# Markov Processes

Winter, 2009/2010, Uni Bonn

Anton Bovier
*Institut für Angewandte Mathematik*
*Rheinische Friedrich-Wilhelms-Universität Bonn*
*Endenicher Allee 60*
*53115 Bonn*

*Version: July 5, 2012*

# Contents

# 1

# Continuous time martingales

In the last course we have seen that martingales play a truly fundamental rôle in the theory of stochastic processes in discrete time, and in particular we have seen an intimate connection between martingales and Markov processes. In this course we will seriously engage in the study of continuous time processes where this relation will play an even more central rôle. Therefore, we begin with the extension of martingale theory to the continuous time setting. We will see that this will go quite smoothly, but we will have to worry about a number of technical details. Most of the material in this Chapter is from Rogers and Williams [13].

## 1.1 Càdlàg functions

In the example of Brownian motion we have seen that we could construct this continuous time process on the space of continuous functions. This setting is, however, too restrictive for the general theory. It is quite important to allow for stochastic processes to have jumps, and thus live on spaces of discontinuous paths. Our first objective is to introduce a sufficiently rich space of such functions that will still be manageable.

**Definition 1.1.1** A function $f : \mathbb{R}_+ \to \mathbb{R}$ is called a *càdlàg* [1] function, iff

(i) for every $t \geq 0$, $f(t) = \lim_{s\downarrow t} f(s)$, and
(ii) for every $t > 0$, $f(t-) = \lim_{s\uparrow t} f(s)$ exists.

Recall that this definition should remind you of distribution functions. In fact, a probability distribution function is a non-decreasing càdlàg function.

[1] From "continue à droite, limites à gauche".

It will be important to be able to extend functions specified on countable sets to càdlàg functions.

**Definition 1.1.2** A function $y : \mathbb{Q}_+ \to \mathbb{R}$ is called *regularisable*, iff

(i) for every $t \geq 0$, $\lim_{q \downarrow t} y(q)$ exists finitely, and
(ii) for every $t > 0$, $y(t-) = \lim_{q \uparrow t} y(s)$ exists finitely.

Regularisability is linked to properties of upcrossings. We define this important concept for functions from the rationals to $\mathbb{R}$.

**Definition 1.1.3** Let $y : \mathbb{Q}_+ \to \mathbb{R}$, $N \in \mathbb{N}$ and let $a < b \in \mathbb{R}$. Then the number $U_N(y, [a,b]) \in \mathbb{N} \cup \{\infty\}$ of upcrossings of $[a,b]$ by $y$ during the interval $[0,N]$ is the supremum over all $k \in \mathbb{N}$, such that there are rational numbers $q_i, r_i \in \mathbb{Q}$, $i \leq k$ with the property that

$$0 \leq q_1 < r_1 < \cdots < q_k < r_k \leq N$$

and

$$y(q_i) < a < b < y(r_i), \quad \text{for all } 1 \leq i \leq k.$$

**Theorem 1.1.1** *Let $y : \mathbb{Q}_+ \to \mathbb{R}$. Then $y$ is regularisable if and only if, for all $N \in \mathbb{N}$ and $a < b \in \mathbb{R}$,*

$$\sup\{|y(q)| : q \in \mathbb{Q} \cap [0,N]\} < \infty, \tag{1.1}$$

*and*

$$U_N(y, [a,b]) < \infty. \tag{1.2}$$

*Proof.* Let us first show that the two conditions are sufficient. To do so, assume that $\limsup_{q \downarrow t} y(q) > \liminf_{q \downarrow t} y(q)$. Then choose $b > a$ such that $\limsup_{q \downarrow t} y(q) > b > a > \liminf_{q \downarrow} y(q)$. Then, for $N > t$, $y(q)$ must cross $[a,b]$ infinitely many times, i.e. $U_N(y, [a,b]) = +\infty$, contradicting assumption (1.2). Thus the limit $\lim_{q \downarrow t} y(q)$ exists, and by (1.1) it is finite. The same argument applies to the limit from below.

Next we show that the conditions are necessary. Assume that for some $N$ $y(q)$ is unbounded on $[0,N]$. Then for any $n$ there exists $q_n$ such that $|y(q_n)| > n$. The set $\cup_n \{q_n\}$ must be infinite, since otherwise $q$ will be infinite on a finite set, contradicting the assumption that it takes values in $\mathbb{R}$. Hence this set has at least one accumulation point, $t$. But then either $\lim_{q \uparrow t} y(q)$ or $\lim_{q \downarrow t} y(q)$ must be infinite, hence $y$ is not regularisable.

