon \quad = \quad b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\,dB_t, \qquad X_0^\varepsilon \;=\; 0, \tag{4.50}$$

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. Before, we give a complete proof of the Cameron-Martin Theorem.

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 $W_t(\omega) = \omega(t)$.

Translations 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 $\mu_h := \mu \circ \tau_h^{-1}$.

Theorem 4.17 (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

$$\frac{d\mu_h}{d\mu} \quad = \quad \exp\Big(\int_0^t h'_s \cdot dW_s - \frac{1}{2}\int_0^t |h'_s|^2\,ds\Big). \tag{4.51}$$

Proof. "$\Rightarrow$" 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

$$\mu_h \quad = \quad \mu \circ (W+h)^{-1} \quad = \quad \nu \circ W^{-1} \quad = \quad \nu.$$

"$\Leftarrow$" 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 4.18 (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,

$$\begin{aligned}\mu\big[||W - h|| < \delta\big] \quad &= \quad \mu_{-h}\big[||W|| < \delta\big] \\ &= \quad E\Big[\exp\Big(-\int_0^1 h' \cdot dW - \frac{1}{2}\int_0^1 |h'|^2\, ds\Big)\ ;\ \max_{s \le 1} |W_s| < \delta\Big]\end{aligned}$$

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 (4.50) 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 4.19 (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,

$$\begin{aligned} \mu\big[\sqrt{\varepsilon} W \in B(h,\delta)\big] &= \mu\big[W \in B(h/\sqrt{\varepsilon}, \delta/\sqrt{\varepsilon})\big] \\ &= \mu_{-h/\sqrt{\varepsilon}}\big[B(0,\delta/\sqrt{\varepsilon})\big] \\ &= E\Big[\exp\Big(-\frac{1}{\sqrt{\varepsilon}}\int_0^1 h' \cdot dW - \frac{1}{2\varepsilon}\int_0^1 |h'|^2\, ds\Big) \ ; \ B\Big(0, \frac{\delta}{\sqrt{\varepsilon}}\Big)\Big] \\ &\geq \exp\Big(-\frac{1}{\varepsilon} I(h) - \frac{c}{\sqrt{\varepsilon}}\Big)\, \mu\Big[\Big\{\int_0^1 h' \cdot dW \leq c\Big\} \cap B(0, \frac{\delta}{\sqrt{\varepsilon}})\Big] \\ &\geq \frac{1}{2}\exp\left(-\frac{1}{\varepsilon} I(h) - \sqrt{\frac{8I(h)}{\varepsilon}}\right) \end{aligned}$$

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 4.20 (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

$$\begin{aligned} \mu[I(\sqrt{\varepsilon}W^{(n)}) \geq \lambda] \quad &= \quad \mu\Big[\sum |\eta_k|^2 \geq 2\lambda/\varepsilon\Big] \\ &\leq \quad \exp(-2\lambda c/\varepsilon)\, E\Big[\exp\Big(c \sum |\eta_k|^2\Big)\Big], \end{aligned}$$

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

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in A] \quad \leq \quad -\lambda. \tag{4.52}$$

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 (4.52) now follows from

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[I(\sqrt{\varepsilon}W^{(n)}) > \lambda] \quad \leq \quad -\lambda, \quad \text{and} \tag{4.53}$$

$$\limsup_{\varepsilon \downarrow 0}\ \varepsilon \log\ \mu[||W - W^{(n)}||_{\sup} > \delta/\sqrt{\varepsilon}] \quad \leq \quad -\lambda. \tag{4.54}$$

The bound (4.53) holds by 1) for any $n \in \mathbb{N}$. The proof of (4.54) 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 (4.54) holds if n is large enough. □

Remark (Large deviation principle for Wiener measure). Theorems 4.19 and 4.20 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. [8] or [9]. 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

$$dX_t^\varepsilon \quad = \quad b(X_t^\varepsilon)\,dt + \sqrt{\varepsilon}\,dB_t, \qquad X_0^\varepsilon \;=\; 0. \tag{4.55}$$

This SDE can be solved pathwise:

Lemma 4.21 (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*

$$x(t) \quad = \quad \int_0^t b(x(s))\,ds + \omega(t) \qquad \forall\, t \in [0,1]. \tag{4.56}$$

The function x is absolutely continuous if and only if ω is absolutely continuous, and in this case,

$$x'(t) \quad = \quad b(x(t)) + \omega'(t) \qquad \textit{for a.e. } t \in [0,1]. \tag{4.57}$$

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 (4.56) 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 (4.56) w.r.t. driving paths $\omega, \widetilde{\omega} \in C[0,1],\mathbb{R}^d)$. Then for $t \in [0,1]$,

$$\begin{aligned} |x(t) - \widetilde{x}(t)| \quad &= \quad \Big| \int_0^t (b(\omega(s)) - b(\widetilde{\omega}(s)))\,ds + \sqrt{\varepsilon}(\omega(t) - \widetilde{\omega}(t)) \Big| \\ &\le \quad L \int_0^t |\omega(s) - \widetilde{\omega}(s)|\,ds + \sqrt{\varepsilon}|(t)m - \widetilde{\omega}(t)|. \end{aligned}$$

where $L \in \mathbb{R}_+$ is a Lipschitz constant for b. Gronwall's Lemma now implies

$$|x(t) - \widetilde{x}(t)| \quad \le \quad \exp(tL)\,\sqrt{\varepsilon}\,||\omega - \widetilde{\omega}||_{\sup} \qquad \forall\, t \in [0,1],$$

and hence

$$||x - \widetilde{x}||_{\sup} \quad \le \quad \exp(L)\,\sqrt{\varepsilon}\,||\omega - \widetilde{\omega}||_{\sup}.$$

This shows that the control map $\mathcal{J}$ is even Lipschitz continuous. □

For $\varepsilon > 0$, the unique solution of the SDE (4.55) 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 4.22 (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 4.19. Similarly, if A is closed then $\mathcal{J}^{-1}(A)$ is closed, and hence the corresponding upper bound holds by Theorem 4.20. 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 4.22 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. [8].

4.6 Change of measure for jump processes

A change of the underlying probability measure by an exponential martingale can also be carried out for jump processes. In this section, we first consider absolutely continuous measure transformations for Poisson point processes. We then apply the results to Lévy processes, and finally we prove a general result for semimartingales.

Poisson point processes

Let $(N_t)_{t\geq 0}$ be a Poisson point process on an 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 $\mathcal{L}^2_{\mathrm{loc}}(P \otimes \lambda \otimes \nu)$. 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

$$Z_t \quad := \quad \mathcal{E}^L_t \qquad \text{where} \qquad L_t \quad := \quad (H_\bullet \widetilde{N})_t. \tag{4.58}$$

Lemma 4.23. *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

$$\begin{aligned}
\mathcal{E}^L &= e^{L-[L]^c/2} \prod (1+\Delta L) e^{-\Delta L} \\
&= \exp\Big(L + \sum (\log(1+\Delta L) - \Delta L)\Big) \\
&= \exp\Big(G_\bullet \widetilde{N} + \int (G-H)\, ds\, \nu(dy)\Big).
\end{aligned}$$

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. $\square$

The exponential $Z_t = \mathcal{E}_t^L$ is a strictly positive local martingale w.r.t. Q, and hence a supermartingale. As usual, we fix $t_0 \in \mathbb{R}_+$ and assume:

Assumption. $(Z_t)_{t \leq t_0}$ is a martingale w.r.t. Q, i.e. $E_Q[Z_{t_0}] = 1$.

Also tThen 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 4.24 (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*

$$Z_t \quad = \quad \mathcal{E}_t^{\widetilde{N}^h} \quad = \quad \exp\Big(\widetilde{N}_t^g + t \int (g-h)\, d\nu\Big)$$

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 4.1) it suffices to show that the process

$$M_t^f = \exp\left(iN_t^f + t\psi(f)\right), \qquad \psi(f) = \int \left(1 - e^{if}\right) d\mu,$$

is a local martingale w.r.t. P for any elementary functions $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$. Furthermore, by Lemma 4.9, 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 4.24 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 (4.58), the process

$$M_t^f = \exp\left(iN_t^f + \int (1 - e^{if(y)})(1 + H_t(y))\, dy\right)$$

is a local martingale for any elementary functions $f \in \mathcal{L}^1(\nu)$. This property can be used as a definition of a point process with predictable intensity $(1 + H_t(y))\, dt\, \nu(dy)$.

Application to 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 $\int (|y| \wedge |y|^2)\, \nu(dy) < \infty$, and let

$$\mu(dy) = (1 + h(y))\, \nu(dy).$$

Recall that if (N_t, Q) is a Poisson point process with intensity measure ν, then

$$X_t = \int y\, \widetilde{N}_t(dy), \qquad \widetilde{N}_t = N_t - t\nu,$$

is a Lévy martingale with Lévy measure μ w.r.t. Q.

Corollary 4.25. *Suppose that $h \in \mathcal{L}^2(\nu)$ satisfies $\inf h > -1$ and $\sup h < \infty$. Then the process*

$$\overline{X}_t \quad = \quad \int y\,\overline{N}_t\,dy + t\int y\,h(y)\,\nu(dy), \qquad \overline{N}_t \quad = \quad N_t - t\mu,$$

is a Lévy martingale with Lévy measure μ w.r.t. P provided $P \ll Q$ on $\mathcal{F}_t$ with relative density Z_t for any $t \le 0$.

Example. Suppose that (X, Q) is a compound Poisson process with finite jump intensity measure ν, and let

$$N_t^h \quad = \quad \sum_{s\le t} h(\Delta X_s).$$

with h as above. Then (X, P) is a compound Poisson process with jump intensity measure $d\mu = (1+h)\,d\nu$ provided

$$\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} \quad = \quad \mathcal{E}_t^{\widetilde{N}^h} \quad = \quad e^{-t\int h\,d\nu}\prod_{s\le t}(1+h(\Delta X_s)).$$

A general theorem

We finally state a general change of measure theorem for possibly discontinuous semimartingales:

Theorem 4.26 (P.A. Meyer). *Suppose that the probability measures P and Q are equivalent on $\mathcal{F}_t$ for any $t \ge 0$ with relative density $\left.\frac{dP}{dQ}\right|_{\mathcal{F}_t} = Z_t$. 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 4.9, 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

$$\left[\frac{1}{Z}, [Z, M]\right] \quad = \quad \sum \Delta \frac{1}{Z}\, \Delta[Z, M].$$

□

Remark. Note that the process $\int \frac{1}{Z}\, d[Z, M]$ is not predictable in general. For a predictable counterpart to Theorem 4.26 cf. e.g. [32].

Chapter 5

Variations of stochastic differential equations

This chapter contains a first introduction to basic concepts and results of Malliavin calculus. For a more thorough introduction to Malliavin calculus we refer to [31], [30], [37], [21], [29] and [7].

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 $||\omega|| = \sup\{|\omega(t)| : t \in [0,1]\}$. We consider an SDE of type

$$dX_t \quad = \quad b(X_t)\, dt + \sigma(X_t)\, dW_t, \qquad X_0 \;=\; x, \tag{5.1}$$

driven by the canonical Brownian motion $W_t(\omega) = \omega(t)$. In this chapter, we will be interested in variations of the SDE and its solutions respectively. We will study the relations between different types of variations of (5.1):

- Variations of the initial condition: $\quad x \to x(\varepsilon)$
- Variations of the coefficients: $\quad b(x) \to b(\varepsilon, x), \quad \sigma(x) \to \sigma(\varepsilon, x)$
- Variations of the driving paths: $\quad W_t \to W_t + \varepsilon H_t, \quad (H_t)$ adapted
- Variations of the underlying probability measure: $\quad \mu \to \mu^\varepsilon = Z^\varepsilon \cdot \mu$

We first prove differentiability of the solution w.r.t. variations of the initial condition and the coefficients, see Section 5.1. In Section 5.2, we introduce 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 5.3, Section 5.4 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 5.5 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. [19], [30].

5.1 Variations of parameters in SDE

We now consider a stochastic differential equation

$$dX_t^\varepsilon \quad = \quad b(\varepsilon, X_t^\varepsilon)\,dt + \sum_{k=1}^d \sigma_k(\varepsilon, X_t^\varepsilon)\,dW_t^k, \qquad X_0^\varepsilon \;=\; x(\varepsilon), \tag{5.2}$$

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 (5.2). 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 5.1. *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*

$$\begin{aligned} dY_t^\varepsilon &= \left(\frac{\partial b}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial b}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon\right) dt && (5.3) \\ &\quad + \sum_{k=1}^d \left(\frac{\partial \sigma_k}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial \sigma_k}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon\right) dW_t^k, \\ Y_0^\varepsilon &= x'(\varepsilon), && (5.4) \end{aligned}$$

that is obtained by formally differentiating (5.2) 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 (5.3) 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, (5.3) is a deterministic ODE with coefficients depending on X^ε.

Proof of 5.1. We prove the stronger statement that there is a constant $M_p \in (0, \infty)$ such that

$$\left|\left|X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h\right|\right|_p \quad \leq \quad M_p\, |h|^2 \tag{5.5}$$

holds for any $\varepsilon, h \in \mathbb{R}^m$ with $\varepsilon, \varepsilon + h \in U$, where Y^ε is the unique strong solution of (5.3). 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

$$E\big[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_t^{\star\,p}\big] \quad \le \quad C_p \cdot \left(|\mathrm{I}|^p + \int_0^t E\big[|\mathrm{II}|^p + \sum_{k=1}^d |\mathrm{III}_k|^p\big]\,ds \right). \tag{5.6}$$

Since x, b and σ_k are C^2 with bounded derivatives, there exist finite constants C_{I}, C_{II}, C_{III} such that

$$|\mathrm{I}| \quad \le \quad C_{\mathrm{I}}|h|^2, \tag{5.7}$$

$$|\mathrm{II}| \quad \le \quad C_{\mathrm{II}}|h|^2 + \big|\frac{\partial b}{\partial x}(\varepsilon, X^\varepsilon)(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)\big|, \tag{5.8}$$

$$|\mathrm{III}_k| \quad \le \quad C_{\mathrm{III}}|h|^2 + \big|\frac{\partial \sigma_k}{\partial x}(\varepsilon, X^\varepsilon)(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)\big|. \tag{5.9}$$

Hence there exist finite constants $\widetilde{C}_p$, $\hat{C}_p$ such that

$$E[|\mathrm{II}|^p + \sum_{k=1}^d |\mathrm{III}_k|^p] \quad \le \quad \widetilde{C}_p \left(|h|^{2p} + E\left[\big|X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h\big|^p\right]\right),$$

and thus, by (5.6) and (5.7),

$$E\big[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_t^{\star p}\big] \quad \le \quad \hat{C}_p|h|^{2p} + C_p\widetilde{C}_p \int_0^t E\big[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_s^{\star p}\big]\,ds$$

for any $t \le 1$. The assertion (5.5) 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

$$d\xi_t^x \quad = \quad b(\xi_t^x)\,dt + \sum_{k=1}^d \sigma_k(\xi_t^x)\,dW_t^k, \qquad \xi_0^x \;=\; x. \tag{5.10}$$

Assuming that b and σ_k ($k = 1, \dots, d$) are C^2 with bounded derivatives, Theorem 5.1 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

$$dY_t^x \quad = \quad b'(\xi_t^x)\, Y_t^x\, dt + \sum_{k=1}^d \sigma_k'(\xi_t^x)\, Y_t^x\, dW_t^k, \qquad Y_0^x \;=\; I_n. \tag{5.11}$$

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. [24] or [14].

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 5.2 (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 $t \ge 0$ and $x, y, v \in \mathbb{R}^n$,

$$|\partial_v \xi_t^x| \quad \le \quad e^{\kappa t}|v|, \qquad \textit{and} \qquad |\xi_t^x - \xi_t^y| \quad \le \quad e^{\kappa t}|x-y|.$$

The theorem shows in particular that exponential stability holds if $\kappa < 0$.

Proof. The derivative $Y^x_{v,t} = \partial_v \xi^x_t$ 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

$$|\partial_v \xi^x_t|^2 \quad = \quad |Y^x_{v,t}|^2 \quad \leq \quad e^{2\kappa t} |v|^2, \qquad \text{and thus}$$

$$|\xi^x_t - \xi^y_t| \quad = \quad \left| \int_0^1 \partial_{x-y} \xi_t^{(1-s)x+sy} \, ds \right| \quad \leq \quad e^{\kappa t} |x - y|.$$

□

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 5.3. *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 $Y^x_{v,t} = \partial_v \xi^x_t$ is in $L^2(\Omega, \mathcal{A}, P)$, and*

$$\frac{d}{dt} E[|Y^x_{v,t}|^2] \quad = \quad 2E[Y^x_{v,t} \cdot K(\xi^x_t) Y^x_{v,t}]$$

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

$$\begin{aligned} d|Y_v|^2 &= 2Y_v \cdot dY_v + \sum_k d[Y_v^{(k)}] \\ &\overset{(5.11)}{=} 2Y_v \cdot b'(\xi)Y_v\, dt + 2\sum_k Y_v \cdot \sigma_k'(\xi)\, dW^k + \sum_k |\sigma_k'(\xi)Y_v|^2\, dt. \end{aligned}$$

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

$$E\left[|Y_{v,t\wedge T_n}|^2\right] = |v|^2 + 2E\left[\int_0^{t\wedge T_n} Y_v \cdot K(\xi)Y_v\, ds\right].$$

The assertion follows as $n \to \infty$. □

Theorem 5.4 (Exponential stability II). *Suppose that the assumptions in Lemma 5.3 hold, and let*

$$\kappa := \sup_{x\in\mathbb{R}^n} \sup_{\substack{v\in\mathbb{R}^n \\ |v|=1}} v \cdot K(x)v. \tag{5.12}$$

Then for any $t \geq 0$ and $x, y, v \in \mathbb{R}^n$,

$$E[|\partial_v \xi_t^x|^2] \leq e^{2\kappa t}|v|^2, \quad \textit{and} \tag{5.13}$$

$$E[|\xi_t^x - \xi_t^y|^2]^{1/2} \leq e^{\kappa t}|x - y|. \tag{5.14}$$

Proof. Since $K(x) \leq \kappa I_n$ holds in the form sense for any x, Lemma 5.3 implies

$$\frac{d}{dt}E[|Y_{v,t}|^2] \leq 2\kappa E[|Y_{v,t}|^2].$$

(5.13) now follows immediately by Gronwell's lemma, and (5.14) follows from (5.13) 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 (5.1) with assumptions as in Lemma 5.3 and Theorem 5.4. 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 (5.12).

Corollary 5.5. *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 5.4. □

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 5.3.

Exercise (Lipschitz bound). Prove that for any Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$,

$$||p_t f||_{\text{Lip}} \quad \leq \quad e^{\kappa t} ||f||_{\text{Lip}} \qquad \forall\, t \geq 0,$$

where $||f||_{\text{Lip}} = \sup\{|f(x) - f(y)|/|x-y| : x, y \in \mathbb{R}^n \text{ s.t. } x \neq y\}$.

For continuously differentiable functions f, we even obtain an explicit formula for the gradient of $p_t f$:

Corollary 5.6 (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*

$$v \cdot \nabla_x\, p_t f \quad = \quad E\big[Y_{v,t}^x \cdot \nabla_{\xi_t^x} f\big] \qquad \forall\, x, v \in \mathbb{R}^n. \tag{5.15}$$

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 5.6. 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. □

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 5.4, we will see that by applying an integration by parts on the right hand side of (5.15), 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. (**??**).

5.2 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

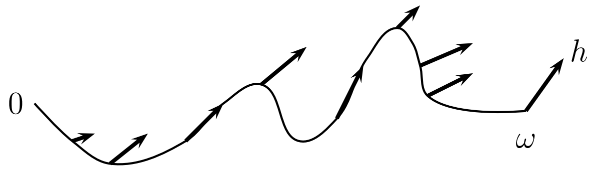
$$(h, g)_H \quad = \quad \int_0^1 h_t' \cdot g_t' \, dt, \qquad h, g \in H_{CM}.$$

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 $C^1_b(\Omega)$ 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.,

$$||l||_{\Omega'} \quad = \quad \sup \{ l(h) \ : \ h \in \Omega \ \text{ with } \ ||h|| \le 1 \} \, .$$

Definition (Malliavin Gradient I). *Let $F \in C^1_b(\mathbb{R})$ and $\omega \in \Omega$.*

1) *The **H-gradient** $(D^H F)(\omega)$ is the unique element in H_{CM} satisfying*

$$\big((D^H F)(\omega), h \big)_H \quad = \quad \frac{\partial F}{\partial h}(\omega) \quad = \quad (d_\omega F)(h) \qquad \text{for any } \ h \in H_{CM}. \tag{5.16}$$

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*

$$(D_t F)(\omega) \quad = \quad \frac{d}{dt}(D^H F)(\omega)(t) \qquad \text{for a.e. } \ t \in [0,1]. \tag{5.17}$$

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 $\omega \in \Omega$,

$$\begin{aligned} \frac{\partial F}{\partial h}(\omega) &= \big(h,(D^H F)(\omega)\big)_H = (h',(DF)(\omega))_{L^2} \\ &= \int_0^1 h'_t \cdot (D_t F)(\omega)\,dt, \end{aligned} \tag{5.18}$$

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 (5.16) holds.
2) By definition of the Malliavin gradient,

$$||D^H F(\omega)||_H^2 = \int_0^1 |D_t F(\omega)|^2\,dt.$$

3) Informally, one may think of $D_t F$ as a directional derivative of F in direction $I_{(t,1]}$, because

$$\text{“}\ D_t F = \frac{d}{dt} D^H F(t) = \int_0^1 (D^H F)'\,I'_{(t,1]} = \partial_{I_{(t,1]}} F\ \text{”}.$$

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

$$\frac{\partial}{\partial h} W_s^i = \frac{d}{d\varepsilon}\big(W_s^i + \varepsilon h_s^i\big)\Big|_{\varepsilon=0} = h_s^i = \int_0^1 h'_t \cdot e_i\, I_{(0,s)}(t)\,dt$$

for any $h \in H_{CM}$. Therefore, by the characterization in (5.18), the Malliavin gradient of F is given by

$$\big(D_t W_s^i\big)(\omega) \quad = \quad e_i \, I_{(0,s)}(t) \qquad \text{for every } \omega \in \Omega \text{ and a.e. } t \in (0,1).$$

Since the function $F : \Omega \to \mathbb{R}$ is linear, the gradient is deterministic. The H-gradient is obtained by integrating DW_s^i:

$$D_t^H W_s^i \quad = \quad \int_0^t D_r W_s^i \, dr \quad = \quad \int_0^t e_i \, I_{(0,s)} \quad = \quad (s \wedge t) \, e_i.$$

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

$$F \quad = \quad g_1 \cdot W_1 - \int_0^1 g_s' \cdot W_s \, ds \qquad \text{almost surely.} \tag{5.19}$$

The function on the right hand side of (5.19) 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 (5.18),

$$D_t F \quad = \quad g_t \qquad \text{and} \qquad D_t^H F \quad = \quad \int_0^t g_s \, ds.$$

Theorem 5.7 (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*

$$E\Big[\int_0^1 D_t F \cdot G_t \, dt\Big] \quad = \quad E\Big[F \int_0^1 G_t \cdot dW_t\Big]. \tag{5.20}$$

To recognize (5.20) as an integration by parts identity on Wiener space let $H_t = \int_0^t G_s, ds$. Then

$$\int_0^1 D_t F \cdot G_t \, dt \quad = \quad \left(D^H F, H\right)_H \quad = \quad \partial_H F.$$

Replacing F in (5.20) by $F \cdot \widetilde{F}$ with $F, \widetilde{F} \in C_b^1(\Omega)$, we obtain the equivalent identity

$$E[F \, \partial_H \widetilde{F}] \quad = \quad -E[\partial_H F \, \widetilde{F}] + E\Big[F \widetilde{F} \int_0^1 G_t \cdot dW_t\Big] \tag{5.21}$$

by the product rule for the directional derivative.

Proof of Theorem 5.7. The formula (5.21) is an infinitesimal version of Girsanov's Theorem. Indeed, suppose first that G is bounded. Then, by Novikov's criterion,

$$Z_t^\varepsilon \quad = \quad \exp\Big(\varepsilon \int_0^t G_s \cdot dW_s - \frac{\varepsilon^1}{2} \int_0^t |G_s|^2 \, ds\Big)$$

is a martingale for any $\varepsilon \in \mathbb{R}$. Hence for $H_t = \int_0^t G_s \, ds$,

$$E[F(W + \varepsilon H)] \quad = \quad E[F(W) Z_1^\varepsilon].$$

The equation (5.21) now follows formally by taking the derivative w.r.t. ε at $\varepsilon = 0$. Rigorously, we have

$$E\Big[\frac{F(W+\varepsilon H) - F(W)}{\varepsilon}\Big] \quad = \quad E\Big[F(W) \frac{Z_1^\varepsilon - 1}{\varepsilon}\Big]. \tag{5.22}$$

As $\varepsilon \to 0$, the right hand side in (5.22) 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 (5.22) converges to the left hand side in (5.21):

$$E\Big[\frac{1}{\varepsilon}(F(W+\varepsilon H) - F(W))\Big] \; = \; E\Big[\int_0^\varepsilon (\partial_H F)(W + sH) \, ds\Big] \; \longrightarrow \; E[(\partial_H F)(W)]$$

as $\varepsilon \to 0$ since $F \in C_b^1(\Omega)$. We have shown that (5.21) holds for bounded adapted G. Moreover, the identity extends to any $G \in \mathcal{L}_a^2(P \otimes \lambda)$ because both sides of (5.21) 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

$$\begin{aligned} D : C_b^1(\Omega) \subseteq L^2(\Omega, \mathcal{A}, P) &\longrightarrow L^2(\Omega \times [0,1] \to \mathbb{R}^d, \mathcal{A} \otimes \mathcal{B}, P \otimes \lambda), \\ F \quad &\longmapsto \qquad (D_t F)_{0 \le t \le 1}, \end{aligned}$$

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 (5.21), 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 5.8 (**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

$$D^H F \quad := \quad \lim_{n\to\infty} D^H F_n \tag{5.23}$$

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 (5.23), 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

$$(L - \widetilde{L}, h)_H \quad = \quad 0 \qquad \text{almost surely for any } \ h \in H. \tag{5.24}$$

Hence fix $h \in H$, and let $\varphi \in C_b^2(\Omega)$. Then by (5.21),

$$\begin{aligned}
E[(L - \widetilde{L}, h)_H \cdot \varphi] \quad &= \quad \lim_{n\to\infty} E[\partial_h(F_n - \widetilde{F}_n) \cdot \varphi] \\
&= \quad \lim_{n\to\infty} \Big\{ E\Big[(F_n - \widetilde{F}_n)\varphi \int_0^1 h' \cdot dW\Big] - E\Big[(F_n - \widetilde{F}_n)\partial_h \varphi\Big] \Big\} \\
&= \quad 0
\end{aligned}$$

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 (5.24) 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 5.9. *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*

$$\sup_{n\in\mathbb{N}} E[||D^H F_n||_H^2] \quad < \quad \infty \tag{5.25}$$

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

$$\frac{1}{k}\sum_{i=1}^{k} F_{n_i} \quad \to \quad F \qquad \textit{w.r.t. the (1,2) norm}. \tag{5.26}$$

The functional analytic proof is based on the theorems of Banach-Alaoglu and Banach-Saks, cf. e.g. the appendix in [27].

Proof. By (5.25), 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 5.9 can be used to extend the product and the chain rule to functions in $\mathbb{D}^{1,2}$.

Theorem 5.10. *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 $\varphi : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable with bounded derivatives then $\varphi(F^{(1)}, \dots, F^{(m)})$ is in $\mathbb{D}^{1,2}$, and

$$D\,\varphi(F^{(1)}, \dots, F^{(m)}) \quad = \quad \sum_{i=1}^{m} \frac{\partial \varphi}{\partial x_i}(F^{(1)}, \dots, F^{(m)})\,DF^{(i)}.$$

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,

$$\begin{aligned} D^H(F_nG_n) \quad &= \quad F_nD^HG_n + G_nD^HF_n \qquad \text{for any } n \in \mathbb{N}, \quad \text{and} \qquad (5.27)\\ ||D^H(F_nG_n)||_H \quad &\leq \quad |F_n|\,||D^HG_n||_H + |G_n|\,||D^HF_n||_H. \end{aligned}$$

Thus the sequence $(D^H(F_nG_n))_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega \to H; P)$. By Lemma 5.9, 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 (5.27). □

5.3 Digression on Representation Theorems

We now prove basic representation theorems for functions and martingales on Wiener space. The Bismut integration by parts identity can then be applied to obtain a more explicit form of the classical Itô Representation Theorem. Throughout this section, $W_t(\omega) = \omega_t$ denotes the canonical Brownian motion on Wiener space $(\Omega, \mathcal{A}, P)$, and

$$\mathcal{F}_t \quad = \quad \sigma(W_s : s \in [0,t])^P, \qquad t \geq 0,$$

is the completed filtration generated by (W_t).

Itôs Representation Theorem

Itô's Representation Theorem states that functions on Wiener space that are measurable w.r.t. the *Brownian filtration* $\mathcal{F}_t = \mathcal{F}_t^{W,P}$ can be represented as stochastic integrals:

Theorem 5.11 (Itô). *For any function $F \in \mathcal{L}^2(\Omega, \mathcal{F}_1, P)$ there exists a unique process $G \in L_a^2(0,1)$ such that*

$$F \quad = \quad E[F] + \int_0^1 G_s \cdot dW_s \qquad P\text{-almost surely.} \tag{5.28}$$

An immediate consequence of Theorem 5.11 is a corresponding representation for martingales *w.r.t. the Brownian filtration* $\mathcal{F}_t = \mathcal{F}_t^{W,P}$:

Corollary 5.12 (Itô representation for martingales). *For any 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*

$$M_t \quad = \quad M_0 + \int_0^t G_s \cdot dW_s \qquad P\text{-a.s.} \qquad \text{for any } t \in [0,1].$$

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 5.11, and then we prove the theorem:

Proof of Corollary 5.12. 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 5.11, there exists a unique process $G \in L^2_a(0,1)$ such that

$$M_1 \quad = \quad E[M_1] + \int_0^1 G \cdot dW \quad = \quad M_0 + \int_0^1 G \cdot dW \qquad \text{a.s.,}$$

and thus

$$M_t \quad = \quad E[M_1|\mathcal{F}_t] \quad = \quad M_0 + \int_0^t G \cdot dW \qquad \text{a.s. for any } t \geq 0.$$

□

Proof of Theorem 5.11. Uniqueness. Suppose that (5.28) holds for two processes $G, \widetilde{G} \in L^2_a(0,1)$. Then

$$\int_0^1 G \cdot dW \quad = \quad \int_0^1 \widetilde{G} \cdot dW,$$

and hence, by Itô's isometry,

$$||G - \widetilde{G}||_{L^2(P\otimes\lambda)} \quad = \quad \Big|\Big| \int (G - \widetilde{G}) \cdot dW \Big|\Big|_{L^2(P)} \quad = \quad 0.$$

Hence $G_t(\omega) = \widetilde{G}_t(\omega)$ for almost every (t, ω).

Existence. We prove the existence of a representation as in (5.28) in several steps $-$ starting with "simple" functions F.

1. Suppose that $F = \exp(ip \cdot (W_t - W_s))$ for some $p \in \mathbb{R}^d$ and $0 \leq s \leq t \leq 1$. By Itô's formula,

$$\exp(ip \cdot W_t + \frac{1}{2}|p|^2 t) \quad = \quad \exp(ip \cdot W_s + \frac{1}{2}|p|^2 s) + \int_s^t \exp\left(ip \cdot W_r + \frac{1}{2}|p|^2 r\right) ip \cdot dW_r.$$

Rearranging terms, we obtain an Itô representation for F with a bounded adapted integrand G.

2. Now suppose that $F = \prod_{k=1}^n F_k$ where $F_k = \exp\left(ip_k \cdot (W_{t_k} - W_{t_{k-1}})\right)$ for some $n \in \mathbb{N}$, $p_1, \ldots, p_n \in \mathbb{R}^d$, and $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n \leq 1$. Denoting by G_k the bounded adapted process in the Itô representation for F_k, we have

$$F \quad = \quad \prod_{k=1}^n \left(E[F_k] + \int_{t_k}^{t_{k+1}} G^k \cdot dW \right).$$

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. W. For this purpose, it suffices to verify that the product of two stochastic integrals $X_t = \int_0^t G \cdot dW$ and $Y_t = \int_0^t H \cdot dW$ with bounded adapted processes G and H is the stochastic integral of a process in $L_a^2(0,1)$ provided $\int_0^1 G_t \cdot H_t \, dt = 0$. This holds true, since by the product rule,

$$X_1 Y_1 \quad = \quad \int_0^1 X_t H_t \cdot dW_t + \int_0^1 Y_t G_t \cdot dW_t + \int_0^1 G_t \cdot H_t \, dt,$$

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

$$E\Big[\varphi \prod_{k=1}^n \exp(ip_k \cdot (W_{t_k} - W_{t_{k-1}}))\Big] \quad = \quad 0$$

for any $n \in \mathbb{N}$, $p_1, \ldots, p_n \in \mathbb{R}^d$ and $0 \le t_0 \le t_1 \le \cdots \le t_n \le 1$. By Fourier inversion, this implies

$$E[\varphi \mid \sigma(W_{t_k} - W_{t_{k-1}} : 1 \le k \le n)] \quad = \quad 0 \qquad \text{a.s.}$$

for any $n \in \mathbb{N}$ and $0 \le t_0 \le \cdots \le t_n \le 1$, and hence $\varphi = 0$ 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

$$F_n - E[F_n] \quad = \quad \int_0^1 G^{(n)} \cdot dW \tag{5.29}$$

with processes $G^{(n)} \in L_a^2(0,1)$. As $n \to \infty$,

$$F_n - E[F_n] \quad \longrightarrow \quad F - E[F] \qquad \text{in } L^2(P).$$

Hence, by (5.29) 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

$$F - E[F] \quad = \quad \int_0^1 G \cdot dW$$

by taking the L^2 limit on both sides of (5.29). □

Clark-Ocone formula

If F is in $\mathbb{D}^{1,2}$ then the process G in the Itô representation can be identified explicitly:

Theorem 5.13 (Clark-Ocone). *For any $F \in \mathbb{D}^{1,2}$,*

$$F - E[F] \quad = \quad \int_0^1 G \cdot dW$$

where

$$G_t \quad = \quad E[D_t F \mid \mathcal{F}_t].$$

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,

$$\begin{aligned} E\Big[\int_0^1 G_t \cdot H_t \, dt\Big] \quad &= \quad E\Big[\int_0^1 G \cdot dW \int_0^1 H \, dW\Big] \quad = \quad E\Big[\int_0^1 D_t F \cdot H_t \, dt\Big] \\ &= \quad E\Big[\int_0^1 E[D_t F|\mathcal{F}_t] \cdot H_t \, dt\Big] \end{aligned}$$

for all Setting $H_t := G_t - E[D_t F|\mathcal{F}_t]$ we obtain

$$G_t(\omega) \quad = \quad E[D_t F|\mathcal{F}_t](\omega) \qquad P \otimes \lambda \, - \, \text{a.e.}$$

□

5.4 First applications to stochastic differential equations

5.5 Existence and smoothness of densities

Index

Bibliography

[1]

[2]

[3] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge UP, Cambridge, 2004.

[4] Ole E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.*, 2(1):41–68, 1998.

[5] Richard F. Bass. *Stochastic Processes*. Cambridge University Press, Cambridge, 2011.

[6] Jean Bertoin. *Lévy Processes*. Cambridge UP, Cambridge, 1998.

[7] Jean-Michel Bismut. *Large deviations and the Malliavin calculus*. Birkhaeuser, Boston, 1994.

[8] Amir Dembo and Ofur Zeitouni. *Large Deviations Techniques and Applications*. Springer, 1998, Berlin.

[9] Frank den Hollander. *Large Deviations*. American Mathematical Society, Providence, 2008.

[10] Wolfgang Doeblin and Marc Yor. *pli cacheté*.

[11] Andreas Eberle. Markov processes. `http://www.uni-bonn.de/~eberle/MarkovProcesses/MPSkript.pdf`, 2009.

[12] Andreas Eberle. Wahrscheinlichkeitstheorie und Stochastische Prozesse. http://www.uni-bonn.de/~eberle/StoPr/WthSkriptneu.pdf, 2010.

[13] Andreas Eberle. Introduction to Stochastic Analysis. http://www.uni-bonn.de/~eberle/StoAn1011/StoAnSkriptneu.pdf, 2011.

[14] David Elworthy. *On the Geometry of Diffusion Operators and Stochastic Flows*. Springer, 1999, Berlin.

[15] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes*. Wiley-Interscience, Hoboken, NJ, 2005.

[16] William Feller. *Introduction to Probability Theory and Its Applications*. Wiley, New York, 1957.

[17] Peter K. Friz and Nicolas Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge UP, Cambridge, 2010.

[18] W. Hackenbroch and A. Thalmaier. *Stochastic Analysis*. Teubner, 1991.

[19] Martn Hairer. On malliavin's proof of hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[20] Elton Hsu. *Stochastic Analysis on Manifolds*. Oxford University Press, Oxford, 2002.

[21] Nobuyuki Ikeda and Takehiko Ed. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam, 1981.

[22] Jean Jacod and Albert Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2003.

[23] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Springer, New York, 2010.

[24] Hiroshi Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1997.

[25] Andreas Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin, 2006.

[26] Terry J. Lyons. *Differential Equations Driven by Rough Paths, Ecole d'Eté de Probabilités de Saint-Flour XXXIV*. Springer, 2004.

[27] Zhi-Ming Ma and Michael Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Springer, Berlin, 1992.

[28] Dilip B. Madan and Eugene Seneta. The variance gamma model for share market returns. *Journal of Business*, 63(4):511–524, 1990.

[29] Paul Malliavin. *Stochastic Analysis*. Springer, Berlin, 2002.

[30] David Nualart. *The Malliavin Calculus and Related Topics*. Springer, Berlin, 2006.

[31] Giuseppe Da Prato. *Introduction to Stochastic Analysis and Malliavin Calculus*. Edizioni della Normale, Pisa, 2009.

[32] Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, 2005.

[33] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer, Berlin, 2005.

[34] L. C. G. Rogers and David Williams. *Diffusions, Markov processes and martingales, Vol. 2: Ito calculus*. Cambridge UP, Camebridge, 2008.

[35] Shreve. *Stochastic calculus for Finance II*. Springer, New York, 2004.

[36] Daniel W. Stroock and S.R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, Berlin, 2005.

[37] Shinzo Watanabe. *Lectures on stochastic differential equations and Malliavin calculus*. Springer, Berlin, 1984.