

# Stochastic Analysis

Prof. Dr. Andreas Eberle

March 13, 2012

# Contents

<b>Contents</b>	<b>2</b>
<b>1 Lévy processes and Poisson point processes</b>	<b>6</b>
1.1 Lévy processes . . . . .	7
Characteristic exponents . . . . .	8
Martingales of Lévy processes . . . . .	9
1.2 Compound Poisson processes . . . . .	10
Poisson random measures . . . . .	11
Compound Poisson processes . . . . .	13
Superpositions and subdivisions of Poisson processes . . . . .	16
1.3 Poisson point processes and Lévy jump processes . . . . .	17
Poisson point processes . . . . .	18
Construction of compound Poisson processes from Poisson point processes . . . . .	20
Construction of Lévy processes with infinite jump intensity . . . . .	22
1.4 Examples and Lévy-Itô representation . . . . .	27
The Lévy-Itô decomposition . . . . .	28
Subordinators . . . . .	30
Finite variation Lévy processes . . . . .	33
Stable processes . . . . .	33
<b>2 Stochastic integrals and Itô calculus for semimartingales</b>	<b>38</b>
2.1 Finite variation calculus . . . . .	41

Lebesgue-Stieltjes integrals revisited . . . . .	42
Product rule . . . . .	44
Chain rule . . . . .	46
Exponentials of finite variation functions . . . . .	47
2.2 Stochastic integration for semimartingales . . . . .	55
Integrals with respect to bounded martingales . . . . .	55
Localization . . . . .	61
Integration w.r.t. semimartingales . . . . .	63
2.3 Quadratic variation and covariation . . . . .	65
Covariation and integration by parts . . . . .	66
Quadratic variation and covariation of local martingales . . . . .	67
Covariation of stochastic integrals . . . . .	73
The Itô isometry for stochastic integrals w.r.t. martingales . . . . .	75
2.4 Itô calculus for semimartingales . . . . .	76
Integration w.r.t. stochastic integrals . . . . .	76
Itô's formula . . . . .	77
Application to Lévy processes . . . . .	81
Applications to Itô diffusions . . . . .	83
Stochastic exponentials . . . . .	85
2.5 Stochastic integrals with general predictable integrands . . . . .	88
Definition of stochastic integrals w.r.t. semimartingales . . . . .	90
Localization . . . . .	92
Properties of the stochastic integral . . . . .	93
Dominated Convergence Theorem for stochastic integrals . . . . .	95
Integration w.r.t. Poisson point processes . . . . .	96
<b>3 Transformations and weak solutions of stochastic differential equations</b>	<b>99</b>
3.1 Lévy characterizations . . . . .	101
Lévy's characterization of Brownian motion . . . . .	102
Lévy characterization of Lévy processes . . . . .	106
Lévy characterization of weak solutions . . . . .	108
3.2 Random time change . . . . .	110

	Continuous local martingales as time-changed Brownian motions . . . .	111
	Time-change representations of stochastic integrals . . . . .	113
	Time substitution in stochastic differential equations . . . . .	113
	One-dimensional SDE . . . . .	116
3.3	Girsanov transformation . . . . .	119
	Change of measure on filtered probability spaces . . . . .	121
	Translated Brownian motions . . . . .	123
	First applications to SDE . . . . .	125
	Novikov's condition . . . . .	127
3.4	Drift transformations for Itô diffusions . . . . .	128
	The Markov property . . . . .	129
	Bridges and heat kernels . . . . .	130
	Drift transformations for general SDE . . . . .	133
	Doob's $h$ -transform and diffusion bridges . . . . .	134
3.5	Large deviations on path spaces . . . . .	136
	Translations of Wiener measure . . . . .	136
	Schilder's Theorem . . . . .	138
	Random perturbations of dynamical systems . . . . .	142
3.6	Change of measure for jump processes . . . . .	144
	Poisson point processes . . . . .	144
	Application to Lévy processes . . . . .	146
	A general theorem . . . . .	147
<b>4</b>	<b>Strong solutions and flows</b>	<b>149</b>
4.1	Existence and uniqueness of strong solutions . . . . .	151
	$L^p$ Stability . . . . .	152
	Existence of strong solutions . . . . .	154
	Non-explosion criteria . . . . .	156
4.2	Stratonovich differential equations . . . . .	157
	Itô-Stratonovich formula . . . . .	158
	Stratonovich SDE . . . . .	159
	Brownian motion on hypersurfaces . . . . .	161

---

Doss-Sussmann method . . . . .	163
Wong Zakai approximations of SDE . . . . .	164
4.3 Stochastic flows . . . . .	165
Kolmogorov-Centsov Theorem . . . . .	166
Existence of a continuous flow . . . . .	167
4.4 Local time . . . . .	169
<b>5 Variations of SDE</b>	<b>170</b>
5.1 Variations of parameters in SDE . . . . .	171
Differentiation of solutions w.r.t. a parameter . . . . .	172
Derivative flow and stability of SDE . . . . .	173
Consequences for the transition semigroup . . . . .	177
<b>Index</b>	<b>179</b>
<b>Bibliography</b>	<b>180</b>

# Chapter 1

## Lévy processes and Poisson point processes

In the next chapter we will extend stochastic calculus to processes with jumps. 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 [4].

Apart from simple transformations of Brownian motion, Lévy processes do not have continuous paths. Instead, we will assume that the paths are **càdlàg (continu 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  $X : I \rightarrow \mathbb{R}$  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

$$X_{s-} = \lim_{\varepsilon \downarrow 0} X_{s-\varepsilon}$$

the left limit of  $X$  at  $s$ , and by

$$\Delta X_s = X_s - X_{s-}$$

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  $X : I \rightarrow \mathbb{R}$ .

**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. 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  $X_t : \Omega \rightarrow \mathbb{R}^d$  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)  $X_{s+t} - X_s \sim X_t - X_0$  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. [27, Ch.I Thm.30].

**Remark (Infinitely divisible random variables).** The increments  $X_{s+t} - X_s$  of a Lévy process are **infinitely divisible** random variables, i.e., for any  $n \in \mathbb{N}$  there exist i.i.d. random variables  $Y_1, \dots, Y_n$  such that  $X_{s+t} - X_s$  has the same distribution as  $\sum_{i=1}^n Y_i$ . 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 are normal and Poisson distributions.

**Example (Brownian motion and Gaussian Lévy processes).** If  $(B_t)$  is a  $d$ -dimensional Brownian motion then  $X_t = \sigma B_t + bt$  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, cf. below. Moreover, a linear combination of independent Lévy processes is again a Lévy process:

**Example (Superposition).** 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  $\beta \in \mathbb{R}^{n \times d'}$ .

## 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  $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$  with  $\psi(0) = 0$  such that*

$$E[e^{ip \cdot X_t}] = e^{-t\psi(p)} \quad \text{for any } t \geq 0 \text{ and } p \in \mathbb{R}^d. \quad (1.1)$$

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

$$E[X_t] = it \nabla \psi(0) \quad , \quad \text{Cov}[X_t^k, X_t^l] = t \frac{\partial^2 \psi}{\partial p_k \partial p_l}(0) \quad (1.2)$$

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

*Proof.* Stationarity and independence of the increments implies the identity

$$\begin{aligned} \varphi_{t+s}(p) &= E[\exp(ip \cdot X_{t+s})] = E[\exp(ip \cdot X_t)] \cdot E[\exp(ip \cdot (X_{t+s} - X_s))] \\ &= \varphi_t(p) \cdot \varphi_s(p) \end{aligned} \quad (1.3)$$



for any  $p \in \mathbb{R}^d$  and  $s, t \geq 0$ . For a given  $p \in \mathbb{R}^d$ , right continuity of the paths and dominated convergence imply that  $t \mapsto \varphi_t(p)$  is right-continuous. Since

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

the function  $t \mapsto \varphi_t(p)$  is also left continuous, and hence continuous. By (1.3) 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 (1.1) holds. Arguing by contradiction we then see that  $\psi(0) = 0$  and  $\psi$  is continuous, since otherwise  $\varphi_t$  would not be continuous for all  $t$ .

Moreover, if  $X_t$  is (square) integrable then  $\varphi_t$  is  $\mathcal{C}^1$  (resp.  $\mathcal{C}^2$ ), and hence  $\psi$  is also  $\mathcal{C}^1$  (resp.  $\mathcal{C}^2$ ). The formulae in (1.2) for the first and second moment now follow by computing the derivatives w.r.t.  $p$  at  $p = 0$  in (1.1).  $\square$

The function  $\psi$  is called the **characteristic exponent** of the Lévy process.

**Examples.** 1) In the Brownian motion example above,

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

2) In the superposition example,

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

## 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 1.2.** *If  $(X_t)$  is a Lévy process with  $X_0 = 0$  and characteristic exponent  $\psi$ , 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  $a^{jk} = \frac{\partial^2 \psi}{\partial p_j \partial p_k}(0)$ , ( $j, k = 1, \dots, d$ ) provided  $X_t \in \mathcal{L}^2 \forall t \geq 0$ .

*Proof.* We only prove (ii) and (iii) for  $d = 1$  and leave the remaining assertions as an exercise to the reader. If  $d = 1$  and  $(X_t)$  is integrable then for  $0 \leq s \leq t$ ,

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

by independence and stationarity of the increments and by (1.2). Hence  $M_t = X_t - it\psi'(0)$  is a martingale. Furthermore,

$$M_t^2 - M_s^2 = (M_t + M_s)(M_t - M_s) = 2M_s(M_t - M_s) + (M_t - M_s)^2.$$

If  $(X_t)$  is square integrable then the same holds for  $(M_t)$ , and we obtain

$$\begin{aligned} E[M_t^2 - M_s^2 \mid \mathcal{F}_s] &= E[(M_t - M_s)^2 \mid \mathcal{F}_s] = \text{Var}[M_t - M_s \mid \mathcal{F}_s] \\ &= \text{Var}[X_t - X_s \mid \mathcal{F}_s] = \text{Var}[X_t - X_s] = \text{Var}[X_{t-s}] = (t-s)\psi''(0) \end{aligned}$$

Hence  $M_t^2 - t\psi''(0)$  is a martingale.  $\square$

Note that Corollary 1.2 (ii) shows that an integrable Lévy process is a *semimartingale* with martingale part  $M_t$  and continuous finite variation part  $bt$ . The identity (1.1) can be used to classify all Lévy processes, c.f. e.g. [3]. In particular, we will prove in Section 3.1 below that by Corollary 1.2, 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. We first look at simple examples and then describe a general construction.

## 1.2 Compound Poisson processes

Compound Poisson processes are pure jump Lévy processes, with paths that are constant apart from a finite number of jumps in finite time. We first introduce the standard Poisson process and Poisson random measures. Then we prove that compound Poisson processes are Lévy processes.

The most basic example of a jump process in continuous time is the Poisson process. It takes values in  $\{0, 1, 2, \dots\}$  and jumps up one unit each time after an exponentially

distributed waiting time. Explicitly, a Poisson process  $(N_t)_{t \geq 0}$  with intensity  $\lambda > 0$  is given by

$$N_t = \sum_{n=1}^{\infty} I_{\{S_n \leq t\}} = \#\{n \in \mathbb{N} : S_n \leq t\} \quad (1.4)$$

where  $S_n = T_1 + T_2 + \dots + T_n$  with independent random variables  $T_i \sim \text{Exp}(\lambda)$ . The increments  $N_t - N_s$  of a Poisson process over disjoint time intervals are independent and Poisson distributed with parameter  $\lambda(t - s)$ , cf. [8, Satz 10.12]. Note that by (1.4), the sample paths  $t \mapsto N_t(\omega)$  are càdlàg. In general, any Lévy process with

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

is called a **Poisson process with intensity  $\lambda$** , and can be represented as above. The characteristic exponent is  $\psi(p) = \lambda(1 - e^{ip})$ .

The paths of a Poisson process are increasing and hence of finite variation. The **compensated Poisson process**

$$M_t := N_t - E[N_t] = N_t - \lambda t$$

is an  $(\mathcal{F}_t^N)$  martingale, yielding the semimartingale decomposition

$$N_t = M_t + \lambda t$$

with the continuous finite variation part  $\lambda t$ . 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.

## Poisson random measures

Let

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

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  $N : \Omega \rightarrow \mathcal{M}_c^+([0, \infty))$ .

**Definition.** Let  $\nu$  be a  $\sigma$ -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  $\nu$**  if and only if

- (i)  $B \mapsto N(B)(\omega)$  is a counting measure for any  $\omega \in \Omega$ ,
- (ii) if  $B_1, \dots, B_n \in \mathcal{S}$  are disjoint then the random variables  $N(B_1), \dots, N(B_n)$  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  $\nu$  can be constructed as the empirical measure of a Poisson distributed number of independent samples from the normalized measure  $\nu/\nu(S)$ :

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

**Remark.** If the intensity measure  $\nu$  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*.

## Compound Poisson processes

A **compound Poisson process** is a continuous time Random Walk defined by

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

with a Poisson process  $(N_t)$  of intensity  $\lambda > 0$  and with independent identically distributed random variables  $\eta_j : \Omega \rightarrow \mathbb{R}^d$  ( $j \in \mathbb{N}$ ) 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

$$\nu(dy) = \lambda \pi(dy),$$

where  $\pi$  denotes the joint distribution of the random variables  $\eta_j$ .

**Lemma 1.3.** *A compound Poisson process is a Lévy process with characteristic exponent*

$$\psi(p) = \int (1 - e^{ip \cdot y}) \nu(dy). \quad (1.5)$$

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

$$X_{t_k} - X_{t_{k-1}} = \sum_{j=N_{t_{k-1}}+1}^{N_{t_k}} \eta_j \quad , \quad k = 1, 2, \dots, n, \quad (1.6)$$

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

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

where  $\varphi$  denotes the characteristic function of the jump sizes  $\eta_j$ . By taking the expectation value on both sides, we see that the increments in (1.6) are independent and

stationary, since the same holds for the Poisson process  $(N_t)$ . Moreover, by a similar computation,

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

for any  $p \in \mathbb{R}^d$ , which proves (1.5).  $\square$

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 1.10 below.

**Exercise (Martingales of Compound Poisson processes).** Show that the following processes are martingales:

$$(a) \quad M_t = X_t - bt \quad \text{where } b = \int y \nu(dy) \quad \text{provided } \eta_1 \in \mathcal{L}^1,$$

$$(b) \quad |M_t|^2 - at \quad \text{where } a = \int |y|^2 \nu(dy) \quad \text{provided } \eta_1 \in \mathcal{L}^2.$$

We have shown that a compound Poisson process with jump intensity measure  $\nu(dy)$  is a Lévy process with characteristic exponent

$$\psi_\nu(p) = \int (1 - e^{ip \cdot y}) \nu(dy) \quad , \quad p \in \mathbb{R}^d. \quad (1.7)$$

Since the distribution of a Lévy process on the space  $\mathcal{D}([0, \infty), \mathbb{R}^d)$  of càdlàg paths is uniquely determined by its characteristic exponent, we can prove conversely:

**Lemma 1.4.** *Suppose that  $\nu$  is a finite positive measure on  $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$  with total mass  $\lambda = \nu(\mathbb{R}^d \setminus \{0\})$ , and  $(X_t)$  is a Lévy process with  $X_0 = 0$  and characteristic exponent  $\psi_\nu$ , 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  $\lambda$  on  $(\Omega, \mathcal{A}, P)$  such that almost surely,*

$$X_t = \sum_{j=1}^{N_t} \eta_j \quad . \quad (1.8)$$

*Proof.* Let  $(\tilde{\eta}_j)$  be an arbitrary sequence of i.i.d. random variables with distribution  $\lambda^{-1}\nu$ , and let  $(\tilde{N}_t)$  be an independent Poisson process of intensity  $\nu(\mathbb{R}^d \setminus \{0\})$ , all defined on a probability space  $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ . Then the compound Poisson process  $\tilde{X}_t = \sum_{j=1}^{\tilde{N}_t} \tilde{\eta}_j$  is also a Lévy process with  $\tilde{X}_0 = 0$  and characteristic exponent  $\psi_\nu$ . Therefore, the finite dimensional marginals of  $(X_t)$  and  $(\tilde{X}_t)$ , and hence the distributions of  $(X_t)$  and  $(\tilde{X}_t)$  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

$$S_j = \inf \{s > S_{j-1} : \Delta X_s \neq 0\} \quad \text{for } j \in \mathbb{N}$$

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

$$N_t := |\{s \in (0, t] : \Delta X_s \neq 0\}| = |\{j \in \mathbb{N} : S_j \leq t\}|,$$

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 (1.8) holds. Moreover, for any  $j \in \mathbb{N}$  and  $t \geq 0$ ,  $\eta_j$  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  $\tilde{\eta}_j$  ( $j \in \mathbb{N}$ ) and  $\tilde{N}_t$  ( $t \geq 0$ ), which are the corresponding measurable functions of the process  $(\tilde{X}_t)$ . 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  $\lambda$ .  $\square$

The lemma motivates the following formal definition of a compound Poisson process:

**Definition.** Let  $\nu$  be a finite positive measure on  $\mathbb{R}^d$ , and let  $\psi_\nu : \mathbb{R}^d \rightarrow \mathbb{C}$  be the function defined by (1.7).

1) The unique probability measure  $\pi_\nu$  on  $\mathcal{B}(\mathbb{R}^d)$  with characteristic function

$$\int e^{ip \cdot y} \pi_\nu(dy) = \exp(-\psi_\nu(p)) \quad \forall p \in \mathbb{R}^d$$

is called the **compound Poisson distribution with intensity measure  $\nu$** .

- 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)  $\nu$** .

The compound Poisson distribution  $\pi_\nu$  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  $(\eta_j)$  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

$$\pi_\nu = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \nu^{*k},$$

where  $\nu^{*k}$  denotes the  $k$ -fold convolution of  $\nu$ .

## 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 1.5.** *Let  $K$  be a countable set.*

- 1) *Suppose that  $(N_t^{(k)})_{t \geq 0}$ ,  $k \in K$ , are independent Poisson processes with intensities  $\lambda_k$ . Then*

$$N_t = \sum_{k \in K} N_t^{(k)} \quad , \quad t \geq 0,$$

*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*

$$N_t^{(k)} = \sum_{j=1}^{N_t} I_{\{C_j=k\}} \quad , \quad t \geq 0,$$

*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  $\vec{N}_t : \Omega \rightarrow \mathbb{R}^K$  defined by

$$\vec{N}_t := \left( N_t^{(k)} \right)_{k \in K} = \sum_{j=1}^{N_t} \eta_j \quad \text{with} \quad \eta_j = \left( I_{\{k\}}(C_j) \right)_{k \in K}$$

is a compound Poisson process on  $\mathbb{R}^K$ , and hence a Lévy process. Moreover, by the proof of Lemma 1.3, 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

$$E \left[ \exp(ip \cdot \vec{N}_t) \right] = \prod_{k \in K} \exp(\lambda t q_k (e^{ip_k} - 1)) \quad \text{for any } p \in \mathbb{R}^K \text{ and } t \geq 0.$$

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  $q_k \lambda$ .  $\square$

### 1.3 Poisson point processes and Lévy jump 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. We would like to construct such processes from their jumps.

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

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

or, equivalently, in the random counting measure  $N : \Omega \rightarrow \mathcal{M}_c^+(\mathbb{R}_+ \times (\mathbb{R}^d \setminus \{0\}))$  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  $y$ . If  $(X_t)$  is a pure jump process then it can be recovered from  $N_t$  by adding up the jump sizes:

$$X_t - X_0 = \sum_{s \leq t} \Delta X_s = \int y N_t(dy).$$

We are now conversely going to construct a Lévy jump process from the measure-valued process encoding the jumps.

## Poisson point processes

Let  $(S, \mathcal{S}, \nu)$  be a  $\sigma$ -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  $\nu$**  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  $B_1, \dots, B_n \in \mathcal{S}$  are disjoint then  $(N_t(B_1))_{t \geq 0}, \dots, (N_t(B_n))_{t \geq 0}$  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.

$$N_t(B) = N((0, t] \times B) \quad \text{for any } t \geq 0 \text{ and } B \in \mathcal{S}.$$

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  $\nu$  then

$$(N_t(B_1), \dots, N_t(B_n))_{t \geq 0} \sim (\widetilde{N}_t(B_1), \dots, \widetilde{N}_t(B_n))_{t \geq 0}$$

for any finite collection of disjoint sets  $B_1, \dots, B_n \in \mathcal{S}$ , and, hence, for any finite collection of measurable arbitrary sets  $B_1, \dots, B_n \in \mathcal{S}$ .

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  $f : S \rightarrow T$  be a measurable function. Prove that if  $(N_t)$  is a Poisson point process with intensity measure  $\nu$  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 1.6 (Construction of Poisson point processes).**

1) Suppose that  $\nu$  is a finite measure with total mass  $\lambda = \nu(S)$ . Then

$$N_t = \sum_{j=1}^{K_t} \delta_{\eta_j}$$

is a Poisson point process of intensity  $\nu$  provided the random variables  $\eta_j$  are independent with distribution  $\lambda^{-1}\nu$ , and  $(K_t)$  is an independent Poisson process of intensity  $\lambda$ .

2) If  $(N_t^{(k)})$ ,  $k \in \mathbb{N}$ , are independent Poisson point processes on  $(S, \mathcal{S})$  with intensity measures  $\nu_k$  then

$$\overline{N}_t = \sum_{k=1}^{\infty} N_t^{(k)}$$

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  $\nu$  can be almost surely represented as in the first part of Theorem 1.6, 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 1.4.

### Construction of compound Poisson processes from Poisson point processes

We are now 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  $\nu$ . Then the support of  $N_t$  is almost surely finite for any  $t \geq 0$ . Therefore, we can define

$$X_t = \int_{\mathbb{R}^d \setminus \{0\}} y N_t(dy) = \sum_{y \in \text{supp}(N_t)} y N_t(\{y\}),$$

**Theorem 1.7.** *If  $\nu(\mathbb{R}^d \setminus \{0\}) < \infty$  then  $(X_t)_{t \geq 0}$  is a compound Poisson process with jump intensity  $\nu$ . More generally, for any Poisson point process with finite intensity measure  $\nu$  on a measurable space  $(S, \mathcal{S})$  and for any measurable function  $f : S \rightarrow \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , the process*

$$N_t^f := \int f(y) N_t(dy) \quad , \quad t \geq 0,$$

*is a compound Poisson process with intensity measure  $\nu \circ f^{-1}$ .*

*Proof.* By Theorem 1.6 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,

$$N_t^f = \int f(y) N_t(dy) = \sum_{j=1}^{K_t} f(\eta_j) \quad \text{almost surely.}$$

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 1.8 (Martingales of Poisson point processes).** *Suppose that  $(N_t)$  is a Poisson point process with finite intensity measure  $\nu$ . Then the following processes are martingales w.r.t. the filtration  $\mathcal{F}_t^N = \sigma(N_s(B) \mid 0 \leq s \leq t, B \in \mathcal{S})$ :*

$$(i) \quad \widetilde{N}_t^f = N_t^f - t \int f d\nu \quad \text{for any } f \in \mathcal{L}^1(\nu),$$

$$(ii) \quad \widetilde{N}_t^f \widetilde{N}_t^g - t \int fg d\nu \quad \text{for any } f, g \in \mathcal{L}^2(\nu),$$

$$(iii) \quad \exp(ipN_t^f + t \int (1 - e^{ipf}) d\nu) \quad \text{for any measurable } f : S \rightarrow \mathbb{R} \text{ and } p \in \mathbb{R}.$$

*Proof.* If  $f$  is in  $\mathcal{L}^p(\nu)$  for  $p = 1, 2$  respectively, then

$$\begin{aligned} \int |x|^p \nu \circ f^{-1}(dx) &= \int |f(y)|^p \nu(dy) < \infty, \\ \int x \nu \circ f^{-1}(dx) &= \int f d\nu, \quad \text{and} \quad \int xy \nu \circ (fg)^{-1}(dxdy) = \int fg d\nu. \end{aligned}$$

Therefore (i) and (ii) (and similarly also (iii)) follow from the corresponding statements for compound Poisson processes.  $\square$

With a different proof and an additional integrability assumption, the assertion of Corollary 1.8 extends to  $\sigma$ -finite intensity measures:

**Exercise (Expectation values and martingales for Poisson point processes with infinite intensity).** Let  $(N_t)$  be a Poisson point process with  $\sigma$ -finite intensity  $\nu$ .

a) By considering first elementary functions, prove that for  $t \geq 0$ , the identity

$$E \left[ \int f(y) N_t(dy) \right] = t \int f(y) \nu(dy)$$

holds for any measurable function  $f : S \rightarrow [0, \infty]$ . 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

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

c) Show that the processes considered in Corollary 1.8 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  $\nu$  then the signed measure valued stochastic process

$$\widetilde{N}_t(dy) := N_t(dy) - \nu(dy), \quad t \geq 0,$$

is called a **compensated Poisson point process**. Note that by Corollary 1.8 and the exercise,

$$\widetilde{N}_t^f = \int f(y) \widetilde{N}_t(dy)$$

is a martingale for any  $f \in \mathcal{L}^1(\nu)$ , i.e.,  $(\widetilde{N}_t)$  is a *measure-valued martingale*.

### Construction of Lévy processes with infinite jump intensity

Let  $\nu(dy)$  be a  $\sigma$ -finite positive measure on  $\mathbb{R}^d \setminus \{0\}$  such that

$$\nu(|y| > \varepsilon) < \infty \quad \text{for any } \varepsilon > 0. \quad (1.9)$$

Note that the condition (1.9) is necessary for the existence of a Lévy process with jump intensity  $\nu$ . Indeed, if (1.9) would be violated for some  $\varepsilon > 0$  then a corresponding Lévy process should have infinitely many jumps of size greater than  $\varepsilon$  in finite time. This contradicts the càdlàg property of the paths.

In order to construct a Lévy process with jump intensity  $\nu$ , let  $N_t(dy)$ ,  $t \geq 0$ , be a Poisson point process with intensity measure  $\nu$  defined on a probability space  $(\Omega, \mathcal{A}, P)$ . For a measure  $\mu$  and a measurable set  $A$ , we denote by

$$\mu^A(B) = \mu(B \cap A)$$

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  $\nu^A, \nu^B$  respectively.

For a given  $\varepsilon \in (0, 1)$  we consider the compound Poisson processes

$$X_t^\varepsilon := \int_{|y| > \varepsilon} y N_t(dy) \quad \text{and} \quad X_t^{\varepsilon,1} := \int_{\varepsilon < |y| \leq 1} y N_t(dy). \quad (1.10)$$

A first idea for defining a Lévy jump process  $(X_t)$  with jump intensity  $\nu$  would be to set

$$X_t := \lim_{\varepsilon \downarrow 0} X_t^\varepsilon = X_t^1 + \lim_{\varepsilon \downarrow 0} X_t^{\varepsilon,1}. \quad (1.11)$$

However, we will see below that the limit in (1.11) does not always exist. Therefore, in general, we will have to compensate the jumps of  $(X_t^{\varepsilon,1})$  before taking the limit, i.e., we will consider the compensated processes

$$M_t^{\varepsilon,1} := X_t^{\varepsilon,1} - t \int_{\varepsilon < |y| \leq 1} y \nu(dy).$$

Since

$$X_t^1 = \int y N_t^{(1,\infty)}(dy) \quad \text{and} \quad X_t^{\varepsilon,1} = \int y N_t^{(\varepsilon,1]}(dy),$$

$(X_t^1)$  and  $(X_t^{\varepsilon,1})$  are independent compound Poisson processes with (finite) jump intensity measures  $\nu^{(1,\infty)}$  and  $\nu^{(\varepsilon,1]}$  for any  $\varepsilon > 0$ . Moreover,

$$M_t^{\varepsilon,1} = \int y (N_t^{(\varepsilon,1]}(dy) - t \nu^{(\varepsilon,1]}(dy))$$

is a martingale w.r.t. the completed filtration

$$\mathcal{F}_t^{N,P} = \sigma(N_s(B) \mid s \in [0, t], B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}))^P$$

generated by the Poisson point process.

To study the existence of the limit as  $\varepsilon \downarrow 0$  we use martingale arguments. Let

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

denote the space of all square-integrable càdlàg martingales w.r.t. the filtration  $(\mathcal{F}_t^{N,P})$ . Recall that the  $L^2$  maximal inequality

$$E\left[\sup_{t \in [0,a]} |M_t|^2\right] \leq \left(\frac{2}{2-1}\right)^2 E[|M_a|^2]$$

holds for any right-continuous martingale in  $\mathcal{M}^2([0, a])$ . Since a uniform limit of càdlàg functions is again càdlàg, this implies that the space  $M_d^2([0, a])$  of equivalence classes of indistinguishable martingales in  $\mathcal{M}_d^2([0, a])$  is a *closed* subspace of the Hilbert space  $M^2([0, a])$  w.r.t. the norm

$$\|M\|_{M^2([0,a])} = E[|M_a|^2]^{1/2}.$$

**Lemma 1.9.** *Let  $a \in (0, \infty)$ .*

1) *If  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$  then  $(M^{1/n,1})$  is a Cauchy sequence w.r.t. the  $M^2([0, a])$  norm. In particular, there exists a martingale  $M \in \mathcal{M}_d^2([0, a])$  such that*

$$E\left[\sup_{0 \leq t \leq a} |M_t^{1/n,1} - M_t|^2\right] \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (1.12)$$

2) *The martingale  $(M_t)_{t \in [0,a]}$  is a Lévy process with characteristic exponent*

$$\psi(p) = \int_{|y| \leq 1} (1 - e^{ip \cdot y} + ip \cdot y) \nu(dy) \quad \text{for any } p \in \mathbb{R}^d. \quad (1.13)$$

*Proof.* 1) Let  $\delta, \varepsilon \in (0, 1)$  with  $\delta < \varepsilon$ . Then

$$X_t^{\delta,1} - X_t^{\varepsilon,1} = X_t^{\delta,\varepsilon} = \int_{\delta < |y| \leq \varepsilon} y N_t(dy) = \int y N_t^{(\delta,\varepsilon]}(dy),$$

i.e.,  $X_t^{\delta,1} - X_t^{\varepsilon,1}$  is a compound Poisson process with jump intensity measure  $\nu^{(\delta,\varepsilon]}$ .

Therefore,

$$M_t^{\delta,1} - M_t^{\varepsilon,1} = M_t^{\delta,\varepsilon} = X_t^{\delta,\varepsilon} - t \int y \nu^{(\delta,\varepsilon]}(dy)$$

is a càdlàg martingale, and

$$|M_t^{\delta,1} - M_t^{\varepsilon,1}|^2 - t \int |y|^2 \nu^{(\delta,\varepsilon]}(dy)$$

is a martingale as well. In particular,



$$\|M^{\delta,1} - M^{\varepsilon,1}\|_{M^2([0,a])}^2 = E[|M_a^{\delta,1} - M_a^{\varepsilon,1}|^2] = a \int_{\delta < |y| \leq \varepsilon} |y|^2 \nu(dy).$$

The assumption  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$  hence implies that  $(M^{1/n,1})_{n \in \mathbb{N}}$  is a Cauchy sequence in  $M_d^2([0, a])$ . The first assertion now follows by completeness of  $M_d^2([0, a])$ .

2) In order to identify the limit martingale  $(M_t)$  as a Lévy process with characteristic exponent (1.13), we note that  $(X_t^{\varepsilon,1})$  is a compound Poisson process with characteristic exponent

$$\psi_\varepsilon(p) = \int_{\varepsilon < |y| \leq 1} (1 - e^{ip \cdot y}) \nu(dy),$$

and, therefore,  $M_t^{\varepsilon,1} = X_t^{\varepsilon,1} - E[X_t^{\varepsilon,1}]$  is a compensated compound Poisson process with characteristic function

$$E[\exp(ip \cdot M_t^{\varepsilon,1})] = E[\exp(ip \cdot X_t^{\varepsilon,1})] \cdot \exp(-ip \cdot E[X_t^{\varepsilon,1}]) = \exp(-t\widetilde{\psi}_\varepsilon(p)),$$

where

$$\begin{aligned} \widetilde{\psi}_\varepsilon(p) &= \psi_\varepsilon(p) + ip \cdot \int_{\varepsilon < |y| \leq 1} y \nu(dy) \\ &= \int_{\varepsilon < |y| \leq 1} (1 - e^{ip \cdot y} + ip \cdot y) \nu(dy). \end{aligned}$$

As  $\varepsilon \downarrow 0$ ,  $\widetilde{\psi}_\varepsilon(p)$  converges to  $\psi(p)$  since  $1 - e^{ip \cdot y} + ip \cdot y = \mathcal{O}(|y|^2)$ . Hence the limit martingale  $M_t = \lim_{n \rightarrow \infty} M_t^{1/n,1}$  also has independent and stationary increments, and characteristic function

$$E[\exp(ip \cdot M_t)] = \lim_{n \rightarrow \infty} E[\exp(ip \cdot M_t^{1/n,1})] = \exp(-t\psi(p)).$$

□

We now return to our original goal of taking the limit of  $(X_t^{1/n})$  as  $n \rightarrow \infty$ . Note that for  $n \in \mathbb{N}$ ,

$$X_t^{1/n} = X_t^1 + M_t^{1/n,1} + t \int_{1/n < |y| \leq 1} y \nu(dy). \quad (1.14)$$

The first summand corresponds to big jumps, the second to the compensated small jumps, and the third summand is the compensating drift. For a càdlàg process  $(X_t)_{t \geq 0}$  and  $a \in (0, \infty)$  we define

$$\|X\|_a := E \left[ \sup_{t \leq a} |X_t|^2 \right]^{1/2}.$$

We can now state the fundamental result for the construction of Lévy processes:

**Theorem 1.10 (Existence of Lévy processes with infinite jump intensity).** *Let  $\nu$  be a  $\sigma$ -finite measure on  $\mathbb{R}^d \setminus \{0\}$  satisfying (1.9), and fix  $a \in (0, \infty)$ .*

- 1) *If  $\int_{|y| \leq 1} |y| \nu(dy) < \infty$ , or if  $\nu$  is symmetric (i.e.,  $\nu(B) = \nu(-B)$  for any  $B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ ) and  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$ , then there exists a Lévy process  $(X_t)_{t \in [0, a]}$  with characteristic exponent*

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

*such that  $\|X^{1/n} - X\|_a \rightarrow 0$  as  $n \rightarrow \infty$ .*

- 2) *More generally, if  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$ , then there exists a Lévy process  $(\widetilde{X}_t)$  with characteristic exponent*

$$\widetilde{\psi}(p) = \int (1 - e^{ip \cdot y} + ip \cdot y I_{\{|y| \leq 1\}}) \nu(dy) \quad \forall p \in \mathbb{R}^d$$

*such that  $\|X^1 + M^{1/n,1} - \widetilde{X}\| \rightarrow 0$  as  $n \rightarrow \infty$ .*

In the first case, if  $\int_{|y| \leq 1} |y| \nu(dy) < \infty$ , then  $1 - e^{ip \cdot y}$  is integrable w.r.t.  $\nu$ , and hence

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

Thus the limit process is a Lévy jump process with Lévy measure  $\nu$ .

In the second case, if  $\int_{1/n < |y| \leq 1} y \nu(dy)$  does not converge as  $n \rightarrow \infty$ , then both the process  $X_t^{1/n}$  as well as the compensated drift in (1.14) diverge. Nevertheless, the difference of both processes converges provided  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$ . The limit process

$$\widetilde{X}_t = \lim_{n \rightarrow \infty} \left( X_t^1 + X_t^{1/n,1} - E \left[ X_t^{1/n,1} \right] \right)$$

is then a partially compensated process with “infinite compensating drift”.

*Proof.* 1) First suppose that  $\int_{|y| \leq 1} |y| \nu(dy) < \infty$ , and let  $b := \int_{|y| \leq 1} y \nu(dy)$ . Then  $\int_{1/n < |y| \leq 1} y \nu(dy) \rightarrow b$  as  $n \rightarrow \infty$ , and hence by (1.14) and Lemma 1.9,

$$X^{1/n,1} \longrightarrow X^{0,1} \quad \text{w.r.t. } \|\cdot\|_a, \quad (1.15)$$

where  $X_t^{0,1} := M_t + tb$  and  $(M_t)$  is the limiting martingale in (1.12). Setting  $X_t := X_t^1 + X_t^{0,1}$ , we conclude that

$$\|X^{1/n} - X\|_a = \|(X^1 + X^{1/n,1}) - (X^1 + X^{0,1})\|_a \rightarrow 0 \quad (1.16)$$

Since  $X^1$  and  $X^{1/n,1}$  are independent Lévy processes for every  $n$ ,  $X^1$  and  $X^{0,1}$  are also independent Lévy processes. Therefore  $X$  is a Lévy process. Moreover, by (1.16),  $X$  has characteristic exponent  $\psi(p) = \lim_{n \rightarrow \infty} \psi_{1/n}(p)$ .

If  $\nu$  is symmetric then the compensating drift  $\int_{1/n < |y| \leq 1} y \nu(dy)$  vanishes for any  $n$ . Hence  $X^{1/n,1} = M^{1/n,1}$  for all  $n$ , and the convergence in (1.15) and (1.16) holds with  $b = 0$  under the weaker assumption  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$ .

2) In general, if  $\int_{|y| \leq 1} |y|^2 \nu(dy) < \infty$ , then we still have

$$\|(X^1 + M^{1/n,1}) - (X^1 + M)\|_a \rightarrow 0$$

by Lemma 1.9. The limit  $\tilde{X} = X^1 + M$  of the partially compensated processes is again the sum of two independent Lévy processes. Hence  $\tilde{X}$  is a Lévy process with characteristic exponent

$$\tilde{\psi}(p) = \psi_{X^1}(p) + \psi_M(p) = \int_{|y| > 1} (1 - e^{ip \cdot y}) \nu(dy) + \int_{|y| \leq 1} (1 - e^{ip \cdot y} + ip \cdot y) \nu(dy).$$

□

## 1.4 Examples and Lévy-Itô representation

In the last section, we have constructed Lévy processes corresponding to a given jump intensity measure  $\nu$  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.

## The Lévy-Itô decomposition

Already the classical Lévy-Khinchin formula for infinity divisible random variables (see Cor.1.12 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  $\sigma$ -finite measure  $\nu$  on  $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$  such that

$$\int (1 \wedge |y|^2) \nu(dy) < \infty . \quad (1.17)$$

Note that (1.17) holds if and only if  $\nu$  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  $(a, b, \nu)$ .

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  $\nu$ . We define a Lévy process  $(X_t)$  by setting

$$X_t = \sigma B_t + bt + \int_{|y| > 1} y N_t(dy) + \lim_{\varepsilon \downarrow 0} \int_{\varepsilon < |y| \leq 1} y (N_t(dy) - t\nu(dy)) . \quad (1.18)$$

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. All four summands are independent Lévy processes. The limit of the martingale

$$M_t^{\varepsilon_n, 1} = \int_{\varepsilon_n < |y| \leq 1} y (N_t(dy) - t\nu(dy)) \quad , \quad t \geq 0,$$

exists w.r.t. convergence in  $M_d^2([0, a])$  for any  $a \in (0, \infty)$  and any sequence  $\varepsilon_n \downarrow 0$ . If, moreover,  $\int (1 \wedge |y|) \nu(dy) < \infty$  then  $\int_{\varepsilon < |y| \leq 1} y N_t(dy)$  converges as  $\varepsilon \downarrow 0$  and the compensation can be omitted.

As a sum of independent Lévy processes, the process  $(X_t)$  is a Lévy process with characteristic exponent

$$\psi(p) = \frac{1}{2} p \cdot ap - ib \cdot p + \int_{\mathbb{R}^d \setminus \{0\}} (1 - e^{ip \cdot y} + ip \cdot y I_{\{|y| \leq 1\}}) \nu(dy) . \quad (1.19)$$

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

**Theorem 1.11 (Lévy-Itô decomposition).**

- 1) The expression (1.18) defines a Lévy process with characteristic exponent  $\psi$ .
- 2) Conversely, any Lévy process  $(X_t)$  can be decomposed as in (1.18) with the Poisson point process

$$N_t = \sum_{\substack{s \leq t \\ \Delta X_s \neq 0}} \Delta X_s \quad , \quad t \geq 0, \quad (1.20)$$

an independent Brownian motion  $(B_t)$ , a matrix  $\sigma \in \mathbb{R}^{d \times d}$ , a vector  $b \in \mathbb{R}^d$ , and a  $\sigma$ -finite measure  $\nu$  on  $\mathbb{R}^d \setminus \{0\}$  satisfying (1.17).

We will not prove the second part of the theorem here. The principal way to proceed is to define  $(N_t)$  via (1.17), and to consider the differences of  $(X_t)$  and the integrals w.r.t.  $(N_t)$  on the right hand side of (1.18). 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 3.1 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 1.12 (Lévy-Khinchin formula).** *For a function  $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$  the following statements are all equivalent:*

- (i)  $\psi$  is the characteristic exponent of a Lévy process.
- (ii)  $\exp(-\psi)$  is the characteristic function of an infinitely divisible random variable.
- (iii)  $\psi$  satisfies (1.19) with a non-negative definite symmetric matrix  $a \in \mathbb{R}^{d \times d}$ , a vector  $b \in \mathbb{R}^d$ , and a measure  $\nu$  on  $\mathcal{B}(\mathbb{R}^d \setminus \{0\})$  such that  $\int (1 \wedge |y|^2) \nu(dy) < \infty$ .

*Proof.* (iii) $\Rightarrow$ (i) holds by the first part of Theorem 1.10, 1).

(i) $\Rightarrow$ (ii): If  $(X_t)$  is a Lévy process with characteristic exponent  $\psi$  then  $X_1 - X_0$  is an infinitely divisible random variable with characteristic function  $\exp(-\psi)$ .

(ii) $\Rightarrow$ (iii) is the content of the classical Lévy-Khinchin theorem, see e.g. [12]. □

We are now going to consider several important subclasses of Lévy processes. The class of **Gaussian Lévy processes** of type

$$X_t = \sigma B_t + bt$$

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  $\nu$  is a positive measure on  $(0, \infty)$  with

$$\int_{(0, \infty)} (1 \wedge |y|) \nu(dy) < \infty.$$

Then a Poisson point process  $(N_t)$  with intensity measure  $\nu$  satisfies almost surely

$$\text{supp}(N_t) \subset [0, \infty) \quad \text{for any } t \geq 0.$$

Hence the integrals

$$X_t = \int y N_t(dy) \quad , \quad t \geq 0,$$

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.

Here are a few important examples of subordinators:

**Example (Gamma process).** The Gamma distributions form a convolution semigroup of probability measures on  $(0, \infty)$ , i.e.,

$$\Gamma(r, \lambda) * \Gamma(s, \lambda) = \Gamma(r + s, \lambda) \quad \text{for any } r, s, \lambda > 0.$$

Therefore, for any  $a, \lambda > 0$  there exists an increasing Lévy process  $(\Gamma_t)_{t \geq 0}$  with increment distributions

$$\Gamma_{t+s} - \Gamma_s \sim \Gamma(at, \lambda) \quad \text{for any } s, t \geq 0.$$

Computation of the Laplace transform yields

$$E[\exp(-u\Gamma_t)] = \left(1 + \frac{u}{\lambda}\right)^{-at} = \exp\left(-t \int_0^\infty (1 - e^{-uxy}) ay^{-1} e^{-\lambda y} dy\right) \quad (1.21)$$

for every  $u \geq 0$ , cf. e.g. [21, Lemma 1.7]. Since  $\Gamma_t \geq 0$ , both sides in (1.21) have a unique analytic extension to  $\{u \in \mathbb{C} : \Re(u) \geq 0\}$ . Therefore, we can replace  $u$  by  $-ip$  in (1.21) to conclude that the characteristic exponent of  $(\Gamma_t)$  is

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

Hence the Gamma process is a non-decreasing pure jump process with jump intensity measure  $\nu$ .

**Example (Inverse Gaussian subordinators).** Let  $(B_t)_{t \geq 0}$  be a one-dimensional Brownian motion with  $B_0 = 0$  and let

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

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  $(\mathcal{F}_{T_s}^B)_{s \geq 0}$ . For any  $\omega$ ,  $s \mapsto T_s(\omega)$  is the generalized right inverse of the Brownian path  $t \mapsto B_t(\omega)$ . Moreover, by the strong Markov property, the process

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

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

$$T_{s+u} = T_s + \tilde{T}_u^{(s)} \quad \text{for } s, u \geq 0, \quad (1.22)$$

where  $\tilde{T}_u^{(s)} = \inf \{t \geq 0 : \tilde{B}_t^{(s)} = u\}$  is the first passage time to  $u$  for the process  $\tilde{B}^{(s)}$ . By (1.22), the increment  $T_{s+k} - T_s$  is independent of  $\mathcal{F}_{T_s}^B$  with distribution

$$T_{s+u} - T_s \sim T_u \sim \frac{u}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{u^2}{2x}\right) I_{(0,\infty)}(x) dx.$$

Hence  $(T_s)$  is a subordinator. It is called “**The Lévy subordinator**”. Computation of the characteristic function shows that  $(T_s)$  is a pure jump Lévy process with Lévy measure

$$\nu(dx) = (2\pi)^{-1/2} x^{-3/2} I_{(0,\infty)}(x) dx.$$

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

$$T_s^X = \inf \{t \geq 0 : X_t = s\}, \quad s \geq 0,$$

is called an **Inverse Gaussian subordinator**. An inverse Gaussian subordinator is a Lévy jump process with jump intensity

$$\nu(dx) = (2\pi)^{-1/2} x^{-3/2} \exp(-b^2 x/2) I_{(0,\infty)}(x) dx$$

**Exercise (Time change by subordinators).** Suppose that  $(X_t)$  is a Lévy process with Laplace exponent  $\eta_X : \mathbb{R}_+ \rightarrow \mathbb{R}$ , i.e.,

$$\int E[\exp(-\alpha X_t)] = \exp(-t\eta_X(\alpha)) \quad \text{for any } \alpha \geq 0.$$

Prove that if  $(T_s)$  is an independent subordinator with Laplace exponent  $\eta_T$  then the time-changed process

$$\tilde{X}_s := X_{T_s}, \quad s \geq 0,$$

is again a Lévy process with Laplace exponent

$$\tilde{\eta}(p) = \eta_T(\eta_X(p)).$$

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

$$\nu(dx) = \lambda(2\pi)^{-1/2} \exp(-x^2/2) dx.$$



2) If  $(\Gamma_t)$  is an independent Gamma process then for  $\sigma, b \in \mathbb{R}$  the process

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

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

$$\nu(dx) = \frac{c}{|x|} (e^{-\lambda x} I_{(0,\infty)}(x) + e^{-\mu|x|} I_{(-\infty,0)}(x)) dx$$

with constants  $c, \lambda, \mu > 0$ . Variance Gamma processes have been introduced by Madan and Seneta [23] to include heavy tails in financial models.

### Finite variation Lévy processes

Suppose that  $(N_t)$  is a Poisson point process on  $\mathbb{R} \setminus \{0\}$  with jump intensity measure  $\nu$  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

$$X_t = X_t^{\nearrow} + X_t^{\searrow}, \quad X_t^{\nearrow} = \int y N_t^{(0,\infty)}(dy), \quad X_t^{\searrow} = \int y N_t^{(-\infty,0)}(dy),$$

of the associated Lévy jump process  $X_t = \int y N_t(dy)$  into a subordinator  $X_t^{\nearrow}$  and a decreasing Lévy process  $X_t^{\searrow}$ . Therefore,  $(X_t)$  has almost surely paths of finite variation.

**Example (Variance Gamma process).** A Variance Gamma process  $X_t = \sigma B_{\Gamma_t} + b\Gamma_t$  satisfies  $X_t = \Gamma_t^{(1)} - \Gamma_t^{(2)}$  with two independent Gamma processes, cf. e.g. [21]. In particular, the random variables  $X_t, t \geq 0$ , have exponential tails.

### 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  $\eta_j$ . If the random variables  $\eta_j$  are square-integrable with mean zero then Donsker's invariance principle (the "functional central limit theorem") states that the diffusively rescaled process  $(k^{-1/2} S_{[kt]})_{t \geq 0}$  converges in distribution to  $(\sigma B_t)_{t \geq 0}$  where  $(B_t)$  is a Brownian motion

in  $\mathbb{R}^d$  and  $\sigma$  is a non-negative definite symmetric  $d \times d$  matrix.

The functional central limit theorem does not apply if the increments  $\eta_j$  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.,

$$k^{-1/\alpha} X_{kt} \sim X_t \quad \text{for any } k \in (0, \infty) \text{ and } t \geq 0. \quad (1.23)$$

**Definition.** Let  $\alpha \in (0, 2]$ . A Lévy process  $(X_t)$  satisfying (1.23) is called (**strictly**)  **$\alpha$ -stable**.

The reason for the restriction to  $\alpha \in (0, 2]$  is that for  $\alpha > 2$ , an  $\alpha$ -stable process does not exist. This will become clear by the proof of Theorem 1.13 below. There is a broader class of Lévy processes that is called  **$\alpha$ -stable** in the literature, cf. e.g. [21]. Throughout these notes, by an  **$\alpha$ -stable process** we always mean a strictly  $\alpha$ -stable process as defined above.

**Examples.** 1) For  $b \in \mathbb{R}$ , the deterministic process  $X_t = bt$  is a 1-stable Lévy process. 2) 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  $\sigma \in [0, \infty)$ .

More generally, the following characterization of  $\alpha$ -stable processes holds:

**Theorem 1.13.** 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  $\alpha$ -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

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

*Proof.* (i)  $\Leftrightarrow$  (ii). The process  $(X_t)$  is strictly  $\alpha$ -stable if and only if  $X_{c^\alpha t} \sim cX_t$  for any  $c, t \geq 0$ , i.e., if and only if

$$e^{-t\psi(cp)} = E\left[e^{ipcX_t}\right] = E\left[e^{ipX_{c^\alpha t}}\right] = e^{-c^\alpha t\psi(p)}$$

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

$$\psi(p) = \begin{cases} z_+ |p|^\alpha & \text{for } p \geq 0, \\ z_- |p|^\alpha & \text{for } p \leq 0. \end{cases}$$

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  $\psi$  satisfies  $\psi(-p) = \overline{\psi(p)}$  and  $\Re(\psi(p)) \geq 0$ . Therefore,  $z_- = \bar{z}_+$  and  $\Re(z_+) \geq 0$ .  $\square$

**Example (Symmetric  $\alpha$ -stable processes).** A Lévy process in  $\mathbb{R}^d$  with characteristic exponent

$$\psi(p) = \sigma^\alpha |p|^\alpha$$

for some  $\sigma \geq 0$  and  $a \in (0, 2]$  is called a *symmetric  $\alpha$ -stable process*. It can be shown by Fourier transformation that a symmetric  $\alpha$ -stable process is a Markov process with generator  $-\sigma^\alpha (-\Delta)^{\alpha/2}$ . In particular, Brownian motion is a symmetric 2-stable process.

The jump intensity of strictly  $\alpha$ -stable processes can be easily found by an informal argument. Suppose we rescale in space and time by  $y \rightarrow cy$  and  $t \rightarrow c^\alpha t$ . 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  $\alpha$ -stable process in  $\mathbb{R}^1$  should be given by

$$\nu(dy) = (c_+ I_{(0, \infty)}(y) + c_- I_{(-\infty, 0)}(y)) |y|^{-1-\alpha} dy \quad (1.24)$$

with constants  $c_+, c_- \in [0, \infty)$ . Note that for any  $\alpha \in (0, 2)$ , the measure  $\nu$  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  $\alpha$ -stable process with Lévy measure (1.24). 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 construct the corresponding  $\alpha$ -stable processes, let

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

where  $(N_t)$  is a Poisson point process with intensity measure  $\nu$ . Setting  $\|X\|_a = E[\sup_{t \leq a} |X_t|^2]^{1/2}$ , an application of Theorem 1.10 yields:

**Corollary 1.14 (Construction of  $\alpha$ -stable processes).** *Let  $\nu$  be the probability measure on  $\mathbb{R} \setminus \{0\}$  defined by (1.24) with  $c_+, c_- \in [0, \infty)$ .*

- 1) *If  $c_+ = c_-$  then there exists a symmetric  $\alpha$ -stable process  $X$  with characteristic exponent  $\psi(p) = \gamma |p|^\alpha$ ,  $\gamma = \int (1 - \cos y) \nu(dy) \in \mathbb{R}$ , such that  $\|X^{1/n} - X\|_a \rightarrow 0$  for any  $a \in [0, \infty)$ .*
- 2) *If  $\alpha \in (0, 1)$  then there exists an  $\alpha$ -stable process  $X$  with characteristic exponent  $\psi(p) = z |p|^\alpha$ ,  $z = \int (1 - e^{iy}) \nu(dy) \in \mathbb{C}$ , such that  $\|X^{1/n} - X\|_a \rightarrow 0$  for any  $a \in (0, \infty)$ .*
- 3) *Finally, if  $\alpha \in (1, 2)$  then there exists an  $\alpha$ -stable process  $(\tilde{X}_t)$  with characteristic exponent  $\psi(p) = \tilde{z} \cdot |p|^\alpha$ ,  $\tilde{z} = \int (1 - e^{iy} + iy) \nu(dy)$ , such that the compensated processes*

$$\tilde{X}_t^{1/n} = X_t^{1/n} - t \int y I_{\{|y| > 1/n\}} \nu(dy), \quad t \geq 0,$$

*converge to  $\tilde{X}$  w.r.t.  $\|\cdot\|_a$  for any  $a \in (0, \infty)$ .*

*Proof.* By Theorem 1.10 it is sufficient to prove convergence of the characteristic exponents

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

to  $\psi(p)$ ,  $\tilde{\psi}(p)$  respectively as  $\varepsilon \downarrow 0$ . This is easily verified in cases 1), 2) and 3) 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 cases 2) and 3) above take a similar form (but with a different constant), the processes are actually very different. In particular, for  $\alpha > 1$ , a strictly  $\alpha$ -stable process is a limit of compensated compound Poisson processes and hence a martingale!

**Example** ( $\alpha$ -stable subordinators). For  $c_- = 0$  and  $\alpha \in (0, 1)$  the  $\alpha$ -stable process with jump intensity  $\nu$  is increasing, i.e., it is an  $\alpha$ -stable subordinator. For  $c_- = 0$  and  $\alpha \in (1, 2)$  this is not the case since the jumps are “*compensated by an infinite drift*”. For  $\alpha \in (0, 2)$ , a symmetric  $\alpha$ -stable process has the same law as  $(\sqrt{2}B_{\tau_s})$  where  $(B_t)$  is a Brownian motion and  $(\tau_s)$  is an  $\alpha/2$ -stable subordinator.

## Chapter 2

# Stochastic integrals and Itô calculus for semimartingales

Our aim in this chapter is to develop a stochastic calculus for functions of finitely many real-valued stochastic processes  $X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(d)}$ . In particular, we will make sense of stochastic differential equations of type

$$dY_t = \sum_{k=1}^d \sigma_k(t, Y_{t-}) dX_t^{(k)}$$

with continuous time-dependent vector fields  $\sigma_1, \dots, \sigma_d : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ . 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. [28, Ch.II, §2] and [27, Ch.I Thm.30].

It will turn out that an adequate class of stochastic processes for which an Itô calculus can be developed are *semimartingales*, i.e., sums of local martingales and adapted finite variation processes with *càdlàg* trajectories. To understand why this could be a reasonable class of processes to consider, we first briefly look at the *discrete time case*.

If  $(\mathcal{F}_n)_{n=0,1,2,\dots}$  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. [9, Thm. 2.4]. The decomposition is determined by choosing

$$M_n - M_{n-1} = X_n - X_{n-1} - E[X_n - X_{n-1} | \mathcal{F}_{n-1}],$$

$$A_n^{\nearrow} - A_{n-1}^{\nearrow} = E[X_n - X_{n-1} | \mathcal{F}_{n-1}]^+, \quad \text{and} \quad A_n^{\searrow} - A_{n-1}^{\searrow} = E[X_n - X_{n-1} | \mathcal{F}_{n-1}]^-.$$

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

$$(G \bullet X)_n = \sum_{k=1}^n G_k (X_k - X_{k-1})$$

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 and increasing (respectively decreasing).

For a bounded adapted process  $(H_n)$ , we can define correspondingly the integral

$$(H_- \bullet X)_n = \sum_{k=1}^n H_{k-1} (X_k - X_{k-1})$$

of the predictable process  $H_- = (H_{k-1})_{k \in \mathbb{N}}$  w.r.t.  $X$ .

In continuous time, it is no longer true that any adapted process can be decomposed into a local martingale and an adapted process of finite variation (i.e., the sum of an increasing and a decreasing process). A counterexample is given by fractional Brownian motion, cf. Section 2.3 below. On the other hand, a large class of relevant processes has a corresponding decomposition.

**Definition.** Let  $(\mathcal{F}_t)_{t \geq 0}$  be a filtration. A real-valued  $(\mathcal{F}_t)$ -adapted stochastic process  $(X_t)_{t \geq 0}$  on a probability space  $(\Omega, \mathcal{A}, P)$  is called an  $(\mathcal{F}_t)$  **semimartingale** if and only if it has a decomposition

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

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 any local martingale can be decomposed into a strict local martingale and an adapted finite variation process (“Fundamental Theorem of Local Martingales”, cf. [27]). 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.

Without additional conditions on  $(A_t)$ , the semimartingale decomposition in (2.2) is not unique. Uniqueness holds if, in addition,  $(A_t)$  is assumed to be continuous, this is a consequence of Corollary 2.15 below.

**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.

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) 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.



The following examples show that semimartingales form a sufficiently rich class of stochastic processes. First examples of semimartingales with jumps have been already given in Chapter 1.

**Example (Solutions of SDE w.r.t. Brownian motion).** Suppose that  $X_t : \Omega \rightarrow \mathbb{R}^n$  is a solution of a stochastic differential equation

$$dX_t = \sigma(t, X_t) dB_t + b(t, X_t) dt \quad (2.3)$$

driven by a  $d$ -dimensional Brownian motion  $(B_t)$ . Then (2.3) directly yields a semimartingale decomposition for the components  $X_t^i$ ,  $i = 1, \dots, n$ , into the local martingale part given by the stochastic integral and the finite variation part given by the integral of the drift term. More generally, the Itô-Doeblin formula yields an explicit semimartingale decomposition of  $F(t, X_t)$  for an arbitrary function  $F \in \mathcal{C}^2(\mathbb{R}_+ \times \mathbb{R}^n)$ , cf. [9, 8.5.1] and Corollary 2.24 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

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

cf. e.g. [7] or [11]. 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).

## 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  $A : [0, u) \rightarrow \mathbb{R}$  be a right-continuous function of finite variation. In particular,  $A$  is càdlàg. We recall that there is a  $\sigma$ -finite measure  $\mu_A$  on  $(0, u)$  with distribution function  $A$ , i.e.,

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

The function  $A$  has the decomposition

$$A_t = A_t^c + A_t^d \quad (2.6)$$

into the pure jump function

$$A_t^d := \sum_{s \leq t} \Delta A_s \quad (2.7)$$

and the continuous function  $A_t^c = A_t - A_t^d$ . Indeed, the series in (2.7) converges absolutely since

$$\sum_{s \leq t} |\Delta A_s| \leq V_t^{(1)}(A) < \infty \quad \text{for any } t \in [0, u].$$

The measure  $\mu_A$  can be decomposed correspondingly into

$$\mu_A = \mu_{A^c} + \mu_{A^d}$$

where

$$\mu_{A^d} = \sum_{\substack{s \in (0, u) \\ \Delta A_s \neq 0}} \Delta A_s \cdot \delta_s$$

is the atomic part, and  $\mu_{A^c}$  does not contain atoms. Note that  $\mu_{A^c}$  is not necessarily absolutely continuous!

### Lebesgue-Stieltjes integrals revisited

Let  $\mathcal{L}_{\text{loc}}^1([0, u], \mu_A) := \mathcal{L}_{\text{loc}}^1([0, u], |\mu_A|)$  where  $|\mu_A|$  denotes the positive measure with distribution function  $V_t^{(1)}(A)$ . For  $G \in \mathcal{L}_{\text{loc}}^1([0, u], \mu_A)$ , the Lebesgue-Stieltjes integral of  $H$  w.r.t.  $A$  is defined as

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

A crucial observation is that the function

$$I_t := \int_0^t G_r dA_r = \int_{(0, t]} G_r \mu_A(dr) \quad , \quad t \in [0, u),$$

is the distribution function of the measure

$$\mu_I(dr) = G_r \mu_A(dr)$$

with density  $G_r$  w.r.t.  $\mu_A$ . This has several important consequences:

1) The function  $I$  is again càdlàg and of finite variation with

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

2)  $I$  decomposes into the continuous and pure jump parts

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

3) For any  $\tilde{G} \in \mathcal{L}_{\text{loc}}^1(\mu_I)$ ,

$$\int_0^t \tilde{G}_r dI_r = \int_0^t \tilde{G}_r G_r dA_r,$$

i.e., if “ $dI = G dA$ ” then also “ $\tilde{G} dI = \tilde{G} G dA$ ”.

**Theorem 2.1 (Riemann sum approximations for Lebesgue-Stieltjes integrals).** *Suppose that  $H : [0, u) \rightarrow \mathbb{R}$  is a càdlàg function. Then for any  $a \in [0, u)$  and for any sequence  $(\pi_n)$  of partitions with  $\text{mesh}(\pi_n) \rightarrow 0$ ,*

$$\lim_{n \rightarrow \infty} \sum_{\substack{s \in \pi_n \\ s < t}} H_s (A_{s' \wedge t} - A_s) = \int_0^t H_{s-} dA_s \quad \text{uniformly for } t \in [0, a].$$

**Remark.** If  $(A_t)$  is continuous then

$$\int_0^t H_{s-} dA_s = \int_0^t H_s dA_s,$$

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 left sided Riemann sums takes the modified form

$$\int_0^t H_{s-} dA_s = \int_0^t H_s dA_s^c + \sum_{s \leq t} H_{s-} \Delta A_s.$$

*Proof.* For  $n \in \mathbb{N}$  and  $t \geq 0$ ,

$$\sum_{\substack{s \in \pi_n \\ s < t}} H_s (A_{s' \wedge t} - A_s) = \sum_{\substack{s \in \pi_n \\ s < t}} \int_{(s, s' \wedge t]} H_r dA_r = \int_{(0, t]} H_{[r]_n} dA_r$$

where  $[r]_n := \max \{s \in \pi_n : s < r\}$  is the next partition point strictly below  $r$ . As  $n \rightarrow \infty$ ,  $[r]_n \rightarrow r$  from below, and thus  $H_{[r]_n} \rightarrow H_{r-}$ . Since the càdlàg function  $H$  is uniformly bounded on the compact interval  $[0, a]$ , we obtain

$$\sup_{t \leq a} \left| \int_0^t H_{[r]_n} dA_r - \int_0^t H_{r-} dA_r \right| \leq \int_{(0, a]} |H_{[r]_n} - H_{r-}| |\mu_A|(dr) \rightarrow 0$$

as  $n \rightarrow \infty$  by dominated convergence.  $\square$

### Product rule

The covariation  $[H, A]$  of two functions  $H, A : [0, u) \rightarrow \mathbb{R}$  w.r.t. a sequence  $(\pi_n)$  of partitions with  $\text{mesh}(\pi_n) \rightarrow 0$  is defined by

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

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  $(\pi_n)$  given by*

$$[H, A]_t = \sum_{0 < s \leq t} \Delta H_s \Delta A_s$$

*Proof.* We again represent the sums as integrals:

$$\sum_{\substack{s \in \pi_n \\ s < t}} (H_{s' \wedge t} - H_s)(A_{s' \wedge t} - A_s) = \int_0^t (H_{[r]_n \wedge t} - H_{[r]_n}) dA_r$$

with  $[r]_n$  as above, and  $[r]_n := \min \{s \in \pi_n : s \geq r\}$ . As  $n \rightarrow \infty$ ,  $H_{[r]_n \wedge t} - H_{[r]_n}$  converges to  $H_r - H_{r-}$ , and hence the integral on the right hand side converges to

$$\int_0^t (H_r - H_{r-}) dA_r = \sum_{r \leq t} \Delta H_r \Delta A_r$$

by dominated convergence. □

**w.r.t.** 1) If  $H$  or  $A$  is continuous then  $[H, A] = 0$ .

2) In general, the proof above shows that

$$\int_0^t H_s dA_s = \int_0^t H_{s-} dA_s + [H, A]_t,$$

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  $H, A : [0, u) \rightarrow \mathbb{R}$  are right continuous functions of finite variation. Then*

$$H_t A_t - H_0 A_0 = \int_0^t H_{r-} dA_r + \int_0^t A_{r-} dH_r + [H, A]_t \quad \text{for any } t \in [0, u). \quad (2.9)$$

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  $(\pi_n)$  such that  $\text{mesh}(\pi_n) \rightarrow 0$ .

In differential notation, (2.9) reads

$$d(HA)_r = H_{r-}dA_r + A_{r-}dH_r + d[H, A]_r.$$

As special cases we note that if  $H$  and  $A$  are continuous then  $HA$  is continuous with

$$d(HA)_r = H_r dA_r + A_r dH_r,$$

and if  $H$  and  $A$  are pure jump functions (i.e.  $H^c = A^c = 0$ ) then  $HA$  is a pure jump function with

$$\Delta(HA)_r = H_{r-}\Delta A_r + A_{r-}\Delta H_r + \Delta A_r\Delta H_r.$$

Note that the latter statement is not completely trivial. Indeed, it says that

$$H_t A_t - H_0 A_0 = \sum_{r \leq t} \Delta(HA)_r$$

holds even when the jump times of  $HA$  are a countable dense subset of  $[0, t]$ !

Since the product rule is crucial but easy to prove, we give two proofs

**Proof 1.** For  $(\pi_n)$  with  $\text{mesh}(\pi_n) \rightarrow 0$ , we have

$$\begin{aligned} H_t A_t - H_0 A_0 &= \sum_{\substack{s \in \pi_n \\ s < t}} (H_{s' \wedge t} A_{s' \wedge t} - H_s A_s) \\ &= \sum H_s (A_{s' \wedge t} - A_s) + \sum A_s (H_{s' \wedge t} - H_s) + \sum (A_{s' \wedge t} - A_s) (H_{s' \wedge t} - H_s). \end{aligned}$$

As  $n \rightarrow \infty$ , (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.  $\square$

**Proof 2.** Note that for  $t \in [0, u)$ ,

$$(H_t - H_0)(A_t - A_0) = \int_{(0,t] \times (0,t]} \mu_H(dr) \mu_A(ds)$$

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

$$\int_{s < r} + \int_{s > r} + \int_{s=r} = \int_0^t (A_{r-} - A_0) dH_r + \int_0^t (H_{s-} - H_0) dA_s + \sum_{s \leq t} \Delta H_s \Delta A_s,$$

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  $A : [0, u) \rightarrow \mathbb{R}$  is right continuous with finite variation, and let  $F \in \mathcal{C}^1(\mathbb{R})$ . Then for any  $t \in [0, u)$ ,*

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

or, equivalently,

$$F(A_t) - F(A_0) = \int_0^t F'(A_{s-}) dA_s^c + \sum_{s \leq t} (F(A_s) - F(A_{s-})). \quad (2.11)$$

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

$$F(A_t) - F(A_0) = \sum_{s \leq t} (F(A_s) - F(A_{s-}))$$

**Remark.** Note that by Taylor's theorem, the sum in (2.10) converges absolutely whenever  $\sum_{s \leq t} (\Delta A_s)^2 < \infty$ . 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  $p_n \rightarrow F$  and  $p'_n \rightarrow F'$  uniformly on the bounded set  $\{A_s \mid s \leq t\}$ . Since (2.11) holds for the polynomials  $p_n$ , it also holds for  $F$ .  $\square$

## Exponentials of finite variation functions

Let  $A : [0, \infty) \rightarrow \mathbb{R}$  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

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

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*

$$Z_t = \exp(A_t^c) \cdot \prod_{s \leq t} (1 + \Delta A_s), \quad (2.13)$$

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

*Proof.* 1) We first show convergence of the product

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

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

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

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

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,

$$P_t = P_0 + \sum_{s \leq t} \Delta P_s = 1 + \sum_{s \leq t} P_{s-} \Delta A_s = 1 + \int_0^t P_{s-} dA_s^d, \quad \forall t \geq 0, \quad (2.15)$$

i.e.,  $P$  is the exponential of the pure jump part  $A_t^d = \sum_{s \leq t} \Delta A_s$ .

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

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



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

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

with zero initial condition. Therefore,

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

where  $V_t := V_t^{(1)}(A)$  is the variation of  $A$  and  $M_t := \sup_{s \leq t} |X_s|$ . Iterating the estimate yields

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

by the chain rule, and

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

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  $n \rightarrow \infty$ , the right hand side in (2.16) converges to 0 since  $M_t$  and  $V_t$  are finite. Hence  $X_t = 0$  for each  $t \geq 0$ .  $\square$

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

$$\mathcal{E}_t^A = 1 + \sum_{k=1}^n \int_{(0,t]} \int_{(0,s_1)} \cdots \int_{(0,s_{k-1})} dA_{s_k} dA_{s_{k-1}} \cdots dA_{s_1} + R_t^{(n)},$$

where the remainder term can be estimated by

$$|R_t^{(n)}| \leq M_t V_t^{n+1}/(n+1)!.$$

If  $A$  is continuous then the iterated integrals can be evaluated explicitly:

$$\int_{(0,t]} \int_{(0,s_1)} \cdots \int_{(0,s_{k-1})} dA_{s_k} dA_{s_{k-1}} \cdots dA_{s_1} = (A_t - A_0)^k/k!.$$

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,

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

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

**Theorem 2.6.** *If  $A, B : [0, \infty) \rightarrow \mathbb{R}$  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  $\lambda$ . Geometric Poisson processes are relevant for financial models, cf. e.g. [30]. The equation (2.17) can be interpreted pathwise as the Stieltjes integral equation

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

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

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

which has the unique solution

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

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

$$S_t = (1 + \sigma)^{N_t} e^{-\lambda\sigma t} \quad .$$

Corresponding exponential martingales occur as “likelihood ratio” when the intensity of a Poisson process is modified, cf. Chapter 3 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

$$X_t = \sum_{j=1}^{K_t} \eta_j$$

be a compound Poisson process on  $\mathbb{R}^d$  with jump intensity measure  $\nu = \lambda\mu$  where  $\lambda \in (0, \infty)$  and  $\mu$  is a probability measure on  $\mathbb{R}^d \setminus \{0\}$ . Hence the  $\eta_j$  are i.i.d.  $\sim \mu$ , and  $(K_t)$  is an independent Poisson process with intensity  $\lambda$ . 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”)

$$Z_t = e^{(\lambda - \bar{\lambda})t} \prod_{j=1}^{N_t} \varrho(\eta_j) = e^{(\lambda - \bar{\lambda})t} \prod_{\substack{s \leq t \\ \Delta X_s \neq 0}} \varrho(\Delta X_s).$$

In Chapter 3, 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

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

Here  $N_t = \sum_{j=1}^{K_t} \delta_{\eta_j}$  denotes the corresponding Poisson point process with intensity measure  $\nu$ . Note that  $(A_t)$  is a martingale, since it is a compensated compound Poisson process

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

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

$$Z_t = 1 + \int_0^t Z_{s-} dA_s \quad (2.19)$$

in the equivalent form

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

where  $\tilde{N}(ds dy) := N(ds dy) - ds \nu(dy)$  is the random measure on  $\mathbb{R}^+ \times \mathbb{R}^d$  with  $\tilde{N}((0, t] \times B) = \tilde{N}_t(B)$  for any  $t \geq 0$  and  $B \in \mathcal{B}(\mathbb{R}^d)$ . In differential notation, (2.20) is an SDE driven by the compensated Poisson point process  $(\tilde{N}_t)$ :

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

**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 [29] 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,

$$(\mathcal{L}f)(a) := \sum_{b \in S} \mathcal{L}(a, b)f(b) = \sum_{b \in S, b \neq a} \mathcal{L}(a, b)(f(b) - f(a))$$

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

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

is an  $(\mathcal{F}_t^X)$  martingale for any  $f : S \rightarrow \mathbb{R}$ . Indeed, this is a direct consequence of the Markov property and the Kolmogorov forward equation, which imply

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

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

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

is a martingale, and, in differential notation,

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

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

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

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

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

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

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

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

is the amount of time spent at  $a$  before time  $t$  (“local time at  $a$ ”). In the form of an SDE,

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

More generally, for any function  $g : S \times S \rightarrow \mathbb{R}$ , 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),

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

Finally, the exponentials of these martingales are again local martingales. For example, we find that

$$\mathcal{E}_t^{\alpha N^{a,b}} = (1 + \alpha)^{\mathcal{J}_t^{a,b}} \exp(-\alpha \mathcal{L}(a,b) L_t^a)$$

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  $f : S \rightarrow \mathbb{R}$ . We assume  $\mathcal{L}(a,b) > 0$  for  $a \neq b$ . Let

$$g(a,b) := \tilde{\mathcal{L}}(a,b)/\mathcal{L}(a,b) - 1 \quad \text{for } a \neq b, \quad g(a,a) := 0,$$

where  $\tilde{\mathcal{L}}$  is another Q-matrix.

- 1) Let  $\lambda(a) = \sum_{b \neq a} \mathcal{L}(a, b) = -\mathcal{L}(a, a)$  and  $\tilde{\lambda}(a) = -\tilde{\mathcal{L}}(a, a)$  denote the total jump intensities at  $a$ . We define a “likelihood quotient” for the trajectories of Markov chains with generators  $\tilde{\mathcal{L}}$  and  $\mathcal{L}$  by  $Z_t = \tilde{\zeta}_t / \zeta_t$  where

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

and  $\zeta_t$  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  $\tilde{P}$  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  $f : S \rightarrow \mathbb{R}$ ,

$$\tilde{M}_t^{[f]} := f(X_t) - f(X_0) - \int_0^t (\tilde{\mathcal{L}}f)(X_s) ds$$

is a martingale w.r.t.  $\tilde{P}$ . Hence under the new probability measure  $\tilde{P}$ ,  $(X_t)$  is a Markov chain with generator  $\tilde{\mathcal{L}}$ .

*Hint: You may assume without proof that  $(\tilde{M}_t^{[f]})$  is a local martingale w.r.t.  $\tilde{P}$  if and only if  $(Z_t \tilde{M}_t^{[f]})$  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

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

is left continuous with right limits and  $(\mathcal{F}_t^P)$  adapted. If  $\pi$  is a partition of  $\mathbb{R}_+$  then we can approximate  $H$  and  $H_-$  by the elementary process (step function)

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

respectively. The process  $H^\pi$  is again càdlàg and adapted, whereas  $H_-^\pi$  is left continuous and (hence) previsible. We consider the Riemann sum approximations

$$I_t^\pi := \sum_{\substack{s \in \pi \\ s < t}} H_s (M_{s' \wedge t} - M_s)$$

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

$$I_t^\pi = \int_0^t H_{s-}^\pi dM_s \quad .$$

We remark that a straightforward pathwise approach for the existence of the limit of  $I^\pi(\omega)$  as  $\text{mesh}(\pi) \rightarrow 0$  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  $(\pi_n)$  is a sequence of partitions of  $\mathbb{R}_+$  with  $\text{mesh}(\pi_n) \rightarrow 0$ . 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  $h : [0, t] \rightarrow \mathbb{R}$  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. [22], [13]). Here, following the standard development of stochastic calculus, we also restrict the class of integrands further (to previsible 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  $I^\pi$  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*

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

holds for any partitions  $\pi$  and  $\tilde{\pi}$  of  $\mathbb{R}_+$  with  $\text{mesh}(\pi) < \delta$  and  $\text{mesh}(\tilde{\pi}) < \delta$ .

The constant  $\Delta$  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. [11]), or, alternatively, a lengthy discussion of general semimartingale theory (cf. [29]).

*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  $\pi$  and  $\tilde{\pi}$  contain the end point  $a$ , and  $\pi$  is a refinement of  $\tilde{\pi}$ . 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

$$I_a^\pi - I_a^{\tilde{\pi}} = \sum_{s \in \pi} (H_s - H_{[s]})(M_{s'} - M_s) \quad (2.28)$$

where from now on, we only sum over partition points less than  $a$ ,  $s'$  denotes the successor of  $s$  in the fine partition  $\pi$ , and

$$[s] := \max \{t \in \tilde{\pi} : t \leq s\}$$

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

$$\begin{aligned} \|I^\pi - I^{\tilde{\pi}}\|_{M^2([0,a])}^2 &= E[(I_a^\pi - I_a^{\tilde{\pi}})^2] \\ &= E\left[\sum_{s \in \pi} (H_s - H_{[s]})^2 (M_{s'} - M_s)^2\right] \\ &\leq \varepsilon^2 E\left[\sum_{s \in \pi} (M_{s'} - M_s)^2\right] + (2C)^2 E\left[\sum_{t \in \tilde{\pi}} \sum_{\substack{s \in \pi \\ \tau_t(\varepsilon) \leq s < [t]}} (M_{s'} - M_s)^2\right] \end{aligned} \quad (2.29)$$

where  $[t] := \min\{u \in \tilde{\pi} : u > t\}$  is the next partition point of the rough partition, and

$$\tau_t(\varepsilon) := \min\{s \in \pi, s > t : |H_s - H_t| > \varepsilon\} \wedge [t].$$

is the first time after  $t$  where  $H$  deviates substantially from  $H_s$ . Note that  $\tau_t$  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:

$$E\left[\sum (M_{s'} - M_s)^2\right] = \sum E[M_{s'}^2 - M_s^2] = E[M_a^2 - M_0^2] \leq C^2. \quad (2.30)$$

The second summand is more difficult to handle. Noting that

$$E[(M_{s'} - M_s)^2 | \mathcal{F}_{\tau_t}] = E[M_{s'}^2 - M_s^2 | \mathcal{F}_{\tau_t}] \quad \text{on } \{\tau_t \leq s\},$$

we can rewrite the expectation value as

$$\begin{aligned} &\sum_{t \in \tilde{\pi}} E\left[\sum_{\tau_t \leq s < [t]} E[(M_{s'} - M_s)^2 | \mathcal{F}_{\tau_t}]\right] \\ &= \sum_{t \in \tilde{\pi}} E[E[M_{[t]}^2 - M_{\tau_t}^2 | \mathcal{F}_{\tau_t}]] = E\left[\sum_{t \in \tilde{\pi}} (M_{[t]} - M_{\tau_t})^2\right] =: B \end{aligned} \quad (2.31)$$

Note that  $M_{[t]} - M_{\tau_t} \neq 0$  only if  $\tau_t < [t]$ , i.e., if  $H$  oscillates more than  $\varepsilon$  in the interval  $[t, \tau_t]$ . We can therefore use the càdlàg property of  $H$  and  $M$  to control (2.31). Let

$$D_{\varepsilon/2} := \{s \in [0, a] : |H_s - H_{s-}| > \varepsilon/2\}$$

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)  $|u - v| \leq \delta \Rightarrow |H_u - H_v| \leq \varepsilon$  or  $(u, v] \cap D_{\varepsilon/2} \neq \emptyset$  ,  
(ii)  $r \in D_{\varepsilon/2}$  ,  $u, v \in [r, r + \delta] \Rightarrow |M_u - M_v| \leq \bar{\varepsilon}$ .

Here we have used that  $H$  is càdlàg,  $D_{\varepsilon/2}$  is finite, and  $M$  is right continuous.

By (i) and (ii), the following implication holds on  $\{\Delta \leq \delta\}$ :

$$\tau_t > [t] \Rightarrow |H_{\tau_t} - H_t| > \varepsilon \Rightarrow [t, \tau_t] \cap D_{\varepsilon/2} \neq \emptyset \Rightarrow |M_{[t]} - M_{\tau_t}| \leq \bar{\varepsilon},$$

i.e., if  $\tau_t > [t]$  and  $\Delta \leq \delta$  then the increment of  $M$  between  $\tau_t$  and  $[t]$  is small.

Now fix  $k \in \mathbb{N}$  and  $\bar{\varepsilon} > 0$ . Then we can decompose  $B = B_1 + B_2$  where

$$B_1 = E\left[\sum_{t \in \tilde{\pi}} (M_{[t]} - M_{\tau_t})^2; \Delta \leq \delta, |D_{\varepsilon/2}| \leq k\right] \leq k\bar{\varepsilon}^2, \quad (2.32)$$

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

In the last step we have used the following upper bound for the martingale increments

$\eta_t := M_{[t]} - M_{\tau_t}$ :

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

This estimate holds by the optional sampling, and since  $E[\sum_{u>t} \eta_u^2 \mid \mathcal{F}_t] \leq E[M_u^2 - M_t^2 \mid \mathcal{F}_t] \leq C^2$  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),

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

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  $\gamma$  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  $4\sqrt{6} C^4 P[|D_{\varepsilon/2}| > k]^{1/2} < \gamma/4$ ,
3. Choose  $\bar{\varepsilon} > 0$  such that  $4C^2 k \bar{\varepsilon}^2 < \gamma/4$ , then choose the random variable  $\delta$  depending on  $\varepsilon$  and  $\bar{\varepsilon}$  such that (i) and (ii) hold.
4. Choose  $\Delta > 0$  such that  $4\sqrt{6} C^4 P[\Delta > \delta]^{1/2} < \gamma/4$ .

Then for this choice of  $\Delta$  we finally obtain

$$\|I^\pi - I^{\tilde{\pi}}\|_{M^2([0,a])}^2 < 4 \cdot \frac{\gamma}{4} = \gamma$$

whenever  $\text{mesh}(\tilde{\pi}) \leq \delta$  and  $\pi$  is a refinement of  $\tilde{\pi}$ . □

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

$$H_{\bullet}M|_{[0,a]} = \lim_{n \rightarrow \infty} H_{\bullet}^{\pi_n}M|_{[0,a]} \quad \text{in } M_a^2([0, a])$$

for any  $a \in (0, \infty)$  and for any sequence  $(\pi_n)$  of partitions of  $\mathbb{R}_+$  with  $\text{mesh}(\pi_n) \rightarrow 0$ .

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)  $\Delta(H_{\bullet}M) = H_{\bullet}\Delta M$  almost surely.

3) If  $T : \Omega \rightarrow [0, \infty]$  is a random variable, and  $H, \tilde{H}, M, \tilde{M}$  are processes as above such that  $H_t = \tilde{H}_t$  for any  $t < T$  and  $M_t = \tilde{M}_t$  for any  $t \leq T$  then, almost surely,

$$H_{-\bullet}M = \tilde{H}_{-\bullet}\tilde{M} \quad \text{on } [0, T].$$

*Proof.* The statements follow easily by Riemann sum approximation. Indeed, let  $(\pi_n)$  be a sequence of partitions of  $\mathbb{R}_+$  such that  $\text{mesh}(\pi_n) \rightarrow 0$ . Then almost surely along a subsequence  $(\tilde{\pi}_n)$ ,

$$(H_{-\bullet}M)_t = \lim_{n \rightarrow \infty} \sum_{\substack{s \leq t \\ s \in \tilde{\pi}_n}} H_s (M_{s' \wedge t} - M_s)$$

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

$$\Delta(H_{-\bullet}M)_t = \lim_{n \rightarrow \infty} H_{[t]_n} (M_t - M_{[t]_n}) = H_{t-} \Delta M_t \quad (2.35)$$

almost surely, where  $[t]_n$  denotes the next partition point of  $(\tilde{\pi}_n)$  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.  $\square$

## 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  $(M_t)_{t \geq 0}$  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

$$M_t^T := M_{t \wedge T} \quad \text{for any } t \geq 0.$$

**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.,

$$\sup \{ |\Delta M_t(\omega)| : 0 \leq t \leq T_n(\omega), \omega \in \Omega \} < \infty \quad \forall n \in \mathbb{N}.$$

**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. [27]. 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  $\nu$  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  $M_t = \int y I_{\{|y|\leq 1\}} (N_t(dy) - t\nu(dy))$  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  $\Delta M^{T_n}$  is uniformly bounded for each  $n$ . Then

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

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

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

□

**Definition (Stochastic integrals for 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

$$H_{-\bullet}M|_{[0,T]} = \tilde{H}_{-\bullet}\tilde{M}|_{[0,T]} \quad a.s., \quad (2.36)$$

whenever  $T$  is an  $(\mathcal{F}_t^P)$  stopping time,  $\tilde{H}$  is a bounded càdlàg  $(\mathcal{F}_t^P)$  adapted process with  $H|_{[0,T]} = \tilde{H}|_{[0,T]}$  almost surely, and  $\tilde{M}$  is a bounded càdlàg  $(\mathcal{F}_t^P)$  martingale with  $M|_{[0,T]} = \tilde{M}|_{[0,T]}$  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  $\tilde{H}_t$  and  $H_t$  only have to agree for  $t < T$ , so we may choose  $\tilde{H}_t = H_t \cdot I_{\{t < T\}}$ . 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

$$H_{-\bullet}M = H_{-\bullet}^{(n)}M^{T_n} \quad \text{on } [0, T_n],$$

and, by Lemma 2.8,  $\Delta(H_{-\bullet}^{(n)}M^{T_n}) = H_{-\bullet}^{(n)}\Delta M^{T_n}$  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

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

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 for 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

$$H_{-\bullet}X = H_{-\bullet}M + H_{-\bullet}A,$$

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  $(\pi_n)$  is a sequence of partitions of  $\mathbb{R}_+$  with  $\text{mesh}(\pi_n) \rightarrow 0$ . Then for any  $a \in [0, \infty)$ ,

$$(H_{-\bullet}X)_t = \lim_{n \rightarrow \infty} \sum_{\substack{s \in \pi_n \\ s < t}} H_s (X_{s' \wedge t} - X_s)$$

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

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

## 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  $M_{\text{loc}}$  and FV respectively. Moreover,

$$\mathcal{S} = M_{\text{loc}} + \text{FV}$$

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  $\mathcal{S} \times \mathcal{S} \rightarrow \mathcal{S}$ ,  $(H, X) \mapsto H \bullet X$  on the equivalence classes that maps  $\mathcal{S} \times M_{\text{loc}}$  to  $M_{\text{loc}}$  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

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

By Theorem (2.10), for  $H, X \in \mathcal{S}$  and any sequence of partitions with  $\text{mesh}(\pi_n) \rightarrow 0$ , 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  $\mathcal{S} \times \mathcal{S} \rightarrow \text{FV}$ . Instead of going once more through the Riemann sum approximations, we can use what we have shown for stochastic integrals and define the covariation by the integration by parts identity

$$X_t Y_t - X_0 Y_0 = \int_0^t X_{s-} dY_s + \int_0^t Y_{s-} dX_s + [X, Y]_t.$$

The approximation by sums is then a direct consequence of Theorem 2.10.

**Definition (Covariation of semimartingales).** For  $X, Y \in \mathcal{S}$ ,

$$[X, Y] := XY - X_0 Y_0 - \int X_- dY - \int Y_- dX.$$

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

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

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

$$[X] = [X, X]$$

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

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

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

In particular, the following statements hold almost surely:

- 1)  $[X]$  is non-decreasing, and  $[X, Y]$  has finite variation.
- 2)  $\Delta[X, Y] = \Delta X \Delta Y$ .
- 3)  $[X, Y]^T = [X^T, Y] = [X, Y^T] = [X^T, Y^T]$ .
- 4)  $|[X, Y]| \leq [X]^{1/2} [Y]^{1/2}$ .

*Proof.* (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.  $\square$

Statements 1) and 2) of the corollary show that  $[X, Y]$  is a finite variation process with decomposition

$$[X, Y]_t = [X, Y]_t^c + \sum_{s \leq t} \Delta X_s \Delta Y_s \quad (2.38)$$

into a continuous part and a pure jump part.

If  $Y$  has finite variation then by Lemma 2.2,

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

Thus  $[X, Y]^c = 0$  and if, moreover,  $X$  or  $Y$  is continuous then  $[X, Y] = 0$ .

More generally, if  $X$  and  $Y$  are semimartingales with decompositions  $X = M + A$ ,  $Y = N + B$  into  $M, N \in M_{\text{loc}}$  and  $A, B \in \text{FV}$  then by bilinearity,

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

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  $M \in \mathcal{M}_a^2([0, a])$  if and only if  $M_0 \in \mathcal{L}^2$  and  $[M]_a \in \mathcal{L}^1$ . In this case,  $M_t^2 - [M]_t$  ( $0 \leq t \leq a$ ) is a martingale, and*

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

*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,

$$E[M_{t \wedge T_n}^2] = E[[M]_{t \wedge T_n}] \quad \text{for any } t \geq 0 \quad \text{and any } n \in \mathbb{N}. \quad (2.40)$$

Since  $M^2$  is a submartingale, we have

$$E[M_t^2] \leq \liminf_{n \rightarrow \infty} E[M_{t \wedge T_n}^2] \leq E[M_t^2]$$

by Fatun's lemma. Moreover, by the monotone convergence theorem,

$$E[[M]_t] = \lim_{n \rightarrow \infty} E[[M]_{t \wedge T_n}].$$

Hence by (2.39), we obtain

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

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

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

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

**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  $[M]_\infty = \lim_{t \rightarrow \infty} [M]_t$  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  $M_\infty = \lim_{t \rightarrow \infty} M_t$  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  $M_d^2([0, a])$  in the sense that*

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

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

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

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

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

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,*

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

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

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

The assertion also extends to the case when  $a$  is replaced by a stopping time. Combined with the existence of the quadratic variation, we have now proven:

*»Non-constant strict local martingales have non-trivial quadratic variation«*

**Example (Fractional Brownian motion is not a semimartingale).** Fractional Brownian motion with Hurst index  $H \in (0, 1)$  is defined as the unique continuous Gaussian process  $(B_t^H)_{t \geq 0}$  satisfying

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

for any  $s, t \geq 0$ . It has been introduced by Mandelbrot as an example of a self-similar process and is used in various applications, cf. [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  $\alpha$ -Hölder continuous if and only if  $\alpha < H$ , cf. e.g. [13]. Furthermore,

$$V_t^{(1)}(B^H) = \infty \quad \text{for any } t \geq 0 \text{ almost surely, and}$$

$$[B^H]_t = \lim_{n \rightarrow \infty} \sum_{\substack{s \in \pi_n \\ s < t}} (B_{s' \wedge t}^H - B_s^H)^2 = \begin{cases} 0 & \text{if } H > 1/2, \\ t & \text{if } H = 1/2, \\ \infty & \text{if } H < 1/2. \end{cases}$$

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

$$[M] = [B^H] = 0 \quad \text{almost surely,}$$

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 *nota 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. [27]). 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. [13].

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  $A \in \text{FV}$  such that*

- (i)  $MN - A \in M_{loc}$ , and
- (ii)  $\Delta A = \Delta M \Delta N$ ,  $A_0 = 0$  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  $\tilde{A}$  is another process in FV satisfying (i) and (ii).

Then  $A - \tilde{A}$  is both in  $M_{loc}$  and in FV, and  $\Delta(A - \tilde{A}) = 0$  almost surely. Hence  $A - \tilde{A}$  is a continuous local martingale of finite variation, and thus  $A - \tilde{A} = A_0 - \tilde{A}_0 = 0$  almost surely by Corollary 2.15.  $\square$

The covariation of two local martingales  $M$  and  $N$  yields a semimartingale decomposition for  $MN$ :

$$MN = \text{local martingale} + [M, N].$$

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  $A \in \text{FV}$  such that

- (i)  $MN - A \in M_{loc}$ , and
- (ii)  $\Delta A = 0$ ,  $A_0 = 0$  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 a weaker condition, cf. e.g. [27]. 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,

$$E[\langle M \rangle_t] = \lim_{n \rightarrow \infty} E[\langle M \rangle_{t \wedge T_n}] = E[[M]_{t \wedge T_n}] = E[[M]_t]$$

for any  $t \geq 0$ . The assertions of Theorem 2.13 now follow similarly as above.  $\square$

**Examples (Covariations of Lévy processes).**

1) *Brownian motion*: If  $(B_t)$  is a Brownian motion in  $\mathbb{R}^d$  then the components  $(B_t^k)$  are independent one-dimensional Brownian motions. Therefore, the processes  $B_t^k B_t^l - \delta_{kl}t$  are martingales, and hence almost surely,

$$[B^k, B^l]_t = \langle B^k, B^l \rangle_t = \delta_{kl}t \quad \text{for any } t \geq 0.$$

2) *Lévy processes without diffusion part*: Let

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

with  $b \in \mathbb{R}^d$ , a  $\sigma$ -finite measure  $\nu$  on  $\mathbb{R}^d \setminus \{0\}$  satisfying  $\int (|y|^2 \wedge 1) \nu(dy) < \infty$ , and a Poisson point process  $(N_t)$  of intensity  $\nu$ . Suppose first that  $\text{supp}(\nu) \subset \{y \in \mathbb{R}^d : |y| \geq \varepsilon\}$  for some  $\varepsilon > 0$ . Then the components  $X^k$  are finite variation processes, and hence

$$[X^k, X^l]_t = \sum_{s \leq t} \Delta X_s^k \Delta X_s^l = \int y^k y^l N_t(dy). \quad (2.41)$$

In general, (2.41) 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  $\|X^{(\varepsilon),k} - X^k\|_{M^2([0,a])} \rightarrow 0$  as  $\varepsilon \downarrow 0$  for any  $a \in \mathbb{R}_+$  and  $k \in \{1, \dots, d\}$ , and hence

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

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

$$\langle M^k, M^l \rangle_t = t \cdot \frac{\partial^2 \psi}{\partial p_k \partial p_l}(0),$$

whereas  $[M^k, M^l] = [X^k, X^l]$  is a pure jump process.

3) *Covariations of Brownian motion and Lévy jump processes*: For  $B$  and  $X$  as above we have

$$\langle B^k, X^l \rangle = [B^k, X^l] = 0 \quad \text{almost surely for any } k \text{ and } l. \quad (2.42)$$

Indeed, (2.42) 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 (2.42) 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*

$$\left[ \int H_- dX, Y \right] = \int H_- d[X, Y] \quad \text{almost surely.} \quad (2.43)$$

*Proof.* 1. We first note that (2.43) holds if  $X$  or  $Y$  has finite variation paths. If, for example,  $X \in \text{FV}$  then also  $\int H_- dX \in \text{FV}$ , and hence

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

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

2. Now we show that (2.43) holds if  $X$  and  $Y$  are bounded martingales, and  $H$  is bounded. For this purpose, we fix a partition  $\pi$ , and we approximate  $H$  by the elementary adapted càdlàg process  $H^\pi = \sum_{s \in \pi} H_s \cdot I_{[s, s']}$ . Let

$$I_t^\pi = \int_{(0, t]} H_-^\pi dX = \sum_{s \in \pi} H_s (X_{s' \wedge t} - X_s).$$

We can easily verify that

$$[I^\pi, Y] = \int H_-^\pi d[X, Y] \quad \text{almost surely.} \quad (2.44)$$

Indeed, if  $(\tilde{\pi}_n)$  is a sequence of partitions such that  $\pi \subset \tilde{\pi}_n$  for any  $n$  and  $\text{mesh}(\tilde{\pi}_n) \rightarrow 0$  then

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

Since the outer sum has only finitely many non-zero summands, the right hand side converges as  $n \rightarrow \infty$  to

$$\sum_{s \in \pi} H_s ([X, Y]_{s' \wedge t} - [X, Y]_s) = \int_{(0, t]} H_-^\pi d[X, Y],$$

in the ucp sense, and hence (2.44) holds.

Having verified (2.44) for any fixed partition  $\pi$ , we choose again a sequence  $(\pi_n)$  of partitions with  $\text{mesh}(\pi_n) \rightarrow 0$ . Then

$$\int H_- dX = \lim_{n \rightarrow \infty} I^{\pi_n} \quad \text{in } M^2([0, a]) \text{ for any } a \in (0, \infty),$$

and hence, by 2.14 and (2.44),

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

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, 2.11 and 2.12,

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

Since this holds for all localizing stopping times as above, (2.44) is satisfied as well.

4. Finally, suppose that  $X$  and  $Y$  are arbitrary semimartingales. Then  $X = M + A$  and  $Y = N + B$  with  $M, N \in M_{loc}$  and  $A, B \in \text{FV}$ . The assertion (2.43) 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)  $(G_\bullet M)_0 = 0$  , and
- (ii)  $[G_\bullet M, N] = G_\bullet [M, N]$  for any  $N \in M_{loc}$ .

*Proof.* By Theorem 2.18,  $G_\bullet M$  satisfies (i) and (ii). It remains to prove uniqueness. Let  $L \in M_{loc}$  such that  $L_0 = 0$  and

$$[L, N] = G_\bullet[M, N] \quad \text{for any } N \in M_{loc}.$$

Then  $[L - G_\bullet M, N] = 0$  for any  $N \in M_{loc}$ . 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.,

$$L - G_\bullet M \equiv L_0 - (G_\bullet M)_0 = 0.$$

□

**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 a broader class of integrands: Let  $\mathcal{P}$  denote the  $\sigma$ -algebra on  $[0, \infty) \times \Omega$  generated by all  $(\mathcal{F}_t^P)$  adapted left-continuous processes  $(s, \omega) \mapsto G_s(\omega)$ .

**Definition (Stochastic integrals with general predictable integrands).**

- 1) A stochastic process  $(G_t)_{t \geq 0}$  is called **predictable** iff the function  $(s, \omega) \rightarrow G_s(\omega)$  is measurable w.r.t the  $\sigma$ -algebra  $\mathcal{P}$ .
- 2) Let  $M \in M_{loc}$ , and suppose that  $G$  is a predictable process satisfying  $\int_0^t G_s^2 d[M]_s < \infty$  almost surely for any  $t \geq 0$ . If there exists  $G_\bullet M \in M_{loc}$  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

$$[G_\bullet M, G_\bullet M] = G_\bullet[M, G_\bullet M] = G_\bullet^2[M].$$

**Corollary 2.20 (Itô isometry for martingales).** *If  $M \in M_{loc}$  then  $(\int G dM)^2 - \int G^2 d[M] \in M_{loc}$ , and*

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

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

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

□

The Itô isometry for martingales states that the  $M^2([0, a])$  norm of the stochastic integral  $\int G dM$  coincides with the  $L^2([0, a] \times \Omega, P_{[M]})$  norm of the integrand  $(t, \omega) \mapsto G_t(\omega)$ , where  $P_{[M]}$  is the measure on  $\mathbb{R}_+ \times \Omega$  given by

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

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 variable formula.

### Integration w.r.t. stochastic integrals

Suppose that  $X$  and  $Y$  are semimartingales satisfying  $dY = \tilde{G} dX$ , i.e.,  $Y - Y_0 = \int \tilde{G} dX$  for some predictable integrand  $\tilde{G}$ . 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  $\int G dY = \int G \tilde{G} dX$ :

$$dY = \tilde{G} dX \implies G dY = G \tilde{G} dX$$

The covariation characterization of the 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*

$$G_{\bullet}(\tilde{G}_{\bullet}X) = (G\tilde{G})_{\bullet}X \quad (2.45)$$

*holds for any processes  $G = H_{-}$  and  $\tilde{G} = \tilde{H}_{-}$  with  $H$  and  $\tilde{H}$  càdlàg and adapted.*

**Remark.** The assertion extends with a similar proof to more general predictable integrands.

*Proof.* We already know that (2.45) holds for  $X \in \text{FV}$ . Therefore, and by bilinearity of the stochastic integral, we may assume  $X \in M_{\text{loc}}$ . By the Kunita-Watanabe characterization it then suffices to “test” the identity (2.45) with local martingales. For  $N \in M_{\text{loc}}$ , Corollary 2.19 and the associative law for FV processes imply

$$\begin{aligned} [G_{\bullet}(\tilde{G}_{\bullet}X), N] &= G_{\bullet}[\tilde{G}_{\bullet}X, N] = G_{\bullet}(\tilde{G}_{\bullet}[X, N]) \\ &= (G\tilde{G})_{\bullet}[X, N] = [(G\tilde{G})_{\bullet}X, N]. \end{aligned}$$

Thus (2.45) holds. □

## Itô’s formula

We are now going to prove a change of variable formula for discontinuous semimartingales. To get an idea how the formula looks like we first briefly consider a semimartingale  $X \in \mathcal{S}$  with a finite number of jumps in finite time. Suppose that  $0 < T_1 < T_2 < \dots$  are the jump times, and let  $T_0 = 0$ . Then on each of the intervals  $[T_{k-1}, T_k]$ ,  $X$  is continuous. Therefore, by a similar argument as in the proof of Itô’s formula for continuous paths (cf. [9, Thm.6.4]), we could expect that

$$\begin{aligned} F(X_t) - F(X_0) &= \sum_k (F(X_{t \wedge T_k}) - F(X_{t \wedge T_{k-1}})) \\ &= \sum_{T_{k-1} < t} \left( \int_{T_{k-1}}^{t \wedge T_k^-} F'(X_{s-}) dX_s + \frac{1}{2} \int_{T_{k-1}}^{t \wedge T_k^-} F''(X_{s-}) d[X]_s \right) + \sum_{T_k < t} (F(X_{T_k}) - F(X_{T_k-})) \\ &= \int_0^t F'(X_{s-}) dX_s^c + \frac{1}{2} \int_0^t F''(X_{s-}) d[X]_s^c + \sum_{s \leq t} (F(X_s) - F(X_{s-})) \quad (2.46) \end{aligned}$$

where  $X_t^c = X_t - \sum_{s \leq t} \Delta X_s$  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 (2.46) in the equivalent form

$$\begin{aligned} F(X_t) - F(X_0) & \qquad \qquad \qquad (2.47) \\ &= \int_0^t F'(X_{s-}) dX_s + \frac{1}{2} \int_0^t F''(X_{s-}) d[X]_s^c + \sum_{s \leq t} (F(X_s) - F(X_{s-}) - F'(X_{s-}) \Delta X_s), \end{aligned}$$

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

$$\begin{aligned} F(X_t) - F(X_0) &= \sum_{i=0}^d \int_{(0,t]} \frac{\partial F}{\partial x^i}(X_{s-}) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_{(0,t]} \frac{\partial^2 F}{\partial x^i \partial x^j}(X_{s-}) d[X^i, X^j]_s^c \\ & \quad + \sum_{s \in (0,t]} (F(X_s) - F(X_{s-}) - \sum_{i=0}^d \frac{\partial F}{\partial x^i}(X_{s-}) \Delta X_s^i) \end{aligned} \quad (2.48)$$

for any  $t \geq 0$ , almost surely.

**Remark.** The existence of the quadratic variations  $[X^i]_t$  implies the almost sure absolute convergence of the series on the right hand side of (2.48). Indeed, a Taylor expansion up to order two shows that

$$\begin{aligned} \sum_{s \leq t} |F(X_s) - F(X_{s-}) - \sum_{i=0}^d \frac{\partial F}{\partial x^i}(X_{s-}) \Delta X_s^i| &\leq C_t \cdot \sum_{s \leq t} \sum_i |\Delta X_s^i|^2 \\ &\leq C_t \cdot \sum_i [X^i]_t < \infty, \end{aligned}$$

where  $C_t = C_t(\omega)$  is an almost surely finite random constant depending only on the maximum of  $F''$  on the compact set  $\{X_s : s \in [0, t]\}$ .

It is possible to prove this general version of Itô's formula also by a Riemann sum approximation, cf. [27]. Here, following [29], 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 (2.47) is an algebra, i.e.,

$$F, G \in \mathcal{A} \implies F \bullet G \in \mathcal{A}.$$

This is a consequence of the integration by parts formula

$$\begin{aligned} F(X_t)G(X_t) - F(X_0)G(X_0) &= \int_0^t F(X_-) dG(X) + \int_0^t G(X_-) dF(X) \\ &\quad + [F(X), G(X)]^c + \sum_{(0,t]} \Delta F(X) \Delta G(X), \end{aligned} \quad (2.49)$$

the associative law, which implies

$$\begin{aligned} \int F(X_-) dG(X) &= \int F(X_-) G'(X_-) dX + \frac{1}{2} \int F(X_-) G''(X_-) d[X]^c \\ &\quad + \sum F(X_-) (\Delta G(X) - G'(X_-) \Delta X), \end{aligned} \quad (2.50)$$

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

$$\begin{aligned} [F(X), G(X)]^c &= \left[ \int F'(X_-) dX, \int G'(X_-) dX \right]^c \\ &= \left( \int F'(X_-) G'(X_-) d[X] \right)^c = \int (F'G')(X_-) d[X]^c \end{aligned} \quad (2.51)$$

for the continuous part of the covariation. Both (2.50) and 2.51 follow from (2.48) and the corresponding identity for  $G$ . It is straightforward to verify that (2.49), (2.50) and (2.51) imply the change of variable formula (2.47) for  $FG$ , i.e.,  $FG \in \mathcal{A}$ . Therefore, by induction, the formula (2.47) 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  $A \in \text{FV}$ . In this case,  $X$  is uniformly bounded by a finite constant  $C$ . Therefore, there exists a sequence  $(p_n)$  of

polynomials such that  $p_n \rightarrow F$ ,  $p'_n \rightarrow F'$  and  $p''_n \rightarrow F''$  uniformly on  $[-C, C]$ . For  $t \geq 0$ , we obtain

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

w.r.t. convergence in probability. Here we have used an expression of the jump terms in (2.47) by a Taylor expansion. The convergence in probability holds since  $X = M + A$ ,

$$\begin{aligned} E \left[ \left| \int_0^t p'_n(X_{s-}) dM_s - \int_0^t F'(X_{s-}) dM_s \right|^2 \right] \\ = E \left[ \int_0^t (p'_n - F')(X_{s-})^2 d[M]_s \right] \leq \sup_{[-C, C]} |p'_n - F'|^2 \cdot E[[M]_t] \end{aligned}$$

by Itô's isometry, and

$$\left| \sum_{s \leq t} \int_{X_{s-}}^{X_s} \int_{X_{s-}}^y (p''_n - F'') dz dy \right| \leq \frac{1}{2} \sup_{[-C, C]} |p''_n - F''| \sum_{s \leq t} (\Delta X_s)^2.$$

3. Finally, the change of variables formula for general semimartingales  $X = M + A$  with  $M \in M_{\text{loc}}$  and  $A \in \text{FV}$  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

$$A_t^{T_n-} := \begin{cases} A_t & \text{for } t < T_n \\ A_{T_n-} & \text{for } t \geq T_n \end{cases}$$

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  $X$ .  $\square$

Note that the second term on the right hand side of Itô's formula (2.48) 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 (2.48). 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

$$X_t = x + \sigma B_t + bt + \int y \tilde{N}_t(dy) \quad (2.52)$$

with  $x, \sigma, b \in \mathbb{R}$ , a Brownian motion  $(B_t)$ , and a compensated Poisson point process  $\tilde{N}_t = N_t - t\nu$  with intensity measure  $\nu$ . 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  $\infty$ , 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  $C_t = \int y \tilde{N}_t(dy)$  we first note that almost surely,

$$[X]_t = \sigma^2[B]_t + 2\sigma[B, C]_t + [C]_t = \sigma^2 t + \sum_{s \leq t} (\Delta X_s)^2.$$

Therefore, by (2.53),

$$\begin{aligned} & F(X_t) - F(X_0) \\ &= \int_0^t F'(X_-) dX + \frac{1}{2} \int_0^t F''(X_-) d[X]^c + \sum_{s \leq t} (F(X) - F(X_-) - F'(X) \Delta X_-) \\ &= \int_0^t (\sigma F')(X_{s-}) dB_s + \int_0^t (bF' + \frac{1}{2} \sigma^2 F'')(X_s) ds + \int_{(0,t] \times \mathbb{R}} y F'(X_{s-}) \tilde{N}(ds dy) \\ & \quad + \int_{(0,t] \times \mathbb{R}} (F(X_{s-} + y) - F(X_{s-}) - F'(X_{s-})y) N(ds dy), \end{aligned} \quad (2.53)$$

where  $N(ds dy)$  is the Poisson random measure on  $\mathbb{R}_+ \times \mathbb{R}$  corresponding to the Poisson point process, and  $\tilde{N}(ds dy) = N(ds dy) - ds \nu(dy)$ . Here, we have used a rule for evaluating a stochastic integral w.r.t. the process  $C_t = \int y \tilde{N}_t(dy)$  which is intuitively clear. A rigorous proof for this rule will be given in Section 2.5 below. 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 (2.53), we note that the stochastic integrals w.r.t.  $(B_t)$  and w.r.t.  $(\tilde{N}_t)$  are local martingales. In the case of  $\tilde{N}$ , a rigorous proof of this fact is again still missing and will be postponed to Section 2.5. By splitting the last integral on the right hand side of (2.53) into an integral w.r.t.  $\tilde{N}(ds dy)$  (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*

$$M_t^{[F]} = F(X_t) - F(X_0) - \int_0^t (\mathcal{L}F)(X_s) ds,$$

$$(\mathcal{L}F)(x) = \frac{1}{2}(\sigma F'')(x) + (bF')(x) + \int (F(x+y) - F(x) - F'(x)y) \nu(dy),$$

*is a local martingale vanishing at 0. For  $F \in \mathcal{C}_b^2(\mathbb{R})$ ,  $M^{[F]}$  is a martingale, and*

$$(\mathcal{L}F)(x) = \lim_{t \downarrow 0} \frac{1}{t} E[F(X_t) - F(X_0)].$$

*Proof.*  $M^{[F]}$  is a local martingale by the considerations above and since  $X_s(\omega) = X_{s-}(\omega)$  for almost all  $(s, \omega)$ . For  $F \in \mathcal{C}_b^2$ ,  $\mathcal{L}F$  is bounded since  $|F(x+y) - F(x) - F'(x)y| = \mathcal{O}(|y| \wedge |y|^2)$ . Hence  $M^{[F]}$  is a martingale in this case, and

$$\frac{1}{t} E[F(X_t) - F(X_0)] = E\left[\frac{1}{t} \int_0^t (\mathcal{L}F)(X_s) ds\right] \rightarrow (\mathcal{L}F)(x)$$

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  $\alpha$ -stable process with characteristic exponent  $|p|^\alpha$

is  $\mathcal{L} = -(-\Delta)^{\alpha/2}$ . The behaviour of symmetric  $\alpha$ -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  $a < 0$ ,

$$P[T_a < \infty] = \lim_{b \rightarrow \infty} P[T_a < T_b] \leq \exp(ma/2).$$

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

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t, \quad X_0 = x_0, \quad (2.54)$$

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  $\mathbb{R}^d$ ,  $b, \sigma_1, \dots, \sigma_d : \mathbb{R}^+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  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 (2.54) is a continuous  $(\mathcal{F}_t^P)$  semi-martingale  $(X_t)$  satisfying

$$X_t = x_0 + \int_0^t b(s, X_s) ds + \sum_{k=1}^d \int_0^t \sigma_k(s, X_s) dB_s^k \quad \forall t \geq 0 \text{ a.s.} \quad (2.55)$$

If  $X$  is a solution then

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

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  $(X_t)$  yields

$$\begin{aligned} dF(t, X) &= \frac{\partial F}{\partial t}(t, X) + \nabla_x F(t, X) \cdot dX + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 F}{\partial x^i \partial x^j}(t, X) d[X^i, X^j] \\ &= (\sigma^T \nabla_x F)(t, X) \cdot dB + \left( \frac{\partial F}{\partial t} + \mathcal{L}F \right)(t, X) dt, \end{aligned}$$

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-Doebelin formula*

$$\boxed{F(t, X_t) - F(0, X_0) = \int_0^t (\sigma^T \nabla F)(s, X_s) \cdot dB_s + \int_0^t \left( \frac{\partial F}{\partial t} + \mathcal{L}F \right)(s, X_s) ds} \quad (2.56)$$

Again, the formula provides a semimartingale decomposition for  $F(t, X_t)$ . It establishes a connection between the stochastic differential equation (2.54) and partial differential equations involving the operator  $\mathcal{L}$ .

**Example.** Suppose that  $F \in C^2(\mathbb{R}_+ \times \mathbb{R}^n)$  is a classical solution of the pde

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

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

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

Then by (2.56), 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  $[0, t] \times \bar{U}$ , then the process  $M^T$  stopped at the first exit time  $T = \inf \{t \geq 0 : X_t \notin U\}$  is a martingale, and hence

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

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.

Often, the solution of an SDE is only defined up to some explosion  $\zeta$  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

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

Then  $U = \bigcup U_k$ . A solution  $(X_t)$  of the SDE (2.54) up to explosion is a process  $(X_t)_{t \in [0, \zeta) \cup \{0\}}$  such that for every  $n \in \mathbb{N}$ ,  $T_n < \zeta$  almost surely on  $\{\zeta \in (0, \infty)\}$ , and the stopped process  $X^{T_n}$  is a semimartingale satisfying (2.55) for  $t \leq T_n$ . By applying Itô's formula to the stopped processes, we obtain:

**Corollary 2.24 (Martingale problem for Itô diffusions).** *If  $X_t : \Omega \rightarrow U$  is a solution of (2.54) up to the explosion time  $\zeta$ , then for any  $F \in C^2(\mathbb{R}_+ \times U)$  and  $x_0 \in U$ , the process*

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

*is a local martingale up to the explosion time  $\zeta$ , and the stopped processes  $M^{T_n}$ ,  $n \in \mathbb{N}$ , are localizing martingales.*

*Proof.* We can choose functions  $F_n \in C_b^2([0, a] \times U)$ ,  $n \in \mathbb{N}$ ,  $a \geq 0$ , such that  $F_n(t, x) = F(t, x)$  for  $t \in [0, a]$  and  $x$  in a neighbourhood of  $\bar{U}_n$ . Then for  $t \leq a$ ,

$$M_t^{T_n} = M_{t \wedge T_n} = F_n(t, X_{t \wedge T_n}) - \int_0^t \left( \frac{\partial F_n}{\partial t} + \mathcal{L}F_n \right)(s, X_{s \wedge T_n}) ds.$$

By (2.56), the right hand side is a bounded martingale.  $\square$

## Stochastic exponentials

If  $X$  is a continuous semimartingale then by Itô's formula,

$$\mathcal{E}_t^X = \exp \left( X_t - \frac{1}{2} [X]_t \right)$$

is the unique solution of the exponential equation

$$d\mathcal{E}^X = \mathcal{E}^X dX, \quad \mathcal{E}_0^X = 1.$$

In particular,  $\mathcal{E}^X$  is a local martingale if  $X$  is a local martingale. Moreover, if

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

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

$$H_t^n = h_n([X]_t, X_t) \tag{2.57}$$

solves the SDE

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

for any  $n \in \mathbb{N}$ , cf. Section 6.4 in [9]. In particular,  $H^n$  is an iterated Itô integral:

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

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*

$$Z_t = 1 + \int_0^t Z_{s-} dX_s, \quad t \geq 0, \tag{2.58}$$

is given by

$$Z_t = \exp\left(X_t - \frac{1}{2}[X]_t^c\right) \prod_{s \in (0, t]} (1 + \Delta X_s) \exp(-\Delta X_s). \tag{2.59}$$

**Remarks.** 1) In the finite variation case, (2.59) can be written as

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

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 simple expression as in (2.57). 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

$$A_t = \int_0^t \int_0^s dX_r dY_s - \int_0^t \int_0^s dY_r dX_s$$

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  $X \in \text{FV}$ , cf. Theorem 2.5. The key observation is that the product

$$P_t = \prod_{s \in (0, t]} (1 + \Delta X_s) \exp(-\Delta X_s)$$

exists and defines a finite variation pure jump process. This follows from the estimate

$$\sum_{\substack{0 < s \leq t \\ |\Delta X_s| \leq 1/2}} |\log(1 + \Delta X_s) - \Delta X_s| \leq \text{const.} \cdot \sum_{s \leq t} |\Delta X_s|^2 \leq \text{const.} \cdot [X]_t$$

which implies that

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

defines almost surely a finite variation pure jump process. Therefore,  $(P_t)$  is also a finite variation pure jump process. Moreover, the process  $G_t = \exp\left(X_t - \frac{1}{2}[X]_t^c\right)$  satisfies

$$G = 1 + \int G_- dX + \sum (\Delta G - G_- \Delta X) \quad (2.60)$$

by Itô's formula. For  $Z = GP$  we obtain

$$\Delta Z = Z_- \left( e^{\Delta X} (1 + \Delta X) e^{-\Delta X} - 1 \right) = Z_- \Delta X,$$

and hence, by integration by parts and (2.60),

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

This proves that  $Z$  solves the SDE (2.58). Uniqueness of the solution follows from a general uniqueness result for SDE with Lipschitz continuous coefficients, cf. Section 4.1. □

**Example (Geometric Lévy processes).** Consider a Lévy martingale  $X_t = \int y \tilde{N}_t(dy)$  where  $(N_t)$  is a Poisson point process on  $\mathbb{R}$  with intensity measure  $\nu$  satisfying  $\int (|y| \wedge |y|^2) \nu(dy) < \infty$ , and  $\tilde{N}_t = N_t - t\nu$ . We derive an SDE for the semimartingale

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

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

$$\begin{aligned} Z_t - 1 &= \sigma \int_{(0,t]} Z_{s-} dX + \mu \int_{(0,t]} Z_{s-} ds + \sum_{(0,t]} Z_{s-} (e^{\sigma \Delta X} - 1 - \sigma \Delta X) \\ &= \sigma \int_{(0,t] \times \mathbb{R}} Z_{s-} y \tilde{N}(ds dy) + \mu \int_{(0,t]} Z_{s-} ds + \int_{(0,t] \times \mathbb{R}} Z_{s-} (e^{\sigma y} - 1 - \sigma y) N(ds dy). \end{aligned} \quad (2.61)$$

If  $\int e^{2\sigma y} \nu(dy) < \infty$  then (2.61) leads to the semimartingale decomposition

$$dZ_t = Z_{t-} dM_t^\sigma + \alpha Z_{t-} dt, \quad Z_0 = 1, \quad (2.62)$$

where

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

is a square-integrable martingale, and  $\alpha = \mu + \int e^{\sigma y} - 1 - \sigma y \nu(dy)$ . Hence we see that although  $(Z_t)$  again solves an SDE driven by the compensated process  $(\tilde{N}_t)$ , this SDE can not be written as an SDE driven by the Lévy process  $(X_t)$ .

## 2.5 Stochastic integrals with general predictable integrands

So far, we have considered stochastic integrals w.r.t. a semimartingale only for 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 [27].

Throughout this section, we fix a filtered probability space  $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$ . Recall that the **predictable  $\sigma$ -algebra** is the  $\sigma$ -algebra  $\mathcal{P}$  on  $\Omega \times (0, \infty)$  generated by all left-continuous  $(\mathcal{F}_t)$  adapted processes  $(\omega, t) \mapsto G_t(\omega)$ . A process  $(\omega, t) \mapsto G_t(\omega)$  is called predictable iff it is measurable w.r.t.  $\mathcal{P}$ .

- Exercise.** 1) Prove that  $\mathcal{P}$  is generated by the sets  $A \times (s, t]$  with  $A \in \mathcal{F}_s$  and  $0 \leq s \leq t$ .
- 2) Equivalently,  $\mathcal{P}$  is generated by all sets of the form  $A \times (S, T]$  where  $A \in \mathcal{F}_S$ , and  $0 \leq S \leq T$  are  $(\mathcal{F}_t)$  stopping times.

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

$$G_t(\omega) = \sum_{i=0}^{n-1} Z_i(\omega) I_{(t_i, t_{i+1}]}(t)$$

with  $n \in \mathbb{N}$ ,  $0 \leq t_0 < t_1 < \dots < t_n$ , and  $Z_i : \Omega \rightarrow \mathbb{R}$  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

$$(G \bullet X)_t = \int_0^t G_s dX_s := \sum_{i=0}^{n-1} Z_i (X_{t_{i+1} \wedge t} - X_{t_i \wedge t})$$

is again a semimartingale. If  $A$  is a finite variation process then  $G \bullet A$  has finite variation as well. If  $M \in M_d^2(0, \infty)$  is a square-integrable martingale then  $G \bullet M \in M_d^2(0, \infty)$ , and the **Itô isometry**

$$\begin{aligned} \|G \bullet M\|_{M^2(0, \infty)} &= E \left[ \left( \int_0^\infty G dM \right)^2 \right] \\ &= E \left[ \int_0^\infty G^2 d[M] \right] = \int_{\Omega \times \mathbb{R}_+} G^2 dP_{[M]} \end{aligned} \quad (2.63)$$

holds, where

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

is the **Doléans measure** 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.  $\langle M \rangle$ ).** 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  $0 \leq s \leq t$  then

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

Since the collection of these sets  $C$  is an  $\cap$ -stable generator for the predictable  $\sigma$ -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 an integrable Lévy process  $X_t : \Omega \rightarrow \mathbb{R}$  then

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

where  $\psi$  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  $A \in \text{FV}$  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 [27]:

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

The result implies that the Doob-Meyer semimartingale decomposition

$$X = M + A \tag{2.64}$$

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

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

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 (2.63), we obtain:

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

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

Hence the stochastic integral  $\mathcal{J} : \mathcal{E} \rightarrow \mathcal{H}^2$ ,  $\mathcal{J}_X(G) = G_\bullet X$ , has a unique isometric extension to the closure  $\bar{\mathcal{E}}^X$  of  $\mathcal{E}$  in the space of all predictable processes w.r.t. the norm  $\|\cdot\|_X$ .

*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 (2.63),

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

The Itô isometry yields a definition of the stochastic integral  $G_\bullet X$  for  $G \in \bar{\mathcal{E}}^X$ . For  $G = H_-$  with  $H$  càdlàg and adapted, this definition is consistent with the definition given above since, because by Corollary 2.20, the Itô isometry also holds for the integrals defined above, and the isometric extension is unique. The class  $\bar{\mathcal{E}}^X$  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. [27] 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 undefined predictable  $G$  with  $\|G\|_X < \infty$ , the processes  $G^n := G \cdot I_{\{G \leq n\}}$ ,  $n \in \mathbb{N}$ , are predictable and bounded with  $\|G^n - G\|_X \rightarrow 0$ .

□

## 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

$$X_t^{T-} = \begin{cases} X_t & \text{for } t < T, \\ X_{T-} & \text{for } t \geq T > 0, \\ 0 & \text{for } T = 0. \end{cases}$$

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 [27] 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 finally ready to state the definition of stochastic integrals for general pre-

dictable 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.,  $X^{T_n-} \in \mathcal{H}^2$ , and  $\|G\|_{X^{T_n-}} < \infty$  for any  $n \in \mathbb{N}$ . If  $G$  is integrable w.r.t.  $X$  then the **stochastic integral**  $G \bullet X$  is defined by

$$(G \bullet X)_t = \int_0^t G_s dX_s = \int_0^t G_s dX_s^{T_n-} \quad \text{for any } t \in [0, T_n), \quad n \in \mathbb{N}.$$

Of course, one has again 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 to Chapter IV in [27].

**Exercise (Sufficient conditions for integrability of predictable processes).** 1) Prove that if  $G$  is predictable and *locally bounded* in the sense that  $G^{T_n}$  is bounded for a sequence  $(T_n)$  of stopping times with  $T_n \uparrow \infty$ , then  $G$  is integrable w.r.t. any semimartingale  $X \in \mathcal{S}$ .

2) Suppose that  $X = M + A$  is a continuous semimartingale with  $M \in \mathcal{M}_c^{\text{loc}}$  and  $A \in \text{FV}_c$ . Prove that  $G$  is integrable w.r.t.  $X$  if  $G$  is predictable and

$$\int_0^t G_s^2 d[M]_s + \int_0^t |G_s| |dA_s| < \infty \quad \text{a.s. for any } t \geq 0.$$

## 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 [27] 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)  $H \bullet (G \bullet X) = (HG) \bullet X$ .

In all statements,  $X$  is a semimartingale,  $G$  is a process that is integrable w.r.t.  $X$ ,  $T$  is a stopping time, and  $H$  is a process such that  $HG$  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:

$$\int_0^a |G_s| |H_s| |d[X, Y]_s| \leq \left( \int_0^a G_s^2 d[X]_s \right)^{1/2} \left( \int_0^a H_s^2 d[Y]_s \right)^{1/2}, \quad (2.65)$$

$$E \left[ \int_0^a |G| |H| |d[X, Y]| \right] \leq \left\| \left( \int_0^a G^2 d[X] \right)^{1/2} \right\|_{L^p} \left\| \left( \int_0^a H^2 d[Y] \right)^{1/2} \right\|_{L^q}. \quad (2.66)$$

*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.  $X$ ,

$$\left[ \int G dX, Y \right] = \int G [X, Y] \quad \text{almost surely.} \quad (2.67)$$

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

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

*Proof of Theorem 2.28.* We only sketch the main steps briefly, cf. [27] for details. Firstly, one verifies directly that (2.67) 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  $\|G^n - G\|_X \rightarrow 0$ , and

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

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

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

by 2.14. Moreover,

$$\int G^n d[X, Y] \longrightarrow \int G d[X, Y] \quad \text{u.c.p..}$$

by the Kunita-Watanabe inequality. Hence (2.67) 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.  $X$  by.  $\square$

### Dominated Convergence Theorem for stochastic integrals

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

**Theorem 2.29.** *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*

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

*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*

$$G^n \bullet X \longrightarrow G \bullet X \quad \text{u.c.p., as } n \rightarrow \infty.$$

*If, in addition to the assumptions above,  $X$  is in  $\mathcal{H}^2$  and  $\|H\|_X < \infty$  then even*

$$\|G^n \bullet X - G \bullet X\|_{\mathcal{H}^2} \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*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

$$\|G^n\|_X = E \left[ \int_0^\infty |G^n|^2 d[M] + \left( \int_0^\infty |G| |dA| \right)^2 \right] \longrightarrow 0$$

as  $n \rightarrow \infty$  by the Dominated Convergence Theorem for Lebesgue integrals. Hence by the Itô isometry,

$$G^n \bullet X \longrightarrow 0 \quad \text{in } \mathcal{H}^2 \quad \text{as } n \rightarrow \infty.$$

The general case can now be reduced to this case by localization, where  $\mathcal{H}^2$  convergence is replaced by the weaker ucp-convergence.  $\square$

### Integration w.r.t. Poisson point processes

Instead of integrating w.r.t. real-valued martingales it is also possible to define stochastic integrals w.r.t. measure-valued martingales. Here, we focus exclusively on integration w.r.t. compensated Poisson point processes, which is important for studying stochastic calculus for jump processes.

Let  $(S, \mathcal{S}, \nu)$  be a  $\sigma$ -finite measure space, and let  $(N_t(dy))_{t \geq 0}$  be a Poisson point process on  $(S, \mathcal{S})$  with intensity measure  $\nu$ . Our main interest is the case  $S = \mathbb{R}^d$ . As usual, we denote by  $\tilde{N}_t = N_t - t\nu$  the compensated Poisson point process, and by  $N(dt dy)$  and  $\tilde{N}(dt dy)$  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  $\tilde{N}_t(A)$ ,  $\tilde{N}_t(B)$ , and  $\tilde{N}_t(A)\tilde{N}_t(B) - t\nu(A \cap B)$  are in  $M_d^2(0, a)$  for any  $a \in \mathbb{R}_+$ . In particular,

$$\langle \tilde{N}(A), \tilde{N}(B) \rangle_t = t \cdot \nu(A \cap B),$$

whereas

$$[\tilde{N}(A), \tilde{N}(B)]_t = \sum_{(0,t]} \Delta N(A) \Delta N(B) = N_t(A \cap B). \quad (2.68)$$

Our goal is to define stochastic integrals of type

$$(H \bullet \tilde{N})_t = \int_{(0,t] \times S} H_s(y) \tilde{N}(ds dy)$$

for predictable processes  $(\omega, s, y) \mapsto H_s(y)(\omega)$  defined on  $\Omega \times \mathbb{R}_+ \times S$ . Here a process  $H$  is called **predictable** iff it is measurable w.r.t. the  $\sigma$ -algebra generated by all processes  $(\omega, s, y) \mapsto K_s(y)(\omega)$  that are product measurable in  $(\omega, s, y)$ , and left-continuous in  $s$  for any given  $\omega, y$ . We consider first predictable integrands of the form

$$H_s(y)(\omega) = \sum_{k=1}^m G_t^{(k)}(\omega) I_{A_k}(y), \quad \omega \in \Omega, \quad s \in (0, \infty), \quad y \in S, \quad (2.69)$$

with  $m \in \mathbb{N}$ , predictable processes  $G^{(k)} \in \mathcal{L}^2(P \otimes \lambda)$  depending not on  $y$ , and disjoint sets  $A_k \in \mathcal{S}$  with  $\nu(A_k) < \infty$ . In this case, the stochastic integral  $H \bullet \tilde{N}$  is defined by

$$(H \bullet \tilde{N})_t = \sum_{k=1}^m \int_0^t G_s^{(k)} d\tilde{N}_s(A_k). \quad (2.70)$$



Since the processes  $\tilde{N}(A_k)$ ,  $k = 1, \dots, m$ , are (independent) compensated Poisson processes and  $G^{(1)}, \dots, G^{(m)}$  are square-integrable w.r.t.  $P \otimes \lambda$ , the stochastic integral  $H_{\bullet} \tilde{N}$  is a square-integrable martingale. Moreover, the following Itô type isometry holds:

**Theorem 2.30 (Itô isometry for integrals w.r.t. compensated PPP).** *For any process  $H$  as in (2.69) and for  $a \in [0, \infty]$ ,*

$$E[(H_{\bullet} \tilde{N})_a]^2 = E\left[\int_0^a \int H_s(y)^2 \nu(dy) ds\right].$$

*Proof.* Let  $I^{(k)} := \int G^{(k)} dN(A_k)$ . Then  $I^{(k)} \in M_d^2(0, a)$  for any  $k \in \mathbb{N}$  and  $a \in [0, \infty]$  with

$$[I^{(k)}, I^{(l)}] = \int G^{(k)} G^{(l)} d[\tilde{N}(A_k), \tilde{N}(A_l)] \quad \forall k, l \in \mathbb{N}.$$

In particular, by (2.68), the covariation vanishes for  $k \neq l$ . Hence

$$\begin{aligned} E[(H_{\bullet} \tilde{N})_a^2] &= \sum_{k,l} E[I_a^{(k)} I_l^{(k)}] = \sum_{k,l} E[[I^{(k)}, I^{(l)}]_a] \\ &= \sum_k E\left[\int_0^a |G^{(k)}|^2 d[\tilde{N}(A_k)]\right] = \sum_k E\left[\int_0^a |G^{(k)}|^2 d\langle \tilde{N}(A_k) \rangle\right] \\ &= E\left[\int_0^a \sum_k |G^{(k)}|^2 \nu(A_k) dt\right] = E\left[\int_0^a \int H^2 d\nu dt\right], \end{aligned}$$

because  $G^{(k)}$  is predictable. □

Theorem 2.30 shows that for processes  $H$  as in (2.69),

$$\|H_{\bullet} \tilde{N}\|_{M^2} = \|H\|_{L^2(P \otimes \lambda \otimes \nu)}.$$

On the other hand, any predictable process  $H \in L^2(P \otimes \lambda_{(0,\infty)} \otimes \nu)$  is a limit w.r.t. the  $L^2(P \otimes \lambda_{(0,\infty)} \otimes \nu)$  norm of processes  $H^{(k)}$  as in (2.69). Hence isometric extension of the map  $H \mapsto H_{\bullet} \tilde{N}$  can again be used to define  $H_{\bullet} \tilde{N} \in M_d^2(0, \infty)$  for any predictable  $H \in L^2(P \otimes \lambda_{(0,\infty)} \otimes \nu)$  in such a way that

$$H_{\bullet}^{(k)} \tilde{N} \longrightarrow H_{\bullet} \tilde{N} \quad \text{in } M^2 \quad \text{whenever } H^{(k)} \rightarrow H \text{ in } L^2.$$

Again, the definition of stochastic integrals w.r.t.  $\tilde{N}$  can then be extended to locally square integrable predictable processes  $H$  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

$$(H_\bullet \tilde{N})_t = \int h(y) \tilde{N}_t(dy) = \tilde{N}_t^h,$$

i.e.,  $H_\bullet \tilde{N}$  is a Lévy martingale with jump intensity measure  $\nu \circ h^{-1}$ .

**Remark (Integration w.r.t. uncompensated Poisson point processes).** For integrands  $H \in \mathcal{L}^1(P \otimes \lambda \otimes \nu)$ , the integral

$$(H_\bullet N)_t = \int_0^t H_s(y) N(ds dy) = \sum_{\substack{(s,y) \in \text{supp } N \\ s \leq t}} H_s(y)$$

w.r.t. the uncompensated Poisson point process  $N$  is almost surely well-defined as a Lebesgue integral. The process  $H_\bullet N$  has finite variation paths, and

$$H_\bullet \tilde{N} = H_\bullet N - H_\bullet(\lambda \otimes \nu).$$

Hence the construction via the Itô isometry is not required in this case. It is nevertheless useful, since it shows that  $H_\bullet \tilde{N}$  is a square-integrable martingale.

If  $H$  is in  $\mathcal{L}^2(P \otimes \lambda \otimes \nu)$  but not in  $\mathcal{L}^1(P \otimes \lambda \otimes \nu)$  then  $H_\bullet N$  does not necessarily converge, whereas  $H_\bullet \tilde{N}$  is still a well-defined square integrable martingale. We have already exploited this fact in the construction of Lévy processes in Chapter 1 where  $H_s(y)(\omega) = y$ .

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 [18] for a detailed study. We only state the following extension of the associative law, which has already been used in the last section:

**Theorem 2.31.** *Suppose that  $H : \Omega \times \mathbb{R}_+ \times S \rightarrow \mathbb{R}$  is predictable and square-integrable w.r.t.  $P \otimes \lambda \otimes \nu$ , and  $G : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{R}$  is a bounded predictable process. If*

$$X_t := \int_{(0,t] \times S} H_s(y) \tilde{N}(ds dy)$$

then

$$\int_0^t G_s dX_s = \int_{(0,t] \times S} G_s H_s(y) \tilde{N}(ds dy).$$

The proof based on approximation by more elementary processes is left as an exercise.

## Chapter 3

# Transformations and weak solutions of stochastic differential equations

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

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t \quad (3.1)$$

with a  $d$ -dimensional Brownian motion  $(B_t)$  and measurable coefficients  $b : [0, \infty) \times U \rightarrow \mathbb{R}^n$  and  $\sigma : [0, \infty) \times U \rightarrow \mathbb{R}^{n \times d}$ . In applications one is often not interested in the random variables  $X_t : \Omega \rightarrow \mathbb{R}$  themselves but only in their joint distribution. In that case, it is usually irrelevant w.r.t. which Brownian motion  $(B_t)$  the SDE (3.1) 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 (3.1) w.r.t. *some* Brownian motion  $(B_t)$  that is usually *defined through* (3.1).

**Definition.** A *weak solution* of the stochastic differential equation (3.1) 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  $B_t : \Omega \rightarrow \mathbb{R}^d$  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  $b(\cdot, X) \in \mathcal{L}_{a,loc}^1([0, S], \mathbb{R}^n)$ ,  $\sigma(\cdot, X) \in \mathcal{L}_{a,loc}^2([0, S], \mathbb{R}^{n \times d})$ , and

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s \quad \text{for any } t < S \text{ a.s.}$$

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  $(\sigma(\mathcal{F}_t^{B,P}, X_0))_{t \geq 0}$  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  $F : \mathbb{R}_+ \times \mathbb{R}^n \times C(\mathbb{R}_+, \mathbb{R}^d) \rightarrow \mathbb{R}^n$ ,  $(t, x_0, y) \mapsto F_t(x_0, y)$ , 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

$$X_t = F_t(X_0, B) \quad \text{for any } t \geq 0$$

holds almost surely. Hence strong solutions are (almost surely) functions of the given Brownian motion and the initial value!

There are SDE that have weak but no strong solutions. An example is given in Section 3.1. The definition of weak and strong solutions can be generalized to other types of SDE including in particular functional equations of the form

$$dX_t = b_t(X) dt + \sigma_t(X) dB_t$$

where  $(b_t)$  and  $(\sigma_t)$  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 4.

Different types of transformations of a stochastic process  $(X_t, P)$  are useful for constructing weak solutions. These include:

- *Random time changes:*  $(X_t)_{t \geq 0} \rightarrow (X_{T_a})_{a \geq 0}$  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

$(X_t) \rightarrow (\varphi(X_t))$ , random translations  $(X_t) \rightarrow (X_t + H_t)$  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  $\tilde{P}$  such that both measures are mutually absolutely continuous on each of the  $\sigma$ -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 3.1 play a crucial rôle.

### 3.1 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 3.1.** *Let  $\psi : \mathbb{R}^d \rightarrow \mathbb{C}$  be a given function. An  $(\mathcal{F}_t)$  adapted càdlàg process  $X_t : \Omega \rightarrow \mathbb{R}^d$  is an  $(\mathcal{F}_t)$  Lévy process with characteristic exponent  $\psi$  if and only if the complex-valued processes*

$$Z_t^p := \exp(ip \cdot X_t + t\psi(p)) \quad , \quad t \geq 0,$$

*are  $(\mathcal{F}_t)$  martingales, or, equivalently, local  $(\mathcal{F}_t)$  martingales for any  $p \in \mathbb{R}^d$ .*

*Proof.* By Corollary 1.2, the processes  $Z^p$  are martingales if  $X$  is a Lévy process with characteristic exponent  $\psi$ . Conversely, suppose that  $Z^p$  is a local martingale for any  $p \in \mathbb{R}^d$ . Then, since these processes are uniformly bounded on finite time intervals, they are martingales. Hence for  $0 \leq s \leq t$  and  $p \in \mathbb{R}^d$ ,

$$E[\exp(ip \cdot (X_t - X_s)) | \mathcal{F}_s] = \exp(-(t-s)\psi(p)),$$

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  $\sigma$ -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  $\nu$ .
- (ii) For any function  $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$ , the real valued process

$$N_t^f = \int f(y) N_t(dy), \quad t \geq 0,$$

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

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

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 3.1, an  $\mathbb{R}^d$ -valued process  $(X_t)$  is a Brownian motion if and only if the processes  $\exp(ip \cdot X_t + t|p|^2/2)$  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 3.2 (P. Lévy 1948).** *Suppose that  $M^1, \dots, M^d$  are continuous local  $(\mathcal{F}_t)$  martingales with*

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

*Then  $M = (M^1, \dots, M^d)$  is a  $d$ -dimensional Brownian motion.*

The following proof is due to Kunita and Watanabe (1967):

*Proof.* Fix  $p \in \mathbb{R}^d$  and let  $\Phi_t := \exp(ip \cdot M_t)$ . By Itô's formula,

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

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 3.1,  $M$  is a Brownian motion.  $\square$

Lévy's characterization of Brownian motion has a lot of remarkable direct consequences.

**Example (Random orthogonal transformations).** Suppose that  $X_t : \Omega \rightarrow \mathbb{R}^n$  is a solution of an SDE

$$dX_t = O_t dB_t, \quad X_0 = x_0, \quad (3.2)$$

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 value  $x_0 \in \mathbb{R}^n$ . We verify that  $X$  is an  $n$ -dimensional Brownian motion provided

$$O_t(\omega) O_t(\omega)^T = I_n \quad \text{for any } t \geq 0, \quad \text{almost surely.} \quad (3.3)$$

Indeed, by (3.2) and (3.3), 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

$$dR_t = e_r(B_t) \cdot dB_t + \frac{d-1}{2R_t} dt \quad \text{for any } t < T_0$$

where  $T_0$  is the first hitting time of 0 for  $(B_t)$ . By the last example, the process

$$W_t := \int_0^t e_r(B_s) \cdot dB_s, \quad t \geq 0,$$

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

$$dR_t = dW_t + \frac{d-1}{2R_t} dt \tag{3.4}$$

up to the first hitting time of 0. The equation (3.4) 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 [28] and other works by M. Yor.

**Exercise.** a) Let  $(X_t)_{0 \leq t < \zeta}$  be a solution of the Bessel equation

$$dX_t = -\frac{d-1}{2X_t} dt + dB_t, \quad X_0 = x_0,$$

where  $(B_t)_{t \geq 0}$  is a standard Brownian motion and  $d$  is a real constant.

- i) Find a non-constant function  $u : \mathbb{R} \rightarrow \mathbb{R}$  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  $a < 0$ ,

$$P[T_a < \infty] = \lim_{b \rightarrow \infty} P[T_a < T_b] \leq \exp(ma/2).$$

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

$$dX_t = \operatorname{sgn}(X_t) dB_t \quad (3.5)$$

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

- 1)  $X$  is a weak solution of (3.5) 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  $\sigma$ -algebra generated by  $\mathcal{F}_0$  and  $\mathcal{F}^{|X|, P}$ .
- 3) There is no strong solution to (3.5).

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

$$B_t := \int_0^t \operatorname{sgn}(X_s) dX_s$$

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

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

i.e.,  $X$  is a weak solution to (3.5).

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

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

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  $\sigma$ -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  $\sigma$ -algebra for a Brownian motion  $(X_t)$ .

### Lévy characterization of Lévy processes

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

**Theorem 3.3.** *Let  $a \in \mathbb{R}^{d \times d}$ ,  $b \in \mathbb{R}$ , and let  $\nu$  be a  $\sigma$ -finite measure on  $\mathbb{R}^d \setminus \{0\}$  satisfying  $\int (|y| \wedge |y|^2) \nu(dy) < \infty$ . If  $X_t^1, \dots, X_t^d : \Omega \rightarrow \mathbb{R}$  are càdlàg stochastic processes such that*

- (i)  $M_t^k := X_t^k - b^k t$  is a local  $(\mathcal{F}_t)$  martingale for any  $k \in \{1, \dots, d\}$ ,
- (ii)  $[X^k, X^l]_t^c = a^{kl} t$  for any  $k, l \in \{1, \dots, d\}$ , and
- (iii)  $E \left[ \sum_{s \in (r, t]} I_B(\Delta X_s) \middle| \mathcal{F}_r \right] = \nu(B) \cdot (t - r)$  almost surely for any  $0 \leq r \leq t$  and for any  $B \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ ,

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

$$\psi(p) = \frac{1}{2} p \cdot a p - i p \cdot b + \int (1 - e^{ip \cdot y} + ip \cdot y) \nu(dy). \quad (3.6)$$

For proving the theorem, we assume without proof that a local martingale is a semi-martingale even if it is not strict, and that the stochastic integral of a bounded adapted left-continuous integrand w.r.t. a local martingale is again a local martingale, cf. e.g. [27].

*Proof of Theorem 3.3.* We first remark that (iii) implies

$$E \left[ \sum_{s \in (r, t]} G_s \cdot f(\Delta X_s) \middle| \mathcal{F}_r \right] = E \left[ \int_r^t \int G_s \cdot f(y) \nu(dy) ds \middle| \mathcal{F}_r \right], \quad \text{a.s. for } 0 \leq r \leq t \quad (3.7)$$

for any bounded left-continuous adapted process  $G$ , and for any measurable function  $f : \mathbb{R}^d \setminus \{0\} \rightarrow \mathbb{C}$  satisfying  $|f(y)| \leq \text{const.} \cdot (|y| \wedge |y|^2)$ . This can be verified by first considering elementary functions of type  $f(y) = \sum c_i I_{B_i}(y)$  and  $G_s(\omega) = \sum A_i(\omega) I_{(t_i, t_{i+1}]}(s)$  with  $c_i \in \mathbb{R}$ ,  $B_i \in \mathcal{B}(\mathbb{R}^d \setminus \{0\})$ ,  $0 \leq t_0 < t_1 < \dots < t_n$ , and  $A_i$  bounded and  $\mathcal{F}_{t_i}$ -measurable.

Now fix  $p \in \mathbb{R}^d$ , and consider the semimartingale

$$Z_t = \exp(ip \cdot X_t + t\psi(p)) = \exp(ip \cdot M_t + t(\psi(p) + ip \cdot b)).$$

Noting that  $[M^k, M^l]_t^c = [X^k, X^l]_t^c = a^{kl}t$  by (ii), Itô's formula yields

$$\begin{aligned} Z_t &= 1 + \int_0^t Z_- ip \cdot dM + \int_0^t Z_- (\psi(p) + ip \cdot b - \frac{1}{2} \sum_{k,l} p_k p_l a^{kl}) dt \\ &\quad + \sum_{(0,t]} Z_- \left( e^{ip \cdot \Delta X} - 1 - ip \cdot \Delta X \right). \end{aligned} \quad (3.8)$$

By (3.7) and since  $|e^{ip \cdot y} - 1 - ip \cdot y| \leq \text{const.} \cdot (|y| \wedge |y|^2)$ , the series on the right hand side of (3.8) can be decomposed into a martingale and the finite variation process

$$A_t = \int_0^t \int Z_{s-} (e^{ip \cdot y} - 1 - ip \cdot y) \nu(dy) ds$$

Therefore, by (3.8) and (3.6),  $(Z_t)$  is a martingale for any  $p \in \mathbb{R}^d$ . The assertion now follows again by Lemma 3.1.  $\square$

An interesting consequence of Theorem 3.3 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  $B_t : \Omega \rightarrow \mathbb{R}^d$  and  $X_t : \Omega \rightarrow \mathbb{R}^d$  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

$$\psi(p, q) = \frac{1}{2}|p|_{\mathbb{R}^d}^2 + \psi_X(q), \quad p \in \mathbb{R}^d, \quad q \in \mathbb{R}^{d'}.$$

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

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t \quad (3.9)$$

driven by a  $d$ -dimensional Brownian motion  $(B_t)$ . As a consequence, one can show that a process is a weak solution of (3.9) if and only if it solves the corresponding martingale problem. We assume that the coefficients  $b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  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

$$\mathcal{L} = \frac{1}{2} \sum_{i,j=1}^d a^{ij}(t, x) \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^d b^i(t, x) \frac{\partial}{\partial x^i} \quad (3.10)$$

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 3.4 (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 (3.9) on the setup  $(\Omega, \mathcal{A}, P, (\mathcal{F}_t), (B_t))$ .
- (ii) The processes  $M_t^i := X_t^i - \int_0^t b^i(s, X_s) ds$ ,  $1 \leq i \leq d$ , are continuous local  $(\mathcal{F}_t^P)$  martingales with covariations

$$[M^i, M^j]_t = \int_0^t a^{ij}(s, X_s) ds \quad P\text{-a.s. for any } t \geq 0. \quad (3.11)$$

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

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

*Proof.* (i) $\Rightarrow$ (iv) is a consequence of the Itô-Doebelin formula, cf. equation (2.56) above.

(iv) $\Rightarrow$ (iii) trivially holds.

(iii) $\Rightarrow$ (ii) follows by choosing for  $f$  polynomials of degree  $\geq 2$ . Indeed, for  $f(x) = x^i$ , we obtain  $\mathcal{L}f = b^i$ , hence

$$M_t^i = X_t^i - \int_0^t b^i(s, X_s) ds = M_t^{[f]} \quad (3.12)$$

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

$$X_t^i X_t^j = M_t^{[f]} + \int_0^t (a^{ij}(s, X_s) + X_s^i b^j(s, X_s) + X_s^j b^i(s, X_s)) ds \quad (3.13)$$

On the other hand, by the product rule and (3.12),

$$\begin{aligned} X_t^i X_t^j &= X_0^i X_0^j + \int_0^t X_s^i dX_s^j + \int_0^t X_s^j dX_s^i + [X^i, X^j]_t \\ &= N_t + \int_0^t (X_s^i b^j(s, X_s) + X_s^j b^i(s, X_s)) ds + [X^i, X^j]_t \end{aligned} \quad (3.14)$$

with a continuous local martingale  $N$ . Comparing (3.13) and (3.14) we obtain

$$[M^i, M^j]_t = [X^i, X^j]_t = \int_0^t a^{ij}(s, X_s) ds$$

since a continuous local martingale of finite variation is constant.

(ii) $\Rightarrow$ (i) is a consequence of Lévy's characterization of Brownian motion: If (ii) holds then

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

where  $M_t = (M_t^1, \dots, M_t^d)$  and  $B_t := \int_0^t \sigma(s, X_s)^{-1} dM_s$  are continuous local martingales with values in  $\mathbb{R}^d$  because  $\sigma^{-1}$  is locally bounded. To identify  $B$  as a Brownian motion it suffices to note that

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

for any  $k, l = 1, \dots, d$  by (3.11). □

**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 [29] 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 [31] and by Ethier and Kurtz [11] for much more on the martingale problem and its applications to Markov processes.

## 3.2 Random time change

Random time change is already central to the work of Doebelin from 1940 that has been discovered only recently [6]. Independently, Dambis and Dubins-Schwarz have developed a theory of random time changes for semimartingales in the 1960s [19], [28]. 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.,

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

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

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 3.5 (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*

$$M_t = B_{[M]_t} \quad \text{for any } t \geq 0, \quad \text{almost surely.} \quad (3.15)$$

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  $a > 0$ ,

$$\lim_{\varepsilon \downarrow 0} M_{T_{a-\varepsilon}} = M_{T_{a-}} \quad \text{for any } a \geq 0$$

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

$$[M]_{T_{a-}} = \lim_{\varepsilon \downarrow 0} [M]_{T_{a-\varepsilon}} = \lim_{\varepsilon \downarrow 0} (a - \varepsilon) = a = [M]_{T_a}$$

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

$$E[[M^{T_a}]_\infty] = E[[M]_{T_a}] = a < \infty.$$

Hence  $M^{T_a}$  is in  $M_c^2([0, \infty])$ , and

$$E[B_a^2] = E[M_{T_a}^2] = E[(M_\infty^{T_a})^2] = a \quad \text{for any } a \geq 0.$$

This shows that  $(B_a)$  is square-integrable, and, moreover,

$$E[B_a | \mathcal{F}_{T_r}] = E[M_{T_a} | \mathcal{F}_{T_r}] = M_{T_r} = B_r \quad \text{for any } 0 \leq r \leq a$$

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

$$\begin{aligned} E[B_a^2 - B_r^2 | \mathcal{F}_{T_r}] &= E[M_{T_a}^2 - M_{T_r}^2 | \mathcal{F}_{T_r}] \\ &= E[[M]_{T_a} - [M]_{T_r} | \mathcal{F}_{T_r}] = a - r \quad \text{for } 0 \leq r \leq a. \end{aligned}$$

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.  $\square$

**Remark.** The assumption  $[M]_\infty = \infty$  in Theorem 3.5 ensures  $T_a < \infty$  almost surely. If the assumption is violated then  $M$  can still be represented in the form (3.15) 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. [28].

**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 3.5. 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 3.5 and the remark below the theorem, stochastic integrals w.r.t. Brownian motions are time-changed Brownian motions. For any integrand  $G \in \mathcal{L}_{a,loc}^2(\mathbb{R}_+, \mathbb{R}^d)$ , there exists a one-dimensional Brownian motion  $B$ , possibly defined on an enlarged probability space, such that almost surely,

$$\int_0^t G_s \cdot dW_s = B_{\int_0^t |G_s|^2 ds} \quad \text{for any } t \geq 0.$$

**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 [9].

Doebelin [6] has developed a stochastic calculus based on time substitutions instead of Itô integrals. For example, an SDE in  $\mathbb{R}^1$  of type

$$X_t - X_0 = \int_0^t \sigma(s, X_s) dW_s + \int_0^t b(s, X_s) ds$$

can be rephrased in the form

$$X_t - X_0 = B_{\int_0^t \sigma(s, X_s)^2 ds} + \int_0^t b(s, X_s) ds$$

with a Brownian motion  $B$ . The one-dimensional Itô-Doebelin formula then takes the form

$$f(t, X_t) - f(0, X_0) = B_{\int_0^t \sigma(s, X_s)^2 ds} f'(s, X_s) + \int_0^t \left( \frac{\partial f}{\partial s} + \mathcal{L}f \right) (s, X_s) ds$$

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

$$dY_t = \sigma(Y_t) dB_t \tag{3.16}$$

in  $\mathbb{R}^1$  where  $\sigma : \mathbb{R} \rightarrow (0, \infty)$  is a strictly positive continuous function. If  $Y$  is a weak solution then by Theorem 3.5 and the remark below,

$$Y_t = X_{A_t} \quad \text{with} \quad A_t = [Y]_t = \int_0^t \sigma(Y_r)^2 dr \quad (3.17)$$

and a Brownian motion  $X$ . Note that  $A$  depends on  $Y$ , so at first glance (3.17) seems not to be useful for solving the SDE (3.16). However, the inverse time substitution  $T = A^{-1}$  satisfies

$$T' = \frac{1}{A' \circ T} = \frac{1}{\sigma(Y \circ T)^2} = \frac{1}{\sigma(X)^2},$$

and hence

$$T_a = \int_0^a \frac{1}{\sigma(X_u)^2} du.$$

Therefore, we can construct a weak solution  $Y$  of (3.16) 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 3.6.** *Suppose that  $(X_a)$  on  $(\Omega, \mathcal{A}, P, (\mathcal{F}_t))$  is a weak solution of an SDE of the form*

$$dX_a = \sigma(X_a) dB_a + b(X_a) da \quad (3.18)$$

*with locally bounded measurable coefficients  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  such that  $\sigma(x)$  is invertible for almost all  $x$ , and  $\sigma^{-1}$  is again locally bounded. Let  $\varrho : \mathbb{R}^d \rightarrow (0, \infty)$  be a measurable function such that almost surely,*

$$T_a := \int_0^a \varrho(X_u) du < \infty \quad \forall a \in (0, \infty), \quad \text{and} \quad T_\infty = \infty. \quad (3.19)$$

*Then the time-changed process defined by*

$$Y_t := X_{A_t}, \quad A := T^{-1},$$

*is a weak solution of the SDE*

$$dY_t = \left( \frac{\sigma}{\sqrt{\varrho}} \right) (Y_t) dB_t + \left( \frac{b}{\varrho} \right) (Y_t) dt. \quad (3.20)$$

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

*Proof of 3.6. (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.,

$$M_a^{[f]} = f(X_a) - \int_0^a (\mathcal{L}f)(X_u) du$$

is a local  $(\mathcal{F}_a)$  martingale for any  $f \in C^2$ . Therefore, the time-changed process

$$\begin{aligned} M_{A_t}^{[f]} &= f(X_{A_t}) - \int_0^{A_t} (\mathcal{L}f)(X_u) du \\ &= f(Y_t) - \int_0^t (\mathcal{L}f)(Y_r) A'_r dr \end{aligned}$$

is a local  $(\mathcal{F}_{A_t})$  martingale. Noting that

$$A'_r = \frac{1}{T'(A_r)} = \frac{1}{\varrho(X_{A_r})} = \frac{1}{\varrho(Y_r)},$$

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

$$\tilde{\mathcal{L}} = \frac{1}{\varrho} \mathcal{L} = \frac{1}{2} \sum_{i,j} \frac{a_{ij}}{\varrho} \frac{\partial^2}{\partial x^i \partial x^j} + \frac{b}{\varrho} \cdot \nabla.$$

Since  $\frac{a}{\varrho} = \frac{\sigma}{\varrho} \frac{\sigma^T}{\varrho}$ , this implies that  $Y$  is a weak solution of (3.20).  $\square$

In particular, the theorem shows that if  $X$  is a Brownian motion and condition (3.19) 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

$$dY_t = |Y_t|^\alpha dB_t, \quad Y_0 = 0, \quad (3.21)$$

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

$$\begin{aligned} E[T_a] &= E\left[\int_0^a \varrho(X_u) du\right] = \int_0^a E[|X_u|^{-2\alpha}] du \\ &= E[|X_1|^{-2\alpha}] \cdot \int_0^a u^{-\alpha} du < \infty \end{aligned}$$

for any  $a \in (0, \infty)$ . Hence (3.19) holds, and therefore the process  $Y_t = X_{A_t}$ ,  $A = T^{-1}$ , is a non-trivial weak solution of (3.21). 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 (3.19) 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

$$Z_a = Y_{T_a}, \quad T = A^{-1} \text{ with } A_t = \int_0^t |B_s|^{-2} ds,$$

is a diffusion taking values in the unit sphere  $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$  with generator

$$\mathcal{L}f(x) = \frac{1}{2} \left( \Delta f(x) - \sum_{i,j} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right) - \frac{n-1}{2} \sum_i x_i \frac{\partial f}{\partial x_i}(x), \quad x \in S^{n-1}.$$

### One-dimensional SDE

By combining scale and time transformations, one can carry out a rather complete study of weak solutions for non-degenerate SDE of the form

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt, \quad X_0 = x_0, \quad (3.22)$$

on a real interval  $(\alpha, \beta)$ . We assume that the initial value  $X_0$  is contained in  $(\alpha, \beta)$ , and  $b, \sigma : (\alpha, \beta) \rightarrow \mathbb{R}$  are continuous functions such that  $\sigma(x) > 0$  for any  $x \in (\alpha, \beta)$ . We first simplify (3.22) by a coordinate transformation  $Y_t = s(X_t)$  where

$$s : (\alpha, \beta) \rightarrow (s(\alpha), s(\beta))$$

is  $C^2$  and satisfies  $s'(x) > 0$  for all  $x$ . The scale function

$$s(z) := \int_{x_0}^z \exp \left( - \int_{x_0}^y \frac{2b(x)}{\sigma(x)^2} dx \right) dy$$

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

$$dY_t = (\sigma s')(X_t) dB_t,$$

i.e.,  $Y$  is a solution of the equation

$$dY_t = \tilde{\sigma}(Y_t) dB_t, \quad Y_0 = s(x_0), \quad (3.23)$$

where  $\tilde{\sigma} := (\sigma s') \circ s^{-1}$ . The SDE (3.23) 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 3.7.** *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 (3.22) defined up to a stopping time  $\zeta$ .*
- (ii) *The process  $Y_t = s(X_t)$ ,  $t < \zeta$ , on the same setup is a weak solution of (3.23) up to  $\zeta$ .*
- (iii) *The process  $(Y_t)_{s < \zeta}$  has a representation of the form  $Y_t = \tilde{B}_{A_t}$ , where  $\tilde{B}_t$  is a one-dimensional Brownian motion satisfying  $\tilde{B}_0 = s(x_0)$  and  $A = T^{-1}$  with*

$$T_r = \int_0^r \varrho(\tilde{B}_u) du, \quad \varrho(y) = 1/\tilde{\sigma}(y)^2.$$

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 [29]. Here we only note that 3.7 immediately implies existence and uniqueness of a maximal weak solution of (3.22):

**Corollary 3.8.** *Under the regularity and non-degeneracy conditions on  $\sigma$  and  $b$  imposed above there exists a weak solution  $X$  of (3.22) 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  $(\alpha, \beta)$ . 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  $\sigma$  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

$$dR_t = dW_t + \frac{d-1}{2R_t} dt, \quad W \sim \text{BM}(\mathbb{R}^1),$$

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

$$s(r) = \begin{cases} \frac{1}{2-d} r^{2-d} & \text{for } d \neq 2, \\ \log r & \text{for } d = 2 \end{cases}$$

(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

$$(s(0), s(\infty)) = \begin{cases} (0, \infty) & \text{for } d < 2, \\ (-\infty, \infty) & \text{for } d = 2, \\ (-\infty, 0) & \text{for } d > 2. \end{cases}$$

By applying the scale transformation, we see that

$$P[T_b^R < T_a^R] = P[T_{s(b)}^{s(R)} < T_{s(a)}^{s(R)}] = \frac{s(r_0) - s(a)}{s(b) - s(a)}$$

for any  $a < r_0 < b$ , where  $T_c^X$  denoted the first passage time to  $c$  for the process  $X$ . As a consequence,

$$P\left[\liminf_{t \uparrow \zeta} R_t = 0\right] = P\left[\bigcap_{a \in (0, r_0)} \bigcup_{b \in (r_0, \infty)} \{T_a^R < T_b^R\}\right] = \begin{cases} 1 & \text{for } d \leq 2, \\ 0 & \text{for } d > 2, \end{cases}$$

$$P\left[\limsup_{t \uparrow \zeta} R_t = \infty\right] = P\left[\bigcap_{b \in (r_0, \infty)} \bigcup_{a \in (0, r_0)} \{T_b^R < T_a^R\}\right] = \begin{cases} 1 & \text{for } d \geq 2, \\ 0 & \text{for } d < 2. \end{cases}$$

Note that  $d = 2$  is the critical dimension in both cases. Rewriting the SDE in natural scale yields

$$d s(R) = \tilde{\sigma}(s(R)) dW \quad \text{with} \quad \tilde{\sigma}(y) = s'(s^{-1}(y)).$$

In the *critical case*  $d = 2$ ,  $s(r) = \log r$ ,  $\tilde{\sigma}(y) = e^{-y}$ , and hence  $\varrho(y) = \tilde{\sigma}(y)^{-2} = e^{2y}$ . Thus the speed measure is  $m(dy) = e^{2y} dy$ , and  $\log R_t = \tilde{B}_{T^{-1}(t)}$ , i.e.,

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

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

### 3.3 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 a stochastic differential equation of type

$$dX_t = d(t, X_t) dt + dY_t \quad (3.24)$$

corresponds to a random translation of the paths of  $(Y_t)$ :

$$X_t(\omega) = Y_t(\omega) + H_t(\omega) \quad \text{where} \quad H_t = \int_0^t b(X_s) ds.$$

- 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

$$X = b(X) + Y, \quad Y \sim N(0, I_d) \text{ w.r.t. } P, \quad (3.25)$$

for random variables  $X, Y : \Omega \rightarrow \mathbb{R}^d$  where  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a “predictable” map in the sense that the  $i$ -th component  $b^i(x)$  depends only on the first  $i - 1$  components  $X^1, \dots, X^{i-1}$  of  $X$ . The predictability ensures in particular that the transformation defined by (3.25) is invertible, with  $Y^1 = X^1 - b^1$ ,  $Y^2 = X^2 - b^2(X^1)$ ,  $Y^3 = X^3 - b^3(X^1, X^2), \dots, Y^n = X^n - b^n(X^1, \dots, X^{n-1})$ .

A random variable  $(X, P)$  is a “weak” solution of the equation (3.25) 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

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

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

*(X, P) is a weak solution of (3.25) 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*

$$\frac{dP}{dQ} = \exp(X \cdot b(X) - |b(X)|^2/2). \quad (3.26)$$

In particular, we see that the distribution  $\mu_b$  of a weak solution of (3.25) is uniquely determined, and  $\mu_b$  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 3.17 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 (3.24) 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  $\sigma$ -algebra  $\mathcal{F}_{t_0}$  with relative density

$$Z_{t_0} = \left. \frac{dP}{dQ} \right|_{\mathcal{F}_{t_0}} > 0 \quad Q\text{-almost surely.}$$

Then  $P$  and  $Q$  are also mutually absolutely continuous on each of the  $\sigma$ -algebras  $\mathcal{F}_t$ ,  $t \leq t_0$ , with  $Q$ - and  $P$ -almost surely strictly positive relative densities

$$Z_t = \left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = E_Q[Z_{t_0} | \mathcal{F}_t] \quad \text{and} \quad \left. \frac{dQ}{dP} \right|_{\mathcal{F}_t} = \frac{1}{Z_t}.$$

The process  $(Z_t)_{t \leq t_0}$  is a martingale w.r.t.  $Q$ , and, correspondingly,  $(1/Z_t)_{t \leq t_0}$  is a martingale w.r.t.  $P$ . From now on, we always choose a càdlàg version of these martingales.

**Lemma 3.9.** 1) For any  $0 \leq s \leq t \leq t_0$ , and for any  $\mathcal{F}_t$ -measurable random variable  $X : \Omega \rightarrow [0, \infty]$ ,

$$E_P[X | \mathcal{F}_s] = \frac{E_Q[X Z_t | \mathcal{F}_s]}{E_Q[Z_t | \mathcal{F}_s]} = \frac{E_Q[X Z_t | \mathcal{F}_s]}{Z_s} \quad P\text{-a.s. and } Q\text{-a.s.} \quad (3.27)$$

2) Suppose that  $(M_t)_{t \leq t_0}$  is an  $(\mathcal{F}_t)$  adapted càdlàg stochastic process. Then

- (i)  $M$  is a martingale w.r.t.  $P$   $\Leftrightarrow$   $M \cdot Z$  is a martingale w.r.t.  $Q$ ,
- (ii)  $M$  is a local martingale w.r.t.  $P$   $\Leftrightarrow$   $M \cdot Z$  is a local martingale w.r.t.  $Q$ .

*Proof.* 1) The right-hand side of (3.27) is  $\mathcal{F}_s$ -measurable. Moreover, for any  $A \in \mathcal{F}_s$ ,

$$\begin{aligned} E_P[E_Q[XZ_t|\mathcal{F}_s]/Z_s; A] &= E_Q[E_Q[XZ_t|\mathcal{F}_s]; A] \\ &= E_Q[XZ_t; A] = E_Q[X; A]. \end{aligned}$$

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.,

$$E_P[M_{t \wedge T_n}; A] = E_P[M_{s \wedge T_n}; A] \quad \text{for any } A \in \mathcal{F}_s, \quad 0 \leq s \leq t \leq t_0. \quad (3.28)$$

To verify (3.28), we first note that

$$E_P[M_{t \wedge T_n}; A \cap \{T_n \leq s\}] = E_P[M_{s \wedge T_n}; A \cap \{T_n \leq s\}] \quad (3.29)$$

since  $t \wedge T_n = T_n = s \wedge T_n$  on  $\{T_n \leq s\}$ . Moreover, one verifies from the definition of the  $\sigma$ -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,

$$\begin{aligned} E_P[M_{t \wedge T_n}; A \cap \{T_n > s\}] &= E_Q[M_{t \wedge T_n} Z_{t \wedge T_n}; A \cap \{T_n > s\}] \\ &= E_Q[M_{s \wedge T_n} Z_{s \wedge T_n}; A \cap \{T_n > s\}] = E_P[M_{s \wedge T_n}; A \cap \{T_n > s\}] \end{aligned} \quad (3.30)$$

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  $Z_{t \wedge T_n}$ . (3.28) follows from (3.29) and (3.30).  $\square$

If the probability measures  $P$  and  $Q$  are mutually absolutely continuous on the  $\sigma$ -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

$$L_t := \int_0^t \frac{1}{Z_{s-}} dZ_s$$

denote the **stochastic "logarithm"** of  $Z$ . Note that  $(L_t)_{t \leq t_0}$  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,

$$dZ_t = Z_{t-} dL_t, \quad Z_0 = 1,$$

so  $Z_t$  is the stochastic exponential of the local  $Q$ -martingale  $(L_t)$ :

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

### 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

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

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

$$Z_t = \exp\left(\int_0^t G_s \cdot dX_s - \frac{1}{2} \int_0^t |G_s|^2 ds\right) \quad (3.31)$$

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

$$\frac{dP}{dQ} \Big|_{\mathcal{F}_{t_0}} = Z_{t_0} \quad Q\text{-a.s.} \quad (3.32)$$

Note that  $P$  and  $Q$  are mutually absolutely continuous on  $\mathcal{F}_t$  for any  $t \leq t_0$  with

$$\frac{dP}{dQ} \Big|_{\mathcal{F}_t} = Z_t \quad \text{and} \quad \frac{dQ}{dP} \Big|_{\mathcal{F}_t} = \frac{1}{Z_t}$$

both  $P$ - and  $Q$ -almost surely. We are now ready to prove one of the most important results of stochastic analysis:

**Theorem 3.10 (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 (3.31) with  $G \in \mathcal{L}_{a,loc}^2(\mathbb{R}_+, \mathbb{R}^d)$ . If  $E_Q[Z_{t_0}] = 1$  then the process*

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

*is a Brownian motion w.r.t. any probability measure  $P$  on  $\mathcal{A}$  satisfying (3.32).*

*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 3.9, 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  $1 \leq i \leq d$ ,

$$\begin{aligned} d(B^i \cdot Z) &= B^i dZ + Z dB^i + d[B^i, Z] \\ &= B^i ZG \cdot dX + Z dX^i - Z G^i dt + ZG^i dt, \end{aligned} \tag{3.33}$$

where we have used that

$$d[B^i, Z] = ZG \cdot d[B^i, X] = ZG^i dt \quad Q\text{- a.s.}$$

The right-hand side of (3.33) is a stochastic integral w.r.t. the  $Q$ -Brownian motion  $X$ , and hence a local  $Q$ -martingale.  $\square$

The theorem shows that if  $X$  is a Brownian motion w.r.t.  $Q$ , and  $Z$  defined by (3.31) is a  $Q$ -martingale, then  $X$  satisfies

$$dX_t = G_t dt + dB_t.$$

with a  $P$ -Brownian motion  $B$ . It generalizes the Cameron-Martin Theorem to non-deterministic adapted translation

$$X_t(\omega) \longrightarrow X_t(\omega) - H_t(\omega), \quad H_t = \int_0^t G_s ds,$$

of a Brownian motion  $X$ .

**Remark (Assumptions in Girsanov's Theorem).** 1) Absolute continuity and adapt- edness of the "translation process"  $H_t = \int_0^t G_s ds$  are essential for the assertion of Theorem 3.10.

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 3.12 below. If  $(Z_t)$  is not a martin- gale 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  $\sigma$ -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

$$dX_t = b(t, X_t) dt + dB_t, \quad X_0 = o, \quad B \sim \text{BM}(\mathbb{R}^d), \quad (3.34)$$

where  $b : \mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is continuous, and  $o \in \mathbb{R}^d$  is a fixed initial value. Let  $X_t(x) = x_t$  denote the canonical process on

$$\Omega := \{x \in C([0, \infty), \mathbb{R}^d) : x_0 = o\},$$

and let  $\mu^0$  denote Wiener measure on  $(\Omega, \mathcal{F}_\infty^X)$ . Then  $(X, \mu^0)$  is a Brownian motion. By changing measure, we will transform it into a weak solution of (3.34).

**Assumption (A).** The process

$$Z_t = \exp\left(\int_0^t b(s, X_s) \cdot dX_s - \frac{1}{2} \int_0^t |b(s, X_s)|^2 ds\right), \quad t \geq 0,$$

is a martingale w.r.t.  $\mu^0$ .

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  $\mu^b$  on  $\mathcal{F}_\infty^X$  such that  $\mu^b$  is mutually absolutely continuous w.r.t.  $\mu^0$  on each of the  $\sigma$ -algebras  $\mathcal{F}_t^X$ ,  $t \in [0, \infty)$ , with relative densities

$$\frac{d\mu^b}{d\mu^0} \Big|_{\mathcal{F}_t^X} = Z_t \quad \mu^0\text{-a.s.}$$

By Girsanov's Theorem, the process

$$B_t = X_t - \int_0^t b(s, X_s) ds, \quad t \geq 0,$$

is a Brownian motion w.r.t.  $\mu^b$ . Thus we have shown:

**Corollary 3.11.** *Suppose that (A) holds. Then:*

- 1) *The process  $(X, \mu^b)$  is a weak solution of (3.34).*

2) For any  $t \in [0, \infty)$ , the distribution  $\mu^b \circ X^{-1}$  is absolutely continuous w.r.t. Wiener measure  $\mu^0$  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  $(X, \mu^b)$  of the SDE (3.34): If  $\mu_t^b$  denotes the distribution of  $(X_s)_{s \leq t}$  on  $C_0([0, t], \mathbb{R}^d)$  w.r.t.  $\mu^b$  then

$$\mu_t^b(dx) = \exp\left(\int_0^t b(s, x_s) \cdot dx_s - \frac{1}{2} \int_0^t |b(s, x_s)|^2 ds\right) \mu_t^0(dx). \quad (3.35)$$

By combining (3.35) with the **heuristic** path integral representation

$$\text{“ } \mu_t^0(dx) = \frac{1}{\infty} \exp\left(-\frac{1}{2} \int_0^t |x'_s|^2 ds\right) \delta_0(dx_0) \prod_{0 < s \leq t} dx_s \text{ ”}$$

Wiener measure, we obtain the non-rigorous but very intuitive representation

$$\text{“ } \mu_t^b(dx) = \frac{1}{\infty} \exp\left(-\frac{1}{2} \int_0^t |x'_s - b(s, x_s)|^2 ds\right) \delta_0(dx_0) \prod_{0 < s \leq t} dx_s \text{ ”} \quad (3.36)$$

of  $\mu^b$ . Hence intuitively, the “likely” paths w.r.t.  $\mu^b$  should be those for which the *action functional*

$$I(x) = \frac{1}{2} \int_0^t |x'_s - b(s, x_s)|^2 ds$$

takes small values, and the “most likely trajectory” should be the solution of the deterministic ODE

$$x'_s = b(s, x_s)$$

obtained by setting the noise term in the SDE (3.34) equal to zero. Of course, these arguments do not hold rigorously, because  $I(x) = \infty$  for  $\mu_t^0$ - and  $\mu_t^b$ - almost every  $x$ . Nevertheless, they provide an extremely valuable guideline to conclusions that can then be verified rigorously, e.g. via (3.35).

**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

$$\begin{aligned} H_0 : \quad X_t &= B_t, \\ H_1 : \quad dX_t &= b(X_t) dt + dB_t, \end{aligned}$$

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 (3.35) and Itô's formula is given by

$$\begin{aligned} \frac{d\mu_t^b}{d\mu_t^0} &= \exp\left(\int_0^t b(X_s) \cdot dX_s - \frac{1}{2} \int_0^t |b(X_s)|^2 ds\right) \\ &= \exp\left(H(X_0) - H(X_t) + \frac{1}{2} \int_0^t (\Delta H - |\nabla H|^2)(X_s) ds\right) \end{aligned} \quad (3.37)$$

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

$$H(x_0) - H(x_t) + \frac{1}{2} \int_0^t (\Delta H - |\nabla H|^2)(x_s) ds > \log c. \quad (3.38)$$

Note that the integration by parts in (3.37) shows that the estimation procedure is quite stable, because the log likelihood ratio in (3.38) 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

$$Z_t = \exp(L_t - 1/2 [L]_t)$$

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 3.12 (Novikov 1971).** *Let  $t_0 \in \mathbb{R}_+$ . If  $E[\exp(\frac{p}{2}[L]_{t_0})] < \infty$  for some  $p > 1$  then  $(Z_t)_{t \leq t_0}$  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

$$\begin{aligned}
 & E[Z_{t \wedge T_n} ; Z_{t \wedge T_n} \geq c] \\
 &= E\left[\exp\left(L_{t \wedge T_n} - \frac{p}{2}[L]_{t \wedge T_n}\right) \exp\left(\frac{p-1}{2}[L]_{t \wedge T_n}\right) ; Z_{t \wedge T_n} \geq c\right] \quad (3.39) \\
 &\leq E\left[\exp\left(pL_{t \wedge T_n} - \frac{p^2}{2}[L]_{t \wedge T_n}\right)\right]^{1/p} \cdot E\left[\exp\left(q \cdot \frac{p-1}{2}[L]_{t \wedge T_n}\right) ; Z_{t \wedge T_n} \geq c\right]^{1/q} \\
 &\leq E\left[\exp\left(\frac{p}{2}[L]_t\right) ; Z_{t \wedge T_n} \geq c\right]^{1/q}
 \end{aligned}$$

for any  $n \in \mathbb{N}$ . Here we have used Hölder's inequality and the fact that  $\exp\left(pL_{t \wedge T_n} - \frac{p^2}{2}[L]_{t \wedge T_n}\right)$  is an exponential supermartingale. If  $\exp\left(\frac{p}{2}[L]_t\right)$  is integrable then the right hand side of (3.39) converges to 0 uniformly in  $n$  as  $c \rightarrow \infty$ , because

$$P[Z_{t \wedge T_n} \geq 0] \leq c^{-1} E[Z_{t \wedge T_n}] \leq c^{-1} \rightarrow 0$$

uniformly in  $n$  as  $c \rightarrow \infty$ . 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.

### 3.4 Drift transformations for Itô diffusions

We now consider an SDE

$$dX_t = b(X_t) dt + dB_t, \quad X_0 = o, \quad B \sim \text{BM}(\mathbb{R}^d) \quad (3.40)$$



with initial condition  $o \in \mathbb{R}^d$  and  $b \in \mathcal{L}_{\text{loc}}^\infty(\mathbb{R}^d, \mathbb{R}^d)$ . 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  $\mu^0$  denote Wiener measure on  $(\Omega, \mathcal{F}_\infty^X)$  where  $\Omega = C_0([0, \infty), \mathbb{R}^d)$  and  $X_t(x) = x_t$  is the canonical process on  $\Omega$ . 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.  $\mu^0$ .

We note that by Novikov's criterion, the assumption always holds if

$$|b(x)| \leq c \cdot (1 + |x|) \quad \text{for some finite constant } c > 0 : \quad (3.41)$$

**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  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  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 (3.41) 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  $\mu^b$  on  $(\Omega, \mathcal{F}_\infty^X)$  such that

$$\mu^b[A] = E^0[Z_t; A] \quad \text{for any } t \geq 0 \text{ and } A \in \mathcal{F}_t^X.$$

Here  $E^0$  denotes expectation w.r.t.  $\mu^0$ . By Girsanov's Theorem, the process  $(X, \mu^b)$  is a weak solution of (3.40). Moreover, we can easily verify that  $(X, \mu^b)$  is a Markov process:

**Theorem 3.13 (Markov property).** *If (A) holds then  $(X, \mu^b)$  is a time-homogeneous Markov process with transition function*

$$p_t^b(o, C) = \mu^b[X_t \in C] = E^0[Z_t; X_t \in C] \quad \forall C \in \mathcal{B}(\mathbb{R}^d).$$

*Proof.* Let  $0 \leq s \leq t$ , and let  $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$  be a non-negative measurable function. Then, by the Markov property for Brownian motion,

$$\begin{aligned} E^b[f(X_t) | \mathcal{F}_s^X] &= E^0[f(X_t) Z_t | \mathcal{F}_s^X] / Z_s \\ &= E^0 \left[ f(X_t) \exp \left( \int_s^t b(X_r) \cdot dX_r - \frac{1}{2} \int_s^t |b(X_r)|^2 dr \right) \middle| \mathcal{F}_s^X \right] \\ &= E_{X_s}^0 [f(X_{t-s}) Z_{t-s}] = (p_{t-s}^b f)(X_s) \end{aligned}$$

$\mu^0$ - and  $\mu^b$ -almost surely where  $E_x^0$  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  $(X^b, \mu)$  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  $\sigma$ -algebra on the Banach space  $\Omega$ . Our goal is to condition the diffusion process  $(X_\mu^b)$  on a given terminal value  $X_1 = y$ ,  $y \in \mathbb{R}^d$ . More precisely, we will construct a **regular version  $y \mapsto \mu_y^b$  of the conditional distribution  $\mu^b[\cdot | X_1 = y]$**  in the following sense:

(i) For any  $y \in \mathbb{R}^d$ ,  $\mu_y^b$  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

$$\mu^b[A] = \int_{\mathbb{R}^d} \mu_y^b[A] p_1^b(o, dy).$$

(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 [9]: Under  $\mu^0$ , the process  $X_t^y := X_t - tX_1 + ty$ ,  $0 \leq t \leq 1$ , is independent of  $X_1$  with terminal value  $y$ , and the law  $\mu_y^0$  of  $(X_t^y)_{t \in [0,1]}$  w.r.t.  $\mu^0$  is a regular version of  $\mu^0[\cdot | X_1 = y]$ . The measure  $\mu_y^0$  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:

$$b(x) = -\nabla H(x) \quad \text{with } H \in C^2(\mathbb{R}^d).$$

Then the exponential martingale  $(Z_t)$  takes the form

$$Z_t = \exp\left(H(X_0) - H(X_t) + \frac{1}{2} \int_0^t (\Delta H - |\nabla H|^2)(X_s) ds\right),$$

cf. (3.37). Note that the expression on the right-hand side is defined  $\mu_y^0$ -almost surely for any  $y$ . Therefore,  $(Z_t)$  can be used for changing the measure w.r.t. the Brownian bridge.

**Theorem 3.14 (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

$$p_1^b(o, y) = p_1^0(o, y) \cdot E_y^0[Z_1].$$

2) A regular version of  $\mu^b[\cdot | X_1 = y]$  is given by

$$\mu_y^b(dx) = \frac{p_1^0(o, y)}{p_1^b(o, y)} \frac{\exp H(o)}{\exp H(y)} \exp\left(\frac{1}{2} \int_0^1 (\Delta H - |\nabla H|^2)(x_s) ds\right) \mu_y^0(dx).$$

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  $\mu_y^b$ :

$$\mu_y^b(dx) \propto \exp\left(\frac{1}{2} \int_0^1 (\Delta H - |\nabla H|^2)(x_s) ds\right) \mu_y^0(dx). \quad (3.42)$$

*Proof of 3.14.* Let  $F : \Omega \rightarrow \mathbb{R}_+$  and  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  be measurable functions. By the disintegration of Wiener measure into pinned Wiener measures,

$$E^b[F \cdot g(X_1)] = E^0[Fg(X_1)Z_1] = \int E_y^0[FZ_1] g(y) p_1^0(o, y) dy.$$

Choosing  $F \equiv 1$ , we obtain

$$\int g(y) p_1^b(o, dy) = \int g(y) E_y^0[Z_1] p_1^0(x, y) dy$$

for any non-negative measurable function  $g$ , which implies 1).

Now, by choosing  $g \equiv 1$ , we obtain

$$E^b[F] = \int E_y^0[FZ_1] p_1^0(o, y) dy = \int \frac{E_y^0[FZ_1]}{E_y^0[Z_1]} p_1^b(o, dy) \quad (3.43)$$

$$= \int E_y^b[F] p_1^b(o, dy) \quad (3.44)$$

by 1). This proves 2), because  $X_1 = y$   $\mu_y^b$ -almost surely, and  $y \mapsto \mu_y^b$  is weakly continuous.  $\square$

**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 (3.42) immediately implies the following reversibility property of the diffusion bridge when  $b$  is a gradient: If  $R : C([0, 1], \mathbb{R}^d) \rightarrow C([0, 1], \mathbb{R}^d)$  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 (3.42) is invariant under time reversal.

### Drift transformations for general SDE

We now consider a drift transformation for an SDE of the form

$$dX_t = \sigma(t, X_t) dW_t + b(t, X_t) dt \quad (3.45)$$

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

$$Z_t = \exp\left(\int_0^t \beta(s, X_s) \cdot dW_s - \frac{1}{2} \int_0^t |\beta(s, X_s)|^2 ds\right)$$

where  $b, \beta : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  are continuous functions.

**Corollary 3.15.** *Suppose that  $(X, Q)$  is a weak solution of (3.45). If  $(Z_t)_{t \leq t_0}$  is a  $Q$ -martingale and  $P \ll Q$  on  $\mathcal{F}_{t_0}$  with relative density  $Z_{t_0}$  then  $((X_t)_{t \leq t_0}, P)$  is a weak solution of*

$$dX_t = \sigma(t, X_t) dB_t + (b + \sigma\beta)(t, X_t) dt, \quad B \sim \text{BM}(\mathbb{R}^d). \quad (3.46)$$

*Proof.* By (3.45), the equation (3.46) holds with

$$B_t = W_t - \int_0^t \beta(s, X_s) ds.$$

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

$$\mathcal{L} = \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2}{\partial x^i \partial x^j} + b \cdot \nabla, \quad a = \sigma \sigma^T,$$

then  $(X, P)$  is a solution of the martingale problem for

$$\tilde{\mathcal{L}} = \mathcal{L} + \beta \cdot \sigma^T \nabla.$$

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 (3.41)) with  $h(0, o) = 1$ :

$$\frac{\partial h}{\partial t} + \mathcal{L}h = 0, \quad h(0, o) = 1. \quad (3.47)$$

Then, by Itô's formula, the process

$$Z_t = h(t, X_t), \quad t \geq 0,$$

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 (3.47),

$$dZ_t = (\sigma^T \nabla h)(t, X_t) \cdot dW_t.$$

Hence  $Z_t = \mathcal{E}_t^L = \exp(L_t - \frac{1}{2}[L]_t)$  where

$$L_t = \int_0^t \frac{1}{Z_s} dZ_s = \int_0^t (\sigma^T \nabla \log h)(s, X_s) \cdot dW_s$$

is the stochastic logarithm of  $Z$ . Thus if  $Z$  is a martingale, and  $P \ll Q$  with  $\frac{dP}{dQ}|_{\mathcal{F}_t} = Z_t$  then  $(X, P)$  solves the SDE (3.45) with

$$\beta = \sigma^T \nabla \log h.$$

**Example.** If  $X_t = W_t$  is a Brownian motion w.r.t.  $Q$  then

$$dX_t = \nabla \log h(t, X_t) dt + dB_t, \quad B \sim \text{BM}(\mathbb{R}^d) \text{ w.r.t. } P.$$

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  $\alpha$ .

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  $(X, \mu^b)$  on  $C_0([0, 1], \mathbb{R}^d)$  satisfying

$$dX_t = dB_t + b(X_t) dt, \quad X_0 = o,$$

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

$$\frac{\partial}{\partial t} p_t^b(\cdot, y) = \mathcal{L}^b p_t^b(\cdot, y) \quad \text{for any } t > 0,$$

where  $\mathcal{L}^b = \frac{1}{2}\Delta + b \cdot \nabla$  is the corresponding generator. Hence

$$h(t, z) = p_{1-t}^b(z, y)/p_1^b(o, y), \quad t < 1,$$

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  $\mu^b, \mu_y^b$  respectively are

$$\begin{aligned} (X_{t_1}, \dots, X_{t_k}) &\sim p_{t_1}^b(o, x_1) p_{t_2-t_1}^b(x_1, x_2) \cdots p_{t_k-t_{k-1}}^b(x_{k-1}, x_k) dx^k && \text{w.r.t. } \mu^b, \\ &\sim \frac{p_{t_1}^b(o, x_1) p_{t_2-t_1}^b(x_1, x_2) \cdots p_{t_k-t_{k-1}}^b(x_{k-1}, x_k) p_{1-t_k}^b(x_k, y)}{p_1^b(o, y)} dx^k && \text{w.r.t. } \mu_y^b. \end{aligned}$$

This shows that  $y \rightarrow \mu_y^b$  coincides with the regular version of the conditional distribution of  $\mu^b$  given  $X_1$ , i.e.,  $\mu_y^b$  is the bridge measure from  $o$  to  $y$ . Hence, by Corollary 3.15, we have shown:

**Theorem 3.16 (SDE for diffusion bridges).** *The diffusion bridge  $(X, \mu_y^b)$  is a weak solution of the SDE*

$$dX_t = dB_t + b(X_t) dt + (\nabla \log p_{1-t}^b(\cdot, y))(X_t) dt, \quad t < 1. \quad (3.48)$$

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  $\sigma$ -algebra  $\mathcal{F}_1$ , the measures  $\mu^b$  and  $\mu_y^b$  are singular.

**Remark (Generalized diffusion bridges).** Theorem 3.16 carries over to bridges of diffusion processes with non-constant diffusion coefficients  $\sigma$ . In this case, the SDE (3.48) is replaced by

$$dX_t = \sigma(X_t) dB_t + b(X_t) dt + (\sigma \sigma^T \nabla \log p_{1-t}(\cdot, y))(X_t) dt. \quad (3.49)$$

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.

### 3.5 Large deviations on path spaces

In this section, we apply Girsanov's Theorem to study random perturbations of a dynamical system of type

$$dX_t^\varepsilon = b(X_t^\varepsilon) dt + \sqrt{\varepsilon} dB_t, \quad X_0^\varepsilon = 0, \quad (3.50)$$

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  $\mu$  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  $\tau_h : \Omega \rightarrow \Omega$ ,

$$\tau_h(\omega) = \omega + h,$$

and the translated Wiener measure  $\mu_h := \mu \circ \tau_h^{-1}$ .

**Theorem 3.17 (Cameron, Martin 1944).** *Let  $h \in \Omega$ . Then  $\mu_h \ll \mu$  if and only if  $h$  is contained in the **Cameron-Martin space***

$$H_{CM} = \left\{ h \in \Omega : h \text{ is absolutely contin. with } h' \in L^2([0, 1], \mathbb{R}^d) \right\}.$$

*In this case, the relative density of  $\mu_h$  w.r.t.  $\mu$  is*

$$\frac{d\mu_h}{d\mu} = \exp\left(\int_0^t h'_s \cdot dW_s - \frac{1}{2} \int_0^t |h'_s|^2 ds\right). \quad (3.51)$$

*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,

$$Z_t = \exp\left(\int_0^t h' \cdot dW - \frac{1}{2} \int_0^t |h'|^2 ds\right)$$



is a martingale w.r.t. Wiener measure  $\mu$ . 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

$$\mu_h = \mu \circ (W + h)^{-1} = \nu \circ W^{-1} = \nu.$$

“ $\Leftarrow$ ” 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.  $\mu$ ,  $W - h$  is a Brownian motion w.r.t.  $\mu_h$ . In particular, it is a semimartingale. Moreover,  $W$  is a semimartingale w.r.t.  $\mu$  and hence also w.r.t.  $\mu_h$ . Thus  $h = W - (W - h)$  is also a semimartingale w.r.t.  $\mu_h$ . 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

$$\int_0^1 g \cdot dh = \int_0^1 g \cdot f ds \quad \text{for any } g \in L^2([0, 1], \mathbb{R}^d).$$

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  $\|g_n\|_{L^2} \rightarrow 0$ . Then by Itô's isometry,  $\int g_n dW \rightarrow 0$  in  $L^2(\mu)$ , and hence  $\mu$ - and  $\mu_h$ -almost surely along a subsequence. Thus also

$$\int g_n \cdot dh = \int g_n \cdot d(W + h) - \int g_n \cdot dW \rightarrow 0$$

$\mu$ -almost surely along a subsequence. Applying the same argument to a subsequence of  $(g_n)$ , we see that every subsequence  $(\tilde{g}_n)$  has a subsequence  $(\hat{g}_n)$  such that  $\int \hat{g}_n \cdot dh \rightarrow 0$ . 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  $L^2([0, 1], \mathbb{R}^d)$ .  $\square$

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 3.18 (Support Theorem).** *For any  $h \in \Omega$  and  $\delta > 0$ ,*

$$\mu[\{\omega \in \Omega : \|\omega - h\| < \delta\}] > 0.$$

*Proof.* Since the Cameron-Martin space is dense in  $\Omega$  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

$$\mu[||W - h|| < \delta] = \mu_{-h}[||W|| < \delta] > 0.$$

as  $\mu[||W|| < \delta] > 0$  and  $\mu_{-h} \ll \mu$ . □

**Remark (Quantitative Support Theorem).** More explicitly,

$$\begin{aligned} \mu[||W - h|| < \delta] &= \mu_{-h}[||W|| < \delta] \\ &= E\left[\exp\left(-\int_0^1 h' \cdot dW - \frac{1}{2} \int_0^1 |h'|^2 ds\right); \max_{s \leq 1} |W_s| < \delta\right] \end{aligned}$$

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 (3.50) for  $b \equiv 0$ , i.e.,

$$X_t^\varepsilon = \sqrt{\varepsilon} B_t, \quad t \in [0, 1],$$

with  $\varepsilon > 0$  and a  $d$ -dimensional Brownian motion  $(B_t)$ . Path integral heuristics suggests that for  $h \in H_{CM}$ ,

$$\text{“ } P[X^\varepsilon \approx h] = \mu\left[W \approx \frac{h}{\sqrt{\varepsilon}}\right] \sim e^{-I(h/\sqrt{\varepsilon})} = e^{-I(h)/\varepsilon} \text{”}$$

where  $I : \Omega \rightarrow [0, \infty]$  is the *action functional* defined by

$$I(\omega) = \begin{cases} \frac{1}{2} \int_0^1 |\omega'(s)|^2 ds & \text{if } \omega \in H_{CM}, \\ +\infty & \text{otherwise.} \end{cases}$$

The heuristics can be turned into a rigorous statement asymptotically as  $\varepsilon \rightarrow 0$  on the exponential scale. This is the content of the next two results that together are known as Schilder's Theorem:

**Theorem 3.19 (Schilder's large derivation principle, lower bound).**

1) For any  $h \in H_{CM}$  and  $\delta > 0$ ,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in B(h, \delta)] \geq -I(h).$$

2) For any open subset  $U \subseteq \Omega$ ,

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in U] \geq -\inf_{\omega \in U} I(\omega).$$

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,

$$\begin{aligned} \mu[\sqrt{\varepsilon}W \in B(h, \delta)] &= \mu[W \in B(h/\sqrt{\varepsilon}, \delta/\sqrt{\varepsilon})] \\ &= \mu_{-h/\sqrt{\varepsilon}}[B(0, \delta/\sqrt{\varepsilon})] \\ &= E\left[\exp\left(-\frac{1}{\sqrt{\varepsilon}} \int_0^1 h' \cdot dW - \frac{1}{2\varepsilon} \int_0^1 |h'|^2 ds\right); B\left(0, \frac{\delta}{\sqrt{\varepsilon}}\right)\right] \\ &\geq \exp\left(-\frac{1}{\varepsilon}I(h) - \frac{c}{\sqrt{\varepsilon}}\right) \mu\left[\left\{\int_0^1 h' \cdot dW \leq c\right\} \cap B\left(0, \frac{\delta}{\sqrt{\varepsilon}}\right)\right] \\ &\geq \frac{1}{2} \exp\left(-\frac{1}{\varepsilon}I(h) - \sqrt{\frac{8I(h)}{\varepsilon}}\right) \end{aligned}$$

where  $E$  stands for expectation w.r.t. Wiener measure. Here we have used that

$$\mu\left[\int_0^1 h' \cdot dW > c\right] \leq c^{-2} E\left[\left(\int_0^1 h' \cdot dW\right)^2\right] = 2I(h)/c^2 \leq 1/4$$

by Itô's isometry and the choice of  $c$ .

2) Let  $U$  be an open subset of  $\Omega$ . For  $h \in U \cap H_{CM}$ , there exists  $\delta > 0$  such that  $B(h, \delta) \subset U$ . Hence by 1),

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in U] \geq -I(h).$$

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

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in U] \geq -\inf_{h \in U \cap H_{CM}} I(h) = -\inf_{\omega \in U} I(\omega).$$

□

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 3.20 (Schilder's large deviations principle, upper bound).**

1) For any  $n \in \mathbb{N}$  and  $\lambda \geq 0$ ,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[I(\sqrt{\varepsilon}W^{(n)}) \geq \lambda] \leq -\lambda.$$

2) For any closed subset  $A \subseteq \Omega$ ,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in A] \leq -\inf_{\omega \in A} I(\omega).$$

*Proof.* 1) Let  $\varepsilon > 0$  and  $n \in \mathbb{N}$ . Then

$$I(\sqrt{\varepsilon}W^{(n)}) = \frac{1}{2}\varepsilon \sum_{k=1}^n n (W_{k/n} - W_{(k-1)/n})^2.$$

Since the random variables  $\eta_k := \sqrt{n} \cdot (W_{k/n} - W_{(k-1)/n})$  are independent and standard normally distributed, we obtain

$$\begin{aligned} \mu[I(\sqrt{\varepsilon}W^{(n)}) \geq \lambda] &= \mu\left[\sum |\eta_k|^2 \geq 2\lambda/\varepsilon\right] \\ &\leq \exp(-2\lambda c/\varepsilon) E\left[\exp\left(c \sum |\eta_k|^2\right)\right], \end{aligned}$$

where the expectation on the right hand side is finite for  $c < 1/2$ . Hence for any  $c < 1/2$ ,

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[I(\sqrt{\varepsilon}W^{(n)}) \geq \lambda] \leq -2c\lambda.$$

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

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[\sqrt{\varepsilon}W \in A] \leq -\lambda. \quad (3.52)$$

By the Theorem of Arzéla-Ascoli, the set  $\{I \leq \lambda\}$  is a *compact* subset of the Banach space  $\Omega$ . Indeed, by the Cauchy-Schwarz inequality,

$$|\omega(t) - \omega(s)| = \left| \int_s^t \omega'(u) du \right| \leq \sqrt{2\lambda} \sqrt{t-s} \quad \forall s, t \in [0, 1]$$

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  $\delta$  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  $\lambda$ . Hence for  $\varepsilon > 0$ , we can estimate

$$\mu[\sqrt{\varepsilon}W \in A] \leq \mu[I(\sqrt{\varepsilon}W^{(n)}) > \lambda] + \mu[\|\sqrt{\varepsilon}W - \sqrt{\varepsilon}W^{(n)}\|_{\text{sup}} > \delta].$$

The assertion (3.52) now follows from

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[I(\sqrt{\varepsilon}W^{(n)}) > \lambda] \leq -\lambda, \quad \text{and} \quad (3.53)$$

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log \mu[\|W - W^{(n)}\|_{\text{sup}} > \delta/\sqrt{\varepsilon}] \leq -\lambda. \quad (3.54)$$

The bound (3.53) holds by 1) for any  $n \in \mathbb{N}$ . The proof of (3.54) 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 (3.54) holds if  $n$  is large enough.  $\square$

**Remark (Large deviation principle for Wiener measure).** Theorems 3.19 and 3.20 show that

$$\mu[\sqrt{\varepsilon}W \in A] \simeq \exp\left(-\frac{1}{\varepsilon} \inf_{\omega \in A} I(\omega)\right)$$

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. [?] or [?]. 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

$$dX_t^\varepsilon = b(X_t^\varepsilon) dt + \sqrt{\varepsilon} dB_t, \quad X_0^\varepsilon = 0. \quad (3.55)$$

This SDE can be solved pathwise:

**Lemma 3.21 (Control map).** *Suppose that  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  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*

$$x(t) = \int_0^t b(x(s)) ds + \omega(t) \quad \forall t \in [0, 1]. \quad (3.56)$$

*The function  $x$  is absolutely continuous if and only if  $\omega$  is absolutely continuous, and in this case,*

$$x'(t) = b(x(t)) + \omega'(t) \quad \text{for a.e. } t \in [0, 1]. \quad (3.57)$$

- 2) *The control map  $\mathcal{J} : C([0, 1], \mathbb{R}^d) \rightarrow C([0, 1], \mathbb{R}^d)$  that maps  $\omega$  to the solution  $\mathcal{J}(\omega) = x$  of (3.56) is continuous.*

*Proof.* 1) Existence and uniqueness holds by the classical Picard-Lindelöf Theorem.

2) Suppose that  $x = \mathcal{J}(\omega)$  and  $\tilde{x} = \mathcal{J}(\tilde{\omega})$  are solutions of (3.56) w.r.t. driving paths  $\omega, \tilde{\omega} \in C[0, 1], \mathbb{R}^d$ . Then for  $t \in [0, 1]$ ,

$$\begin{aligned} |x(t) - \tilde{x}(t)| &= \left| \int_0^t (b(\omega(s)) - b(\tilde{\omega}(s))) ds + \sqrt{\varepsilon}(\omega(t) - \tilde{\omega}(t)) \right| \\ &\leq L \int_0^t |\omega(s) - \tilde{\omega}(s)| ds + \sqrt{\varepsilon}|\omega(t) - \tilde{\omega}(t)|. \end{aligned}$$

where  $L \in \mathbb{R}_+$  is a Lipschitz constant for  $b$ . Gronwall's Lemma now implies

$$|x(t) - \tilde{x}(t)| \leq \exp(tL) \sqrt{\varepsilon} \|\omega - \tilde{\omega}\|_{\text{sup}} \quad \forall t \in [0, 1],$$

and hence

$$\|x - \tilde{x}\|_{\text{sup}} \leq \exp(L) \sqrt{\varepsilon} \|\omega - \tilde{\omega}\|_{\text{sup}}.$$

This shows that the control map  $\mathcal{J}$  is even Lipschitz continuous. □

For  $\varepsilon > 0$ , the unique solution of the SDE (3.55) on  $[0, 1]$  is given by

$$X^\varepsilon = \mathcal{J}(\sqrt{\varepsilon}B).$$

Since the control map  $\mathcal{J}$  is continuous, we can apply Schilder's Theorem to study the large deviations of  $X^\varepsilon$  as  $\varepsilon \downarrow 0$ :

**Theorem 3.22 (Fredlin & Wentzel 1970, 1984).** *If  $b$  is Lipschitz continuous then the large deviations principle*

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in U] \geq - \inf_{x \in U} I_b(x) \quad \text{for any open set } U \subseteq \Omega,$$

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in A] \geq - \inf_{x \in A} I_b(x) \quad \text{for any closed set } A \subseteq \Omega,$$

holds, where the rate function  $I_b : \Omega \rightarrow [0, \infty]$  is given by

$$I_b(x) = \begin{cases} \frac{1}{2} \int_0^1 |x'(s) - b(x(s))|^2 ds & \text{for } x \in H_{CM}, \\ +\infty & \text{for } x \in \Omega \setminus H_{CM}. \end{cases}$$

*Proof.* For any set  $A \subseteq \Omega$ , we have

$$P[X^\varepsilon \in A] = P[\sqrt{\varepsilon}B \in \mathcal{J}^{-1}(A)] = \mu[\sqrt{\varepsilon}W \in \mathcal{J}^{-1}(A)].$$

If  $A$  is open then  $\mathcal{J}^{-1}(A)$  is open by continuity of  $\mathcal{J}$ , and hence

$$\liminf_{\varepsilon \downarrow 0} \varepsilon \log P[X^\varepsilon \in A] \geq - \inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega)$$

by Theorem 3.19. Similarly, if  $A$  is closed then  $\mathcal{J}^{-1}(A)$  is closed, and hence the corresponding upper bound holds by Theorem 3.20. Thus it only remains to show that

$$\inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega) = \inf_{x \in A} I_b(x).$$

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,

$$\begin{aligned} \inf_{\omega \in \mathcal{J}^{-1}(A)} I(\omega) &= \inf_{\omega \in \mathcal{J}^{-1}(A) \cap H_{CM}} \frac{1}{2} \int_0^1 |\omega'(s)|^2 ds \\ &= \inf_{x \in A \cap H_{CM}} \frac{1}{2} \int_0^1 |x'(s) - b(x(s))|^2 ds = \inf_{x \in A} I_b(x). \end{aligned}$$

□

**Remark.** The large deviation principle in Theorem 3.22 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. [?].

### 3.6 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  $\sigma$ -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  $\mathcal{L}_{\text{loc}}^2(P \otimes \lambda \otimes \nu)$ . 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  $\omega$  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

$$Z_t := \mathcal{E}_t^L \quad \text{where} \quad L_t := (H_{\bullet} \tilde{N})_t. \quad (3.58)$$

**Lemma 3.23.** *Suppose that  $\inf H > -1$ , and let  $G := \log(1 + H)$ . Then for  $t \geq 0$ ,*

$$\mathcal{E}_t^L = \exp \left( \int_{(0,t] \times S} G_s(y) \tilde{N}(ds dy) - \int_{(0,t] \times S} (H_s(y) - G_s(y)) ds \nu(dy) \right).$$



*Proof.* The assumption  $\inf H > -1$  implies  $\inf \Delta L > -1$ . Since, moreover,  $[L]^c = 0$ , we obtain

$$\begin{aligned}\mathcal{E}^L &= e^{L-[L]^c/2} \prod (1 + \Delta L) e^{-\Delta L} \\ &= \exp \left( L + \sum (\log(1 + \Delta L) - \Delta L) \right) \\ &= \exp \left( G_{\bullet} \tilde{N} + \int (G - H) ds \nu(dy) \right).\end{aligned}$$

Here we have used that

$$\sum (\log(1 + \Delta L) - \Delta L) = \int (\log(1 + H_s(y)) - H_s(y)) N(ds dy)$$

holds, since  $|\log(1 + H_s(y)) - H_s(y)| \leq \text{const} \cdot |H_s(y)|^2$  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 there is a probability measure  $P$  on  $\mathcal{F}_{t_0}$  such that

$$\left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = Z_t \quad \text{for any } t \leq t_0.$$

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

$$\mu(dy) = (1 + h(y)) \nu(dy) :$$

**Theorem 3.24 (Change of measure for Poisson point processes).** *Let  $(N_t, Q)$  be a Poisson point process with intensity measure  $\nu$ , and let  $g := \log(1+h)$  where  $h \in \mathcal{L}^2(\nu)$  satisfies  $\inf h > -1$ . Suppose that the exponential*

$$Z_t = \mathcal{E}_t^{\tilde{N}^h} = \exp \left( \tilde{N}_t^g + t \int (g - h) d\nu \right)$$

*is a martingale w.r.t.  $Q$ , and assume that  $P \ll Q$  on  $\mathcal{F}_t$  with relative density  $\left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = Z_t$  for any  $t \geq 0$ . Then the process  $(N_t, P)$  is a Poisson point process with intensity measure*

$$d\mu = (1 + h) d\nu.$$

*Proof.* By the Lévy characterization for Poisson point processes (cf. the exercise below Lemma 3.1) it suffices to show that the process

$$M_t^f = \exp(iN_t^f + t\psi(f)), \quad \psi(f) = \int (1 - e^{if}) d\mu,$$

is a local martingale w.r.t.  $P$  for any elementary functions  $f \in \mathcal{L}^1(S, \mathcal{S}, \nu)$ . Furthermore, by Lemma 3.9,  $M^f$  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.  $\square$

**Remark (Extension to general measure transformations).** The approach in Theorem 3.24 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 (3.58), 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  $\nu$  is a  $\sigma$ -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  $\nu$ , then

$$X_t = \int y \tilde{N}_t(dy), \quad \tilde{N}_t = N_t - t\nu,$$

is a Lévy martingale with Lévy measure  $\mu$  w.r.t.  $Q$ .

**Corollary 3.25.** *Suppose that  $h \in \mathcal{L}^2(\nu)$  satisfies  $\inf h > -1$  and  $\sup h < \infty$ . Then the process*

$$\bar{X}_t = \int y \bar{N}_t dy + t \int y h(y) \nu(dy), \quad \bar{N}_t = N_t - t\mu,$$

*is a Lévy martingale with Lévy measure  $\mu$  w.r.t.  $P$  provided  $P \ll Q$  on  $\mathcal{F}_t$  with relative density  $Z_t$  for any  $t \leq 0$ .*

**Example.** Suppose that  $(X, Q)$  is a compound Poisson process with finite jump intensity measure  $\nu$ , and let

$$N_t^h = \sum_{s \leq t} h(\Delta X_s).$$

with  $h$  as above. Then  $(X, P)$  is a compound Poisson process with jump intensity measure  $d\mu = (1 + h) d\nu$  provided

$$\left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = \mathcal{E}_t^{\tilde{N}^h} = e^{-\int h d\nu} \prod_{s \leq t} (1 + h(\Delta X_s)).$$

## A general theorem

We finally state a general change of measure theorem for possibly discontinuous semimartingales:

**Theorem 3.26 (P.A. Meyer).** *Suppose that the probability measures  $P$  and  $Q$  are equivalent on  $\mathcal{F}_t$  for any  $t \geq 0$  with relative density  $\left. \frac{dP}{dQ} \right|_{\mathcal{F}_t} = Z_t$ . 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 3.9, 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,

$$\frac{1}{Z}[Z, M] = \int [Z, M]_- d\frac{1}{Z} + \int \frac{1}{Z_-} d[Z, M] + \left[\frac{1}{Z}, [Z, M]\right].$$

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

$$\left[\frac{1}{Z}, [Z, M]\right] = \sum \Delta \frac{1}{Z} \Delta [Z, M].$$

□

**Remark.** Note that the process  $\int \frac{1}{Z} d[Z, M]$  is not predictable in general. For a predictable counterpart to Theorem 3.26 cf. e.g. [27].

# Chapter 4

## Strong solutions and flows

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

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

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

$$c_p \cdot E[[M]_\infty^{p/2}] \leq E[(M_\infty^*)^p] \leq C_p \cdot E[[M]_\infty^{p/2}] \quad (4.1)$$

hold for any continuous local martingale  $M$  satisfying  $M_0 = 0$ , cf. [28]. The inequality shows in particular that for continuous martingales, the  $\mathcal{H}^p$  norm, i.e., the  $L^p$  norm of  $M_\infty^*$ , is equivalent to  $E[[M]_\infty^{p/2}]^{1/p}$ . Note that for  $p = 2$ , by Itô's isometry, equality holds in (4.1) with  $c_p = C_p = 1$ . 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 4.3 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 4.1 (Burkholder's inequality).** *Let  $p \in [2, \infty)$ . Then the estimate*

$$E[(M_T^*)^p]^{1/p} \leq \gamma_p E[[M]_T^{p/2}]^{1/p} \quad (4.2)$$

holds for any  $M \in \mathcal{M}_{loc}$  such that  $M_0 = 0$ , and for any stopping time  $T : \Omega \rightarrow [0, \infty]$ , where

$$\gamma_p = \left(1 + \frac{1}{p-1}\right)^{\frac{p-1}{2}} \frac{p}{\sqrt{2}} \leq \sqrt{\frac{e}{2}} p.$$

**Remark.** The estimate does not depend on the underlying filtered probability space, the local martingale  $M$ , and the stopping time  $T$ . However, the constant  $\gamma_p$  goes to  $\infty$  as  $p \rightarrow \infty$ .

*Proof.* 1) We first assume that  $T = \infty$  and  $M$  is a bounded càdlàg martingale. Then, by the Martingale Convergence Theorem,  $M_\infty = \lim_{t \rightarrow \infty} M_t$  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

$$\begin{aligned} |M_\infty|^p &= \int_0^\infty \varphi'(M_{s-}) dM_s + \frac{1}{2} \int_0^\infty \varphi''(M_s) d[M]_s^c \\ &\quad + \sum_s (\varphi(M_s) - \varphi(M_{s-}) - \varphi'(M_{s-})\Delta M_s), \end{aligned} \quad (4.3)$$

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

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

and the summand in the third term can be estimated by

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

Hence by taking expectation values on both sides of (4.3), we obtain for  $q$  satisfying  $\frac{1}{p} + \frac{1}{q} = 1$ :

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

by Doob's inequality, Hölder's inequality, and since  $[M]_\infty^c + \sum(\Delta M)^2 = [M]_\infty$ . The inequality (4.2) 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

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

Burkholder's inequality for  $M$  now follows as  $n \rightarrow \infty$ .

3) Finally, the inequality for an arbitrary stopping time  $T$  can be derived from that for  $T = \infty$  by considering the stopped process  $M^T$ .  $\square$

For  $p \geq 4$ , the converse estimate in (4.1) 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

$$c_p^{-1} E [[M]_\infty^{p/2}] \leq E [(M_\infty^*)^p] \leq c_p E [[M]_\infty^{p/2}]$$

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

$$P \left[ \sup_{s \leq t} M_s \geq x ; [M]_t \leq c \right] \leq \exp \left( - \frac{x^2}{2c} \right)$$

for any  $c, t, x \in [0, \infty)$ .

## 4.1 Existence and uniqueness of strong solutions

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $(S, \mathcal{S}, \nu)$  a  $\sigma$ -finite measure space, and let  $d, n \in \mathbb{N}$ . Suppose that on  $(\Omega, \mathcal{A}, P)$ , we are given an  $\mathbb{R}^d$ -valued Brownian motion  $(B_t)_{t \geq 0}$  and a Poisson random measure  $N(dt dy)$  on  $\mathbb{R}_+ \times S$  with intensity measure  $\lambda_{(0, \infty)} \otimes \nu$ , and

let  $\tilde{N} := N - \lambda_{(0,\infty)} \otimes \nu$ . We study existence and uniqueness for strong solutions of stochastic differential equations of type

$$dX_t = b_t(X) dt + \sigma_t(X) dB_t + \int c_{t-}(X, y) \tilde{N}(dt dy). \quad (4.4)$$

Here  $b : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \rightarrow \mathbb{R}^n$  and  $\sigma : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \rightarrow \mathbb{R}^{n \times d}$  are functions that are **progressively measurable**, i.e., for any  $t \geq 0$ , the restrictions to  $[0, t] \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n)$  are measurable w.r.t.  $\mathcal{B}([0, t]) \otimes \mathcal{B}_t$  where  $\mathcal{B}_t := \sigma(x \mapsto x_s : s \leq t)$ . Moreover,  $c : \mathbb{R}_+ \times \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n) \times S \rightarrow \mathbb{R}^n$ ,  $(t, x, y) \mapsto c_t(x, y)$ , is a càdlàg function in the first variable for any given values of  $x$  and  $y$ , and  $c_t$  is measurable w.r.t.  $\mathcal{B}_t \otimes \mathcal{S}$  for any  $t \geq 0$ . Note that the assumptions imply that  $b_t(x)$  is a measurable function of the path  $(x_s)_{s \leq t}$  up to time  $t$ . Hence  $b_t(x)$  is also well-defined for càdlàg paths  $(x_s)_{s < \zeta}$  with finite lifetime  $\zeta$  provided  $\zeta > t$ . Corresponding statements hold for  $\sigma_t$  and  $c_t$ . In addition to the assumptions above, we assume a **local Lipschitz condition** for the coefficients:

**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  $\text{Lip}(t_0, U)$  condition holds:

$$|b_t(x) - b_t(\tilde{x})| + \|\sigma_t(x) - \sigma_t(\tilde{x})\| + \int |c_t(x, y) - c_t(\tilde{x}, y)|^2 \nu(dy) \leq L \cdot \sup_{s \leq t} |x_s - \tilde{x}_s|$$

for any  $t \in [0, t_0]$  and  $x, \tilde{x} \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}^n)$  with  $x_s, \tilde{x}_s \in U$  for  $s \leq t_0$ .

In the sequel, we denote by  $x_t^*$  the supremum norm of a càdlàg function  $x$  on the interval  $[0, t]$ :

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

By  $(\mathcal{F}_t)$  we denote the completed filtration generated by the Brownian motion  $(B_t)$  and the Poisson point process  $(N_t)$  corresponding to  $N$ .

## **$L^p$ Stability**

We now prove an a priori estimate for solutions of (4.4) that is crucial for studying existence, uniqueness, and dependence on the initial condition.



**Theorem 4.2 (A priori estimate).** Fix  $p \in [2, \infty)$  and an open set  $U \subset \mathbb{R}^n$ , and let  $T$  be an  $(\mathcal{F}_t)$  stopping time. Suppose that  $(X_t)_{t < T}$  and  $(\tilde{X}_t)_{t < T}$  are solutions of (4.4) taking values in  $U$  for  $t < T$ , and let

$$\varphi(t) := E[(X - \tilde{X})_{t \wedge T}^*{}^p].$$

If the Lipschitz condition  $\text{Lip}(t_0, U)$  holds then there exists a finite constant  $C \in \mathbb{R}_+$  such that for any  $t \leq t_0$ ,

$$\varphi(t) \leq C \cdot \left( \varphi(0) + \int_0^t \varphi(s) ds \right), \quad \text{and} \quad (4.5)$$

$$\varphi(t) \leq C \cdot e^{Ct} \varphi(0). \quad (4.6)$$

*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, (4.6) follows from (4.5) by Gronwell's lemma. To prove (4.5), note that

$$X_t = X_0 + \int_0^t b_s(X) ds + \int_0^t \sigma_s(X) dB_s + \int_{(0,t] \times S} s_{s-}(X, y) \tilde{N}(ds dy) \quad \forall t < T,$$

and an analogue equation holds for  $\tilde{X}$ . Hence for  $t \leq t_0$ ,

$$(X - \tilde{X})_{\bullet}^* \leq \text{I} + \text{II} + \text{III} + \text{IV}, \quad \text{where} \quad (4.7)$$

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

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

$$E[I^2] = \varphi(0), \quad \text{and}$$

$$E[\text{II}^2] \leq L^2 t E \left[ \int_0^{t \wedge T} (X - \tilde{X})_s^*{}^2 ds \right] \leq L^2 t \int_0^t \varphi(s) ds.$$

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

$$\begin{aligned} E[\text{III}^2] &= E[M_{t \wedge T}^*{}^2] \leq 4E[M_{t \wedge T}^2] \\ &= 4E \left[ \int_0^{t \wedge T} \|\sigma_s(X) - \sigma_s(\tilde{X})\|^2 ds \right] \leq 4L^2 \int_0^t \varphi(s) ds, \end{aligned}$$

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

The assertion now follows since by (4.7),

$$\varphi(t) = E[(X - \tilde{X})_{t \wedge T}^*{}^2] \leq 4 \cdot E[\text{I}^2 + \text{II}^2 + \text{III}^2 + \text{IV}^2].$$

□

As an immediate consequence of the a priori estimate, we note the following stability property for strong solutions:

**Corollary 4.3 (Dependence of solutions on the initial condition).** *Suppose that a global Lipschitz condition  $\text{Lip}(t_0, \mathbb{R}^n)$  holds for  $t_0 \in \mathbb{R}_+$ , and let  $(X_t)_{t \leq t_0}$  and  $(\tilde{X})_{t \leq t_0}$  be strong solutions of (4.1) with  $X_0 = \tilde{X}_0$  almost surely. Then*

$$P[X_t = \tilde{X}_t \text{ for any } t \leq 0] = 1.$$

*Proof.* For any open bounded set  $U \subset \mathbb{R}^n$  and  $t_0 \in \mathbb{R}_+$ , the a priori estimate 4.2 implies that  $X$  and  $\tilde{X}$  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  $t_0 \in \mathbb{R}_+$ ,

$$\sup_{t < t_0} \left( |b_t(0)| + \|\sigma_t(0)\| + \int |c_t(0, y)|^2 \nu(dy) \right) < \infty.$$

Here 0 denotes the constant path  $X_0 \equiv 0$  in  $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^n)$ .

**Theorem 4.4 (Itô).** Let  $\xi : \Omega \rightarrow \mathbb{R}^n$  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 (4.1) has a strong solution  $(X_t)_{t < \zeta}$  with initial condition  $X_0 = \xi$  that is defined up to the explosion time

$$\zeta = \sup T_n, \quad \text{where } T_n = \inf \{t \geq 0 : |X_t| \geq n\}.$$

- 2) If, moreover, the global Lipschitz condition  $\text{Lip}(t_0, \mathbb{R}^n)$  holds for any  $t_0 \in \mathbb{R}_+$ , then  $\zeta = \infty$  almost surely.

*Proof of 4.4.* We first prove 2), and then we show that 2) implies 1).

2) To prove the second assertion we may assume w.l.o.g. that  $\xi$  is bounded. We then construct a sequence  $(X^n)$  of approximate solutions of (4.1) by a Picard-Lindelöf iteration, i.e., for  $t \geq 0$  and  $n \in \mathbb{N}$  we define inductively

$$\begin{aligned} X_t^1 &:= \xi, \\ X_t^{n+1} &:= \xi + \int_0^t b_s(X^n) ds + \int_0^t \sigma_s(X^n) dB_s + \int_{(0,t] \times S} c_{s-}(X^n, y) \tilde{N}(ds dy). \end{aligned} \quad (4.8)$$

Let

$$\Delta_t^n := E[(X^{n+1} - X^n)_t^*{}^2].$$

Then by (4.1), the global Lipschitz condition and Itô's isometry, for any  $t_0 \in \mathbb{R}_+$  there exists a finite constant  $C(t_0)$  such that

$$\Delta_t^{n+1} \leq C(t_0) \int_0^t \Delta_s^n ds \quad \text{for any } n \geq 0 \text{ and } t \leq t_0.$$

Hence by induction,

$$\Delta_t^n \leq C(t)^n \frac{t^n}{n!} \Delta_t^0 \quad \forall n \in \mathbb{N}, t \geq 0.$$

In particular,  $\sum_{n=1}^{\infty} \Delta_t^n < \infty$ . An application of the Borel-Cantelli Lemma now shows that for  $t \geq 0$ , the limit  $X_s = \lim_{n \rightarrow \infty} X_s^n$  exists uniformly for  $s \in [0, t]$  with probability one. Moreover,  $X$  is a fixed point of the Picard-Lindelöf iteration, and hence a solution of the SDE (4.1).

1) If  $b, \sigma$  and  $c$  only satisfy the local Lipschitz condition (A1) then for  $k \in \mathbb{N}$  and  $t_0 \in \mathbb{R}_+$ , we find functions  $b^k, \sigma^k$  and  $c^k$  that are globally Lipschitz continuous and that agree with  $b, \sigma$  and  $c$  on paths  $(x_t)$  taking values in the ball  $B(0, k)$  for  $t \leq t_0$ . The solution of the SDE with coefficients  $b^k, \sigma^k, c^k$  is then a solution of (4.1) up to  $t \wedge T_k$  where  $T_k$  denotes the first exit time from  $B(0, k)$ . By strong uniqueness, the local solutions obtained in this way are consistent. Hence they can be combined to construct a solution of (4.1) that is defined up to the explosion time  $\zeta = \sup T_n$ .  $\square$

### Non-explosion criteria

Theorem 4.4 shows that under a global Lipschitz and linear growth condition on the coefficients, the solution of (4.1) 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  $\zeta$  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

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t \quad (4.9)$$

where  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  are locally Lipschitz continuous, and let

$$\mathcal{L} = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + b(x) \cdot \nabla, \quad a(x) = \sigma(x)\sigma(x)^T,$$

denote the corresponding generator.

#### Exercise (Lyapunov condition for non-explosion).

- 1) 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)  $\varphi(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$ , and

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

Prove that in this case, the strong solution of (4.1) exists up to  $\zeta = \infty$  almost surely.

(Hint: Show that  $\exp(-\lambda t)\varphi(X_t)$  is a local supermartingale up to the explosion time  $\zeta$ , and apply the Optional Stopping Theorem.)

2) Conclude that if there exists  $\lambda \in \mathbb{R}_+$ , such that

$$2x \cdot b(x) + \text{tr}(a(x)) \leq \lambda \cdot (1 + |x|^2)$$

then  $\zeta = \infty$  almost surely.

Note that the condition in the exercise is satisfied if

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

for sufficiently large  $x \in \mathbb{R}^n$ , i.e., if the outward component of the drift is growing at most linearly, and the trace of the diffusion matrix is growing at most quadratically.

## 4.2 Stratonovich differential equations

Replacing Itô by Stratonovich 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. [27] 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

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

Note that a Stratonovich integral w.r.t. a martingale is not a local martingale in general. The Stratonovich integral is a limit of trapezoidal Riemann sum approximations:

**Lemma 4.5.** *If  $(\pi_n)$  is a sequence of partitions of  $\mathbb{R}_+$  with  $\text{mesh}(\pi_n) \rightarrow 0$  then*

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

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

### Itô-Stratonovich formula

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

**Theorem 4.6.** *Let  $X = (X^1, \dots, X^d)$  with continuous semimartingales  $X^i$ . Then for any function  $F \in C^2(\mathbb{R}^d)$ ,*

$$F(X_t) - F(X_0) = \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x^i}(X_s) \circ dX_s^i \quad \forall t \geq 0. \quad (4.10)$$

*Proof.* To simplify the proof we assume  $F \in C^3$ . Under this condition, (4.10) is just a reformulation of the Itô rule

$$F(X_t) - F(X_0) = \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x^i}(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s) d[X^i, X^j]_s. \quad (4.11)$$

Indeed, applying Itô's rule to the  $C^2$  function  $\frac{\partial F}{\partial x^i}$  shows that

$$\frac{\partial F}{\partial x^i}(X_t) = A_t + \sum_j \int \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s) dX_s^j$$

for some continuous finite variation process  $A$ . Hence the difference between the Stratonovich integral in (4.10) and the Itô integral in (4.11) is

$$\frac{1}{2} \left[ \frac{\partial F}{\partial x^i}(X), X^i \right]_t = \frac{1}{2} \sum_j \int \frac{\partial^2 F}{\partial x^i \partial x^j}(X_s) d[X^j, X^i]_s.$$

$\square$

**Remark.** For the extension of the proof to  $C^2$  functions  $F$  see e.g. [27], where also a generalization to càdlàg semimartingales is considered.

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

**Corollary 4.7.** *For continuous semimartingales  $X, Y$ ,*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s \circ dY_s + \int_0^t Y_s \circ dX_s \quad \forall t \geq 0.$$

**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

$$\circ dX_t = b(X_t) dt + \sum_{k=1}^d \sigma_k(X_t) \circ dB_t^k, \quad X_0 = x_0, \quad (4.12)$$

with  $x_0 \in \mathbb{R}^n$ , continuous vector fields  $b, \sigma_1, \dots, \sigma_d \in C(\mathbb{R}^n, \mathbb{R}^n)$ , and an  $\mathbb{R}^d$ -valued Brownian motion  $(B_t)$ .

**Exercise (Stratonovich to Itô conversion).** 1) Prove that for  $\sigma_1, \dots, \sigma_d \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ , the Stratonovich SDE (4.12) is equivalent to the Itô SDE

$$dX_t = \tilde{b}(X_t) dt + \sum_{k=1}^d \sigma_k(X_t) dB_t^k, \quad X_0 = x_0, \quad (4.13)$$

where

$$\tilde{b} := b + \frac{1}{2} \sum_{k=1}^d \sigma_k \cdot \nabla \sigma_k.$$

2) Conclude that if  $\tilde{b}$  and  $\sigma_1, \dots, \sigma_d$  are Lipschitz continuous, then there is a unique strong solution of (4.12).

**Theorem 4.8 (Martingale problem for Stratonovich SDE).** *Let  $b \in C(\mathbb{R}^n, \mathbb{R}^n)$  and  $\sigma_1, \dots, \sigma_d \in C^2(\mathbb{R}^n, \mathbb{R}^n)$ , and suppose that  $(X_t)_{t \geq 0}$  is a solution of (4.12) 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*

$$\begin{aligned} M_t^F &= F(X_t) - \int_0^t (\mathcal{L}F)(X_s) ds, \\ \mathcal{L}F &= \frac{1}{2} \sum_{k=1}^d \sigma_k \cdot \nabla (\sigma_k \cdot \nabla F) + b \cdot \nabla F, \end{aligned}$$

is a local  $(\mathcal{F}_t^P)$  martingale.

*Proof.* By the Stratonovich chain rule and by (4.12),

$$\begin{aligned} F(X_t) - F(X_0) &= \int_0^t \nabla F(X) \cdot \circ dX \\ &= \int_0^t (b \cdot \nabla F)(X) dt + \sum_k \int_0^t (\sigma_k \cdot \nabla F)(X) \circ dB^k. \end{aligned} \quad (4.14)$$

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

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

with a continuous finite variation process  $(A_t)$ . Hence

$$\begin{aligned} \int_0^t (\sigma_k \cdot \nabla F)(X) \circ dB^k &= \int_0^t (\sigma_k \cdot \nabla F)(X) dB^k + [(\sigma_k \cdot \nabla F)(X), B^k]_t \\ &= \text{local martingale} + \int_0^t \sigma_k \cdot \nabla (\sigma_k \cdot \nabla F)(X) dt. \end{aligned} \quad (4.15)$$

The assertion now follows by (4.14) and (4.15).  $\square$

The theorem shows that the generator of a diffusion process solving a Stratonovich SDE is in sum of squares form. In geometric notation, one briefly writes  $B$  for the derivative  $b \cdot \nabla$  in the direction of the vector field  $b$ . The generator then takes the form

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



### Brownian motion on hypersurfaces

One important application of Stratonovich calculus is stochastic differential geometry. Itô calculus can not be used directly for studying stochastic differential equations on manifolds, because the classical chain rule is essential for ensuring that solutions stay on the manifold if the driving vector fields are tangent vectors. Instead, one considers Stratonovich equations. These are converted to Itô form when computing expectation values. To avoid differential geometric terminology, we only consider Brownian motion on a hypersurface in  $\mathbb{R}^{n+1}$ , cf. [29], [14] and [16] 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

$$M_c = f^{-1}(c) = \{x \in \mathbb{R}^{n+1} : f(x) = c\}$$

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

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

is the **unit normal** to  $M_c$  at  $x$ . The **tangent space** to  $M_c$  at  $x$  is the orthogonal complement

$$T_x M_c = \text{span} \{\mathbf{n}(x)\}^\perp.$$

Let  $P(x) : \mathbb{R}^{n+1} \rightarrow T_x M_c$  denote the orthogonal projection onto the tangent space w.r.t. the Euclidean metric, i.e.,

$$P(x)v = v - v \cdot \mathbf{n}(x) \mathbf{n}(x), \quad v \in \mathbb{R}^{n+1}.$$

(insert graphic)

For  $k \in \{1, \dots, n+1\}$ , we set  $P_k(x) = P(x)e_k$ .

**Definition.** A **Brownian motion on the hypersurface**  $M_c$  with initial value  $x_0 \in M_c$  is a solution  $(X_t)$  of the Stratonovich SDE

$$\circ dX_t = P(X_t) \circ dB_t = \sum_{k=1}^{n+1} P_k(X_t) \circ dB_t^k, \quad X_0 = x_0, \quad (4.16)$$

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 (4.16) 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 4.9.** *If  $X$  is a solution of (4.16) with  $x_0 \in M_c$  then almost surely,  $X_t \in M_c$  for any  $t \geq 0$ .*

The proof is very simple, but it relies on the classical chain rule in an essential way:

*Proof.* We have to show that  $f(X_t)$  is constant. This is an immediate consequence of the Stratonovich formula:

$$f(X_t) - f(X_0) = \int_0^t \nabla f(X_s) \cdot \circ dX_s = \sum_{k=1}^{n+1} \int_0^t \nabla f(X_s) \cdot P_k(X_s) \circ dB_s^k = 0$$

since  $P_k(x)$  is orthogonal to  $\nabla f(x)$  for any  $x$ . □

Although we have defined Brownian motion on the Riemannian manifold  $M_c$  in a non-intrinsic way, one can verify that it actually is an intrinsic object and does not depend on the embedding of  $M_c$  into  $\mathbb{R}^{n+1}$  that we have used. We only convince ourselves that the corresponding generator is an intrinsic object. By Theorem 4.8, the Brownian motion  $(X_t)$  constructed above is a solution of the martingale problem for the operator

$$\mathcal{L} = \frac{1}{2} \sum_{k=1}^{n+1} (P_k \cdot \nabla) P_k \cdot \nabla = \frac{1}{2} \sum_{k=1}^{n+1} P_k^2.$$

From differential geometry it is well-known that this operator is  $\frac{1}{2}\Delta_{M_c}$  where  $\Delta_{M_c}$  denotes the (intrinsic) **Laplace-Beltrami operator** on  $M_c$ .

**Exercise (Itô SDE for Brownian motion on  $M_c$ ).** Prove that the SDE (4.16) can be written in Itô form as

$$dX_t = P(X_t) dB_t - \frac{1}{2} \kappa(X_t) \mathbf{n}(X_t) dt$$

where  $\kappa(x) = \frac{1}{n} \operatorname{div} \mathbf{n}(x)$  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

$$\circ dX_t = b(X_t) dt + \sigma(X_t) \circ dB_t, \quad X_0 = x_0, \quad (4.17)$$

where  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is Lipschitz continuous and  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $C^2$  with bounded derivatives. Recall that (4.17) is equivalent to the Itô SDE

$$dX_t = \left( b + \frac{1}{2} \sigma \cdot \nabla \sigma \right) (X_t) dt + \sigma(X_t) dB_t, \quad X_0 = x_0. \quad (4.18)$$

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,

$$\circ dX_t = F'(B_t) \circ dB_t = \sigma(F(B_t)) \circ dB_t$$

provided  $F$  is a solution of the ordinary differential equation

$$F'(s) = \sigma(F(s)). \quad (4.19)$$

Hence a solution of (4.17) with initial condition  $X_0 = x_0$  is given by

$$X_t = F(B_t, x_0)$$

where  $(s, x) \mapsto F(s, x)$  is the *flow* of the vector field  $\sigma$ , i.e.,  $F(\cdot, x_0)$  is the unique solution of (4.19) with initial condition  $x_0$ .

Recall from the theory of ordinary differential equations that the flow of a vector field  $\sigma$  as above defines a diffeomorphism  $x_0 \mapsto F(s, x_0)$  for any  $s \in \mathbb{R}$ . To obtain a solution of (4.17) in the general case, we try the “variation of constants” ansatz

$$X_t = F(B_t, C_t) \quad (4.20)$$

with a continuous semimartingale  $(C_t)$  satisfying  $C_0 = x_0$ . 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 (4.20), 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 (4.17) provided  $(C_t)$  is almost surely absolutely continuous with derivative

$$\frac{d}{dt}C_t = \frac{\partial F}{\partial x}(B_t, C_t)^{-1} b(F(B_t, C_t)). \quad (4.21)$$

For every given  $\omega$ , the equation (4.21) 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 4.10 (Doss 1977, Sussmann 1978).** *Suppose that  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is Lipschitz continuous and  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $C^2$  with bounded derivatives. Then the flow  $F$  of the vector field  $\sigma$  is well-defined,  $F(s, \cdot)$  is a  $C^2$  diffeomorphism for any  $s \in \mathbb{R}$ , and the equation (4.21) has a unique pathwise solution  $(C_t)_{t \geq 0}$  satisfying  $C_0 = x_0$ . Moreover, the process  $X_t = F(B_t, C_t)$  is the unique strong solution of the equation (4.17), (4.18) respectively.*

We refer to [19] for detailed proof.

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

1)

$$dX_t = \frac{1}{2}X_t dt + \sqrt{1 + X_t^2} dB_t, \quad X_0 = 0,$$

2)

$$dX_t = X_t(1 + X_t^2) dt + (1 + X_t^2) dB_t, \quad X_0 = 1.$$

Do the solutions explode in finite time?

## 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 and not the Itô 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

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

Other smooth approximations could be used as well, cf. [19] and [17]. Let  $X^{(k)}$  denote the unique solution of the ordinary differential equation

$$\frac{d}{dt} X_t^{(k)} = b(X_t^{(k)}) + \sigma(X_t^{(k)}) \frac{d}{dt} B_t^{(k)}, \quad X_0^{(k)} = x_0$$

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

**Theorem 4.11 (Wong, Zakai 1965).** *Suppose that  $b$  is  $C^1$  with bounded derivatives and  $\sigma$  is  $C^2$  with bounded derivatives. Then almost surely as  $k \rightarrow \infty$ ,*

$$X_t^{(k)} \longrightarrow X_t \quad \text{uniformly on compact intervals,}$$

where  $(X_t)$  is the unique solution of the Stratonovich equation

$$\circ dX_t = b(X_t) dt + \sigma(X_t) \circ dB_t, \quad X_0 = x_0.$$

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  $\sigma$ , and the processes  $C^{(k)}$  and  $C$  solving (4.21), cf. [19]. The proof in the more interesting general case is much more involved, cf. e.g. Ikeda & Watanabe [17, Ch. VI, Thm. 7.2].

## 4.3 Stochastic flows

Let  $\Omega = C_0(\mathbb{R}_+, \mathbb{R}^d)$  endowed with Wiener measure  $\mu_0$  and the canonical Brownian motion  $W_t(\omega) = \omega(t)$ . We consider the Itô SDE

$$dX_t = b_t(X) dt + \sigma_t(X) dW_t, \quad X_0 = x, \quad (4.22)$$

with progressively measurable coefficients  $b, \sigma : \mathbb{R}_+ \times C(\mathbb{R}_+, \mathbb{R}^n) \rightarrow \mathbb{R}^n, \mathbb{R}^{n \times d}$  respectively satisfying the global Lipschitz condition

$$|b_t(x) - b_t(\tilde{x})| + \|\sigma_t(x) - \sigma_t(\tilde{x})\| \leq L (x - \tilde{x})_t^* \quad \forall t, x, \tilde{x} \quad (4.23)$$

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

$$\sup_{s \in [0, t]} (|b_s(0)| + \|\sigma_s(0)\|) < \infty \quad \forall t. \quad (4.24)$$

Then by Itô's existence and uniqueness theorem, there exists a unique global strong solution  $(X_t^x)_{t \geq 0}$  of (4.22) for any initial condition  $x \in \mathbb{R}^n$ . Our next goal is to show that there is a continuous modification  $(t, x) \mapsto \xi_t^x$  of  $(X_t^x)$ . The proof is based on the Kolmogorov-Centsov continuity theorem for stochastic processes.

### Kolmogorov-Centsov Theorem

**Theorem 4.12.** *Suppose that  $(E, \|\cdot\|)$  is a Banach space,  $C = \prod_{k=1}^d I_k$  is a product of real intervals  $I_1, \dots, I_d \subset \mathbb{R}$ , and  $X_u : \Omega \rightarrow E, u \in C$ , is an  $E$ -valued stochastic process (a random field) indexed by  $C$ . If there exists constants  $\gamma, c, \varepsilon \in \mathbb{R}_+$  such that*

$$E[\|X_u - X_v\|^\gamma] \leq c|u - v|^{d+\varepsilon} \quad \text{for any } u, v \in C, \quad (4.25)$$

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

$$E\left[\left(\sup_{u \neq v} \frac{\|\xi_u - \xi_v\|}{|u - v|^\alpha}\right)^\gamma\right] < \infty \quad \text{for any } \alpha \in [0, \varepsilon/\gamma]. \quad (4.26)$$

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

For the proof cf. e.g. [28, Ch. I, (2.1)]. (XXX include later)

**Example.** Brownian motion satisfies (4.25) with  $d = 1$  and  $\varepsilon = \frac{\gamma}{2} - 1$  for any  $\gamma \in (2, \infty)$ . Letting  $\gamma$  tend to  $\infty$ , we see that almost every Brownian path is  $\alpha$ -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. [9, Thm. 1.20].

### Existence of a continuous flow

We now apply the Kolmogorov-Centsov Theorem to the solution  $x \mapsto (X_s^x)_{s \leq t}$  of the SDE (4.22) as a function of its starting point.

**Theorem 4.13 (Flow of an SDE).** *Suppose that (4.23) and (4.24) hold.*

- 1) *There exists a function  $\xi : \mathbb{R}^n \times \Omega \rightarrow C(\mathbb{R}_+, \mathbb{R}^n)$ ,  $(x, \omega) \mapsto \xi^x(\omega)$  such that*
  - (i)  $\xi^x = (\xi_t^x)_{t \geq 0}$  *is a strong solution of (4.22) for any  $x \in \mathbb{R}^n$ , and*
  - (ii) *the map  $x \mapsto \xi^x(\omega)$  is continuous w.r.t. uniform convergence on finite time intervals for any  $\omega \in \Omega$ .*
- 2) *If  $\sigma(t, x) = \tilde{\sigma}(x_t)$  and  $b(t, x) = \tilde{b}(x_t)$  with Lipschitz continuous functions  $\tilde{\sigma} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  and  $\tilde{b} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  then  $\xi$  satisfies the **cocycle property***

$$\xi_{t+s}^x(\omega) = \xi_s^{\xi_t^x(\omega)}(\Theta^t(\omega)) \quad \forall s, t \geq 0, x \in \mathbb{R}^n \quad (4.27)$$

for  $\mu_0$ -almost every  $\omega$ , where

$$\Theta^t(\omega) = \omega(\cdot + t) \in C(\mathbb{R}_+, \mathbb{R}^d)$$

denotes the shifted path, and the definition of  $\xi$  has been extended by

$$\xi(\omega) := \xi(\omega - \omega(0)) \quad (4.28)$$

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

$$E[(X^x - X^y)_t^*]^p \leq c \cdot e^{ct} |x - y|^p \quad \text{for any } t \geq 0 \text{ and } x, y \in \mathbb{R}^n, \quad (4.29)$$

where  $X^x$  denotes a version of the strong solution of (4.22) with initial condition  $x$ .

Now fix  $t \in \mathbb{R}_+$ . We apply the Kolmogorov-Centsov Theorem with  $E = C([0, t], \mathbb{R}^n)$  endowed with the supremum norm  $\|X\|_t = X_t^*$ . By (4.29), there exists a modification  $\xi$  of  $(X_s^x)_{s \leq t, x \in \mathbb{R}^n}$  such that  $x \mapsto (\xi_s^x)_{s \leq t}$  is almost surely  $\alpha$ -Hölder continuous w.r.t.  $\|\cdot\|_t$  for any  $\alpha < \frac{p}{p-d}$ . Clearly, for  $t_1 \leq t_2$ , the almost surely continuous map  $(s, x) \mapsto \xi_s^x$

constructed on  $[0, t_1] \times \mathbb{R}^n$  coincides almost surely with the restriction of the corresponding map on  $[0, t_2] \times \mathbb{R}^n$ . Hence we can almost surely extend the definition to  $\mathbb{R}_+ \times \mathbb{R}^n$  in a consistent way.

2) Fix  $t \geq 0$  and  $x \in \mathbb{R}^n$ . Then  $\mu_0$ -almost surely, both sides of (4.27) solve the same SDE as a function of  $s$ . Indeed,

$$\begin{aligned}\xi_{t+s}^x &= \xi_t^x + \int_t^{t+s} \tilde{b}(\xi_u^x) du + \int_t^{t+s} \tilde{\sigma}(\xi_u^x) dW_u \\ &= \xi_t^x + \int_0^s \tilde{b}(\xi_{t+r}^x) dr + \int_0^s \tilde{\sigma}(\xi_{t+r}^x) d(W_r \circ \Theta^t),\end{aligned}$$

$$\xi_s^{\xi_t^x \circ \Theta^t} = \xi_t^x + \int_0^s \tilde{b}(\xi_r^{\xi_t^x \circ \Theta^t}) dr + \int_0^s \tilde{\sigma}(\xi_r^{\xi_t^x \circ \Theta^t}) d(W_r \circ \Theta^t)$$

hold  $\mu_0$ -almost surely for any  $s \geq 0$  where  $(\xi_r^{\xi_t^x \circ \Theta^t})(\omega) := \xi_r^{\xi_t^x(\omega)}(\Theta^t(\omega))$ . Strong uniqueness now implies

$$\xi_{t+s}^x = \xi_s^{\xi_t^x \circ \Theta^t} \quad \text{for any } s \geq 0, \quad \text{almost surely.}$$

Continuity of  $\xi$  then implies that the cocycle property (4.27) holds with probability one for all  $s, t$  and  $x$  simultaneously.  $\square$

**Remark.** 1) Since the constant  $p$  in the proof above can be chosen arbitrarily large, the argument yields  $\alpha$ -Hölder continuity of  $x \mapsto \xi^x$  for any  $\alpha < 1$ . It is also possible to prove joint Hölder continuity in  $t$  and  $x$ , cf. XXX. In Chapter 5.1 we will prove that under a slightly stronger assumption on  $b$  and  $\sigma$ , the flow is actually continuously differentiable in  $x$ .

2) For the validity of the cocycle property, strong uniqueness is essential.

Above we have shown the existence of a continuous flow for the SDE (4.22) on the canonical setup. From this we can obtain strong solutions on other setups:

**Exercise.** Show that the unique strong solution of (4.22) w.r.t. an arbitrary driving Brownian motion  $B$  instead of  $W$  is given by  $X_t^x(\omega) = \xi_t^x(B(\omega))$ .

In the time-homogeneous diffusion case, the Markov property for solutions of the SDE (4.22) is a direct consequence of the cocycle property:



**Corollary 4.14.** *Suppose that  $\sigma(t, x) = \tilde{\sigma}(x_t)$  and  $b(t, x) = \tilde{b}(x_t)$  with Lipschitz continuous functions  $\tilde{\sigma} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  and  $\tilde{b} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Then  $(\xi_t^x)_{t \geq 0}$  is a time-homogeneous  $(\mathcal{F}_t^{W,P})$  Markov process with transition function*

$$p_t(x, B) = P[\xi_t^x \in B], \quad t \geq 0, \quad x \in \mathbb{R}^n.$$

*Proof.* Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a measurable function. Then for  $0 \leq s \leq t$ ,

$$\Theta^t(\omega) = \omega(t) + (\omega(t + \cdot) - \omega(t)),$$

and hence, by the cocycle property and by (4.28),

$$f(\xi_{s+t}^x(\omega)) = f(\xi_s^{\xi_s^x(\omega)}(\omega(t + \cdot) - \omega(t)))$$

for a.e.  $\omega$ . Since  $\omega(t + \cdot) - \omega(t)$  is a Brownian motion starting at 0 independent of  $\mathcal{F}_t^{W,P}$ , we obtain

$$E[f(\xi_{s+t}^x) | \mathcal{F}_t^{W,P}](\omega) = E[f(\xi_s^{\xi_s^x(\omega)})] = (p_s f)(\xi_t^x(\omega)) \quad \text{almost surely.}$$

□

**Remark.** Without uniqueness, both the cocycle and the Markov property do not hold in general.

## 4.4 Local time

# Chapter 5

## Variations of SDE

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 [26], [25], [32], [17], [24] and [5].

Let  $\mu$  denote Wiener measure on the Borel  $\sigma$ -algebra of 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

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = x, \quad (5.1)$$

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:  $x \rightarrow x + h$
- Variations of the coefficients:  $b(x) \rightarrow b(\varepsilon, x)$ ,  $\sigma(x) \rightarrow \sigma(\varepsilon, x)$
- Variations of the driving paths:  $W_t \rightarrow W_t + H_t$ ,  $(H_t)$  adapted
- Variations of the underlying probability measure:  $\mu \rightarrow \mu^\varepsilon = Z^\varepsilon \cdot \mu$

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 solutions 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. [15] [25].

## 5.1 Variations of parameters in SDE

We now consider an SDE

$$dX_t^\varepsilon = b(\varepsilon, X_t^\varepsilon) dt + \sum_{k=1}^d \sigma_k(\varepsilon, X_t^\varepsilon) dW_t^k, \quad X_0^\varepsilon = x(\varepsilon), \quad (5.2)$$

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  $b, \sigma_k : U \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  are functions that are Lipschitz continuous in the second variable, and  $x : U \rightarrow \mathbb{R}^n$ . 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  $p \in [0, \infty)$  let

$$\|X^\varepsilon\|_p := E \left[ \sup_{t \in [0,1]} |X_t^\varepsilon|^p \right]^{1/p}.$$

**Exercise (Lipschitz dependence on  $\varepsilon$ ).** Prove that if the maps  $x, b$  and  $\sigma_k$  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

$$\|X^{\varepsilon+h} - X^\varepsilon\|_p \leq L_p |h|, \quad \text{for any } \varepsilon, h \in \mathbb{R}^m \text{ with } \varepsilon, \varepsilon + h \in U.$$

We now prove a stronger result under additional regularity assumptions:

### Differentiation of solutions w.r.t. a parameter

**Theorem 5.1.** *Let  $p \in [2, \infty)$ , and suppose that  $x$ ,  $b$  and  $\sigma_k$  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*

$$\begin{aligned} dY_t^\varepsilon &= \left( \frac{\partial b}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial b}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon \right) dt \\ &\quad + \sum_{k=1}^d \left( \frac{\partial \sigma_k}{\partial \varepsilon}(\varepsilon, X_t^\varepsilon) + \frac{\partial \sigma_k}{\partial x}(\varepsilon, X_t^\varepsilon) Y_t^\varepsilon \right) dW_t^k, \end{aligned} \quad (5.3)$$

$$Y_0^\varepsilon = x'(\varepsilon),$$

that is obtained by formally differentiating (5.2) w.r.t.  $\varepsilon$ .

Here and below  $\frac{\partial}{\partial \varepsilon}$  and  $\frac{\partial}{\partial x}$  denote the differential w.r.t. the  $\varepsilon$  and  $x$  variable, and  $x'$  denotes the (total) differential of the function  $x$ .

**Remark.** Note that if  $(X_t^\varepsilon)$  is given, then (5.3) is a linear SDE for  $(Y_t^\varepsilon)$  (with multiplicative noise). In particular, there is a unique strong solution. The SDE for the derivative process  $Y^\varepsilon$  is particularly simple if  $\sigma$  is constant: In that case, (5.3) is a deterministic ODE with coefficients depending on  $X^\varepsilon$ .

*Proof of 5.1.* We prove the stronger statement that there is a constant  $M_p \in (0, \infty)$  such that

$$\|X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h\|_p \leq M_p |h|^2 \quad (5.4)$$

holds for all  $\varepsilon, h \in \mathbb{R}^m$  with  $\varepsilon, \varepsilon + h \in U$ , where  $Y^\varepsilon$  is the unique strong solution of (5.3). Indeed, by subtracting the equations satisfied by  $X^{\varepsilon+h}$ ,  $X^\varepsilon$  and  $Y^\varepsilon h$ , we obtain for  $t \in [0, 1]$ :

$$|X_t^{\varepsilon+h} - X_t^\varepsilon - Y_t^\varepsilon h| \leq |\mathbf{I}| + \int_0^t |\mathbf{II}| ds + \sum_{k=1}^d \left| \int_0^t \mathbf{III}_k dW_s^k \right|,$$

where

$$\begin{aligned} \text{I} &= x(\varepsilon + h) - x(\varepsilon) - x'(\varepsilon)h, \\ \text{II} &= b(\varepsilon + h, X_s^{\varepsilon+h}) - b(\varepsilon, X_s^\varepsilon) - b'(\varepsilon, X_s^\varepsilon) \begin{pmatrix} 1 \\ Y_s^\varepsilon h \end{pmatrix}, \quad \text{and} \\ \text{III} &= \sigma_k(\varepsilon + h, X_s^{\varepsilon+h}) - \sigma_k(\varepsilon, X_s^\varepsilon) - \sigma'_k(\varepsilon, X_s^\varepsilon) \begin{pmatrix} 1 \\ Y_s^\varepsilon h \end{pmatrix}. \end{aligned}$$

Hence by Burkholder's inequality, there exists a finite constant  $C_p$  such that

$$E[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_t^{*p}] \leq C_p \cdot \left( |\mathbb{I}|^p + \int_0^t E[|\mathbb{II}|^p + |\mathbb{III}|^p] ds \right). \quad (5.5)$$

Since  $x, b$  and  $\sigma_k$  are  $C^2$  with bounded derivatives, there exist finite constants  $C_I, C_{II}, C_{III}$  such that

$$|\mathbb{I}| \leq C_I |h|^2, \quad (5.6)$$

$$|\mathbb{II}| \leq C_{II} |h|^2 + \left| \frac{\partial b}{\partial x}(\varepsilon, X_s^\varepsilon)(X_s^{\varepsilon+h} - X_s^\varepsilon - Y_s^\varepsilon h) \right|, \quad (5.7)$$

$$|\mathbb{III}_k| \leq C_{III} |h|^2 + \left| \frac{\partial \sigma_k}{\partial x}(\varepsilon, X_s^\varepsilon)(X_s^{\varepsilon+h} - X_s^\varepsilon - Y_s^\varepsilon h) \right|. \quad (5.8)$$

Hence there exist finite constants  $\tilde{C}_p, \hat{C}_p$  such that

$$E[|\mathbb{II}|^p + |\mathbb{III}|^p] \leq \tilde{C}_{III} \left( |h|^{2p} + \tilde{C}_{III} \int_0^t E[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_s^{*p}] ds \right),$$

and thus, by (5.6),

$$E[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_t^{*p}] \leq \hat{C}_p |h|^{2p} + \tilde{C}_p \int_0^t E[(X^{\varepsilon+h} - X^\varepsilon - Y^\varepsilon h)_s^{*p}] ds$$

for any  $t \leq 1$ . The assertion (5.4) now follows by Gronwall's lemma.  $\square$

## Derivative flow and stability of SDE

We now apply the general result above to variations of the initial condition, i.e., we consider the flow

$$d\xi_t^x = b(\xi_t^x) dt + \sum_{k=1}^d \sigma_k(\xi_t^x) dW_t^k, \quad \xi_0^x = x. \quad (5.9)$$

Assuming that  $b$  and  $\sigma_k$  ( $k = 1, \dots, d$ ) are  $C^2$  with bounded derivatives, Theorem 5.1 shows that the **derivative flow**

$$Y_t^x := \xi_t'(x) = \left( \frac{\partial}{\partial x^k} \xi_t^{x,l} \right)_{1 \leq k, l \leq n}$$

exists w.r.t.  $\|\cdot\|_p$  and  $(Y_t^x)_{t \geq 0}$  satisfies the SDE

$$dY_t^x = b'(\xi_t^x) Y_t^x dt + \sum_{k=1}^d \sigma_k'(\xi_t^x) Y_t^x dW_t^k, \quad Y_0^x = I_n. \quad (5.10)$$

Note that again, this is a linear SDE for  $Y$  if  $\xi$  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  $\omega$ , cf. [20] or [10].

In the sequel, we will denote the directional derivative of the flow  $\xi_t$  in direction  $v \in \mathbb{R}^n$  by  $Y_{v,t}$ :

$$Y_{v,t} = Y_{v,t}^x = Y_t^x v = \partial_v \xi_t^x.$$

(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

$$d\xi^x = b(\xi^x) dt + dW, \quad \xi_0^x = x;$$

and the derivative flow solves the ODE

$$dY^x = b'(\xi^x) Y dt, \quad Y_0 = I_n.$$

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  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is  $C^2$  with bounded derivatives, and let*

$$\kappa = \sup_{x \in \mathbb{R}^n} \sup_{\substack{v \in \mathbb{R}^n \\ |v|=1}} v \cdot b'(x)v.$$

*Then for any  $t \geq 0$  and  $x, y, v \in \mathbb{R}^n$ ,*

$$|\partial_v \xi_t^x| \leq e^{\kappa t} |v|, \quad \text{and} \quad |\xi_t^x - \xi_t^y| \leq e^{\kappa t} |x - y|.$$

The theorem shows in particular that exponential stability holds if  $\kappa < 0$ .

*Proof.* The derivative  $Y_{v,t}^x = \partial_v \xi_t^x$  satisfies the ODE

$$dY_v = b'(\xi)Y_v dt.$$

Hence

$$d|Y_v|^2 = 2Y_v \cdot b'(\xi)Y_v dt \leq 2\kappa|Y_v|^2 dt,$$

which implies

$$\begin{aligned} |\partial_v \xi_t^x|^2 &= |Y_{v,t}^x|^2 \leq e^{2\kappa t}|v|^2, \quad \text{and thus} \\ |\xi_t^x - \xi_t^y| &= \left| \int_0^1 \partial_{x-y} \xi_t^{(1-s)x+sy} ds \right| \leq e^{\kappa t}|x-y|. \end{aligned}$$

□

**Example (Ornstein-Uhlenbeck process).** Let  $A \in \mathbb{R}^{n \times n}$ . The generalized Ornstein-Uhlenbeck process solving the SDE

$$d\xi_t = A\xi_t dt + dW_t$$

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  $b, \sigma_1, \dots, \sigma_d : \mathbb{R}^n \rightarrow \mathbb{R}^n$  are  $C^2$  with bounded derivatives. Then for any  $t \geq 0$  and  $x, v \in \mathbb{R}^n$ , the derivative flow  $Y_{v,t}^x = \partial_v \xi_t^x$  is in  $L^2(\Omega, \mathcal{A}, P)$ , and*

$$\frac{d}{dt} E[|Y_{v,t}^x|^2] = 2E[Y_{v,t}^x \cdot K(\xi_t^x)Y_{v,t}^x]$$

where

$$K(x) = b'(x) + \frac{1}{2} \sum_{k=1}^d \sigma_k'(x)^T \sigma_k'(x).$$

*Proof.* Let  $Y_v^{(k)}$  denote the  $k$ -th component of  $Y_v$ . The Itô product rule yields

$$\begin{aligned} d|Y_v|^2 &= 2Y_v \cdot dY_v + \sum_k d[Y_v^{(k)}] \\ &\stackrel{(5.10)}{=} 2Y_v \cdot b'(\xi)Y_v dt + 2 \sum_k Y_v \cdot \sigma'_k(\xi) dW^k + \sum_k |\sigma'_k(\xi)Y_v|^2 dt. \end{aligned}$$

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

$$E[|Y_{v,t \wedge T_n}|^2] = |v|^2 + 2E\left[\int_0^{t \wedge T_n} Y_v \cdot K(\xi)Y_v ds\right].$$

The assertion follows as  $n \rightarrow \infty$ .  $\square$

**Theorem 5.4 (Exponential stability II).** *Suppose that the assumptions in Lemma 5.3 hold, and let*

$$\kappa := \sup_{x \in \mathbb{R}^n} \sup_{\substack{v \in \mathbb{R}^n \\ |v|=1}} v \cdot K(x)v. \quad (5.11)$$

Then for any  $t \geq 0$  and  $x, y, v \in \mathbb{R}^n$ ,

$$E[|\partial_v \xi_t^x|^2] \leq e^{2\kappa t} |v|^2, \quad \text{and} \quad (5.12)$$

$$E[|\xi_t^x - \xi_t^y|^{1/2}] \leq e^{\kappa t} |x - y|. \quad (5.13)$$

*Proof.* Since  $K(x) \leq \kappa I_n$  holds in the form sense for any  $x$ , Lemma 5.3 implies

$$\frac{d}{dt} E[|Y_{v,t}|^2] \leq 2\kappa E[|Y_{v,t}|^2].$$

(5.12) now follows immediately by Gronwall's lemma, and (5.13) follows from (5.12) since  $\xi_t^x - \xi_t^y = \int_0^1 \partial_{x-y} \xi_t^{(1-s)x+sy} ds$ .  $\square$

**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  $(\xi_t^x)$  of the SDE (5.1) with assumptions as in Lemma 5.3 and Theorem 5.4. Let

$$p_t(x, B) = P[\xi_t^x \in B], \quad x \in \mathbb{R}^n, \quad B \in \mathcal{B}(\mathbb{R}^n),$$

denote the transition function of the diffusion process on  $\mathbb{R}^n$ . For two probability measures  $\mu, \nu$  on  $\mathbb{R}^n$ , we define the  $L^2$  Wasserstein distance

$$\mathcal{W}_2(\mu, \nu) = \inf_{\substack{(X,Y) \\ X \sim \mu, Y \sim \nu}} E[|X - Y|^2]^{1/2}$$

as the infimum of the  $L^2$  distance among all couplings of  $\mu$  and  $\nu$ . Here a coupling of  $\mu$  and  $\nu$  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  $\kappa$  be defined as in (5.11).

**Corollary 5.5.** *For any  $t \geq 0$  and  $x, y \in \mathbb{R}^n$ ,*

$$\mathcal{W}_2(p_t(x, \cdot), p_t(y, \cdot)) \leq e^{\kappa t} |x - y|.$$

*Proof.* The flow defines a coupling between  $p_t(x, \cdot)$  and  $p_t(y, \cdot)$  for any  $t, x$  and  $y$ :

$$\xi_t^x \sim p_t(x, \cdot), \quad \xi_t^y \sim p_t(y, \cdot).$$

Therefore,

$$\mathcal{W}_2(p_t(x, \cdot), p_t(y, \cdot))^2 \leq E[|\xi_t^x - \xi_t^y|^2].$$

The assertion now follows from Theorem 5.4.  $\square$

**Exercise (Exponential convergence to equilibrium).** Suppose that  $\mu$  is a stationary distribution for the diffusion process, i.e.,  $\mu$  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  $x \in \mathbb{R}^d$ ,  $\mathcal{W}_2(p_t(x, \cdot), \mu) \rightarrow 0$  exponentially fast with rate  $\kappa$  as  $t \rightarrow \infty$ .

Besides studying convergence to a stationary distribution, the derivative flow is also useful for computing and controlling derivatives of transition 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  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . We still assume the conditions from Lemma 5.3.

**Exercise (Lipschitz bound).** Prove that for any Lipschitz continuous function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$\|p_t f\|_{\text{Lip}} \leq e^{\kappa t} \|f\|_{\text{Lip}} \quad \forall t \geq 0,$$

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

$$v \cdot \nabla_x p_t f = E[Y_{v,t}^x \cdot \nabla_{\xi_t^x} f] \quad \forall x, v \in \mathbb{R}^n. \quad (5.14)$$

Here  $\nabla_x p_t f$  denotes the gradient evaluated at  $x$ . Note that  $Y_{t,v}^x \cdot \nabla_{\xi_t^x} f$  is the directional derivative of  $f$  in the direction of the derivative flow  $Y_{t,v}^x$ .

*Proof of 5.6.* For  $\lambda \in \mathbb{R} \setminus \{0\}$ ,

$$\frac{(p_t f)(x + \lambda v) - (p_t f)(x)}{\lambda} = \frac{1}{\lambda} E[f(\xi_t^{x+\lambda v}) - f(\xi_t^x)] = \frac{1}{\lambda} \int_0^\lambda E[Y_{v,t}^{x+sv} \cdot \nabla_{\xi_t^{x+sv}} f] ds.$$

The assertion now follows since  $x \mapsto \xi_t^x$  and  $x \mapsto Y_{v,t}^x$  are continuous,  $\nabla f$  is continuous and bounded, and the derivative flow is bounded in  $L^2$ .  $\square$

The first Bismut-Elworthy Formula shows that the gradient of  $p_t f$  can be controlled by the gradient of  $f$  for all  $t \geq 0$ . In Section ??, we will see that by applying an integration by parts on the right hand side of (5.14), for  $t > 0$  it is even possible to control the gradient of  $p_t f$  in terms of the supremum norm of  $f$ , provided a non-degeneracy condition holds, cf. (??).

# Index

- $\alpha$ -stable, 34
- càdlàg, 6
- characteristic exponent, 9
- compensated Poisson point process, 22
- compensated Poisson process, 11
- compound Poisson distribution, 15
- compound Poisson process, 13
- Gamma process, 30
- Gaussian Lévy processes, 30
- infinitely divisible, 7
- Inverse Gaussian subordinator, 32
- Lévy measure, 16
- Lévy subordinator, 32
- Lévy-Itô decomposition, 29
- Lévy-Khinchin formula, 29
- Mapping theorem, 19
- Poisson point process, 18
- Poisson process, 11
- Poisson random measure, 12
- semimartingale, 39
- spatial Poisson process, 12
- subordinator, 30
- Variance Gamma process, 33

# Bibliography

- [1]
- [2]
- [3] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge UP, Cambridge, 2004.
- [4] Jean Bertoin. *Lévy Processes*. Cambridge UP, Cambridge, 1998.
- [5] Jean-Michel Bismut. *Large deviations and the Malliavin calculus*. Birkhaeuser, Boston, 1994.
- [6] Wolfgang Doeblin and Marc Yor. *pli cacheté*.
- [7] Andreas Eberle. Markov processes. <http://www.uni-bonn.de/~eberle/MarkovProcesses/MPSkript.pdf>, 2009.
- [8] Andreas Eberle. Wahrscheinlichkeitstheorie und Stochastische Prozesse. <http://www.uni-bonn.de/~eberle/StoPr/WthSkriptneu.pdf>, 2010.
- [9] Andreas Eberle. Introduction to Stochastic Analysis. <http://www.uni-bonn.de/~eberle/StoAn1011/StoAnSkriptneu.pdf>, 2011.
- [10] David Elworthy. *On the Geometry of Diffusion Operators and Stochastic Flows*. Springer, 1999, Berlin.
- [11] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes*. Wiley-Interscience, Hoboken, NJ, 2005.

- 
- [12] William Feller. *Introduction to Probability Theory and Its Applications*. Wiley, New York, 1957.
- [13] Peter K. Friz and Nicolas Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge UP, Cambridge, 2010.
- [14] W. Hackenbroch and A. Thalmaier. *Stochastic Analysis*. Teubner, 1991.
- [15] Martn Hairer. On malliavin’s proof of hörmander’s theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.
- [16] Elton Hsu. *Stochastic Analysis on Manifolds*. Oxford University Press, Oxford, 2002.
- [17] Nobuyuki Ikeda and Takehiko Ed. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, Amsterdam, 1981.
- [18] Jean Jacod and Albert Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, Berlin, 2003.
- [19] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Springer, New York, 2010.
- [20] Hiroshi Kunita. *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge, 1997.
- [21] Andreas Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin, 2006.
- [22] Terry J. Lyons. *Differential Equations Driven by Rough Paths, Ecole d’Eté de Probabilités de Saint-Flour XXXIV*. Springer, 2004.
- [23] Dilip B. Madan and Eugene Seneta. *The Variance Gamma Process and Option Pricing*.
- [24] Paul Malliavin. *Stochastic Analysis*. Springer, Berlin, 2002.

- [25] David Nualart. *The Malliavin Calculus and Related Topics*. Springer, Berlin, 2006.
- [26] Giuseppe Da Prato. *Introduction to Stochastic Analysis and Malliavin Calculus*. Edizioni della Normale, Pisa, 2009.
- [27] Philip E. Protter. *Stochastic Integration and Differential Equations*. Springer, Berlin, 2005.
- [28] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer, Berlin, 2005.
- [29] L. C. G. Rogers and David Williams. *Diffusions, Markov processes and martingales, Vol. 2: Ito calculus*. Cambridge UP, Cambridge, 2008.
- [30] Shreve. *Stochastic calculus for Finance II*. Springer, New York, 2004.
- [31] Daniel W. Stroock and S.R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer, Berlin, 2005.
- [32] Shinzo Watanabe. *Lectures on stochastic differential equations and Malliavin calculus*. Springer, Berlin, 1984.