Assume now that $U_N(y; [a,b]) = \infty$. Define $t \equiv \inf\{r \in \mathbb{R}_+ : U_r(y; [a,b]) = \infty\}$. Then there are infinitely many upcrossings of $[a,b]$

in any interval $[t-\varepsilon, t]$ or in the interval $[t, t+\varepsilon]$, for any $\varepsilon > 0$. In the first case, this implies that $\limsup_{q\uparrow t} y(y) \geq b$ and $\liminf_{q\uparrow t} y(y) \leq a$, which precludes the existence of that limit. In the second case, the same argument precludes the existence of the limit $\lim_{q\downarrow t} y(y)$. □

One of the main points of Theorem 1.1.1 is that it can be used to show that the property to be regularisable is measurable.

**Corollary 1.1.2** *Let $\{Y_q, q \in \mathbb{Q}_+\}$ be a stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and let*

$$G \equiv \{\omega \in \Omega : q \to Y_q(\omega) \text{ is regularisable}\} \tag{1.3}$$

*Then $G \in \mathcal{F}$.*

*Proof.* By Theorem 1.1.1, to check regularisability we have to take countable intersections and unions of finite dimensional cylinder sets which are all measurable. Thus regularisability is a measurable property. □

Next we observe that from a regularisable function we can readily obtain a càdlàg function by taking limits from the right.

**Theorem 1.1.3** *Let $y : \mathbb{Q}_+ \to \mathbb{R}$ be a regularisable function. Define, for any $t \in \mathbb{R}_+$,*

$$f(t) \equiv \lim_{q\downarrow t} y(q). \tag{1.4}$$

*Then $f$ is càdlàg .*

The proof is obvious and left to the reader.

## 1.2 Filtrations, supermartingales, and càdlàg processes

We begin with a probability space $(\Omega, \mathcal{G}, \mathbb{P})$. We define a continuous time filtration $\mathcal{G}_t, t \in \mathbb{R}_+$ essentially as in the discrete time case.

**Definition 1.2.1** A filtration $(\mathcal{G}_t, t \in \mathbb{R}_+)$ of $(\Omega, \mathcal{G}, \mathbb{P})$ is an increasing family of sub-$\sigma$-algebras $\mathcal{G}_t$, such that, for $0 \leq s < t$,

$$\mathcal{G}_s \subset \mathcal{G}_t \subset \mathcal{G}_\infty \equiv \sigma\left(\bigcup_{r\in\mathbb{R}_+} \mathcal{G}_r\right) \subset \mathcal{G}. \tag{1.5}$$

We call $(\Omega, \mathcal{G}, \mathbb{P}; (\mathcal{G}_t, t \in \mathbb{R}_+))$ a filtered space.

**Definition 1.2.2** A stochastic process, $\{X_t, t \in \mathbb{R}_+\}$, is called *adapted* to the filtration $\{\mathcal{G}_t, t \in \mathbb{R}_+\}$, if, for every $t$, $X_t$ is $\mathcal{G}_t$-measurable.

**Definition 1.2.3** A stochastic process, $X$, on a filtered space is called a *martingale*, if and only if the following hold:

(i) The process $X$ is adapted to the filtration $\{\mathcal{G}_t, t \in \mathbb{R}_+\}$;
(ii) For all $t \in \mathbb{R}_+$, $\mathbb{E}|X_t| < \infty$;
(iii) For all $s \leq t \in \mathbb{R}_+$,

$$\mathbb{E}(X_t|\mathcal{G}_s) = X_s, \text{ a.s.}. \tag{1.6}$$

Sub- and super-martingales are define in the same way, with "=" in (1.6) replaced by "≥" resp. "≤".

We see that so far almost nothing changed with respect to the discrete time setup. Note in particular that if we take a monotone sequence of points $t_n$, then $Y_n \equiv X_{t_n}$ is a discrete time martingale (sub, super) whenever $X_t$ is a continuous time martingale (sub, super).

The next lemma is important to connect martingale properties to càdlàg properties.

**Lemma 1.2.4** *Let $Y$ be a supermartingale on a filtered space $(\Omega, \mathcal{G}, \mathbb{P}; (\mathcal{G}_t, t \in \mathbb{R}_+))$. Let $t \in \mathbb{R}_+$ and let $q(-n)$, $n \in \mathbb{N}$, such that $q(-n) \downarrow t$, as $n \uparrow \infty$. Then*

$$\lim_{q(-n)\downarrow t} Y_{q(-n)}$$

*exists a.s. and in $\mathcal{L}^1$.*

*Proof.* This is an application of the Lévy-Doob downward theorem (see [2], Thm. 4.2.9). □

Spaces of càdlàg functions are the natural setting for stochastic processes. We define this in a strict way.

**Definition 1.2.4** A stochastic process is called a càdlàg process, if all its sample paths are càdlàg functions. càdlàg processes that are (super,sub) martingales are called càdlàg (super,sub) martingales.

**Remark 1.2.1** Note that we do not just ask that almost all sample paths are càdlàg .

## 1.3 Examples

**Brownian motion** We have already seen that Brownian motion is defined in such a way that all its sample paths are continuous, and thus a fortiori càdlàg . We had also argued that Brownian motion is a martingale, and from the definition of continuous time martingales given above, we see that we checked exactly the right things. Thus Brownian motion is our first example of a càdlàg martingale.

**Poisson process.** As a second example we will construct a Poisson counting process. We begin with a $\sigma$-finite measure space $(W, \mathcal{W}, \lambda)$ where $\mathcal{W}$ is assumed to contain all points (think of $W \subset \mathbb{R}, \mathcal{W} = \mathcal{B}(\mathbb{R})$). Assume first that $\lambda(W) < \infty$. Then we can construct, on a probability space $(\Omega, \mathcal{G}, \mathbb{P})$, a family of independent random variables

$$N, Z_1, Z_2, \ldots,$$

where

(i) $N$ is a Poisson random variable with parameter $\lambda(W)$, i.e.

$$\mathbb{P}[N = n] = \frac{\lambda(W)^n}{n!} e^{-\lambda(W)},$$

for all $n \in \mathbb{N}_0$, and

(ii)

$$\mathbb{P}[Z_k \in B] = \frac{\lambda(B)}{\lambda(W)},$$

for all $B \in \mathcal{W}$. Then we can construct the *random measure*, $\Lambda$, by

$$\Lambda_W(B, \omega) \equiv \sum_{n=1}^{N(\omega)} \mathbb{1}_B(Z_n(\omega)),$$

for $B \in \mathcal{W}$ and $\omega \in \Omega$.

**Exercise:** Verify by direct computation that if $W_1 \subset W$, and $W_2 \equiv W \backslash W_1$, then

$$\Lambda_W = \Lambda_{W_1} + \Lambda_{W_2}.$$

where $\Lambda_{W_i}$ are independent of each other.

On the basis of the exercise, we can readily extend the construction to the case where $\lambda$ is only $\sigma$-finite. In that case, there exists a disjoint partition $W_i$, $\cup_i W_i = W$, with $\lambda(W_i) < \infty$. Thus we can construct independent random measures $\Lambda_{W_i}$ and set

$$\Lambda_W(B,\omega) \equiv \sum_i \Lambda_{W_i}(B \cap W_i, \omega). \tag{1.7}$$

This defines the Poisson process. Note that the result of the exercise is crucial to guarantee that this construction is consistent and independent of the choice of the partition $W_i$.

Now let $W = \mathbb{R}_+$, and $\lambda$ the Lebesgue measure. We can define the random functions

$$N_t(\omega) \equiv \Lambda_{\mathbb{R}_+}([0,t],\omega) \equiv \Lambda([0,t],\omega).$$

By construction, these functions are càdlàg for every $\omega$, and so $N_t$ is a càdlàg process. This process is called a Poisson counting process. Moreover, by the properties of the Poisson process, for $s < t$,

$$N_t - N_s = \Lambda((s,t])$$

is independent of $\mathcal{G}_s \equiv \sigma(N_r, r \leq s)$ and $\mathbb{E}(N_t - N_s) = t - s$. Therefore, the process $C_t \equiv N_t - t$ is a càdlàg martingale.

**Lévy processes** An important class of càdlàg processes generalizes both Brownian motion as well as the Poisson counting process. Their characterization is the independence of increments. Quite naturally, they generalize the notion of sums of independent random variables to continuous time processes, and it will not be surprising that they appear as limits of these in non-central limit theorems. An excellent presentation of these *Lévy processes* was given by Kiyosi Itô in his Aarhus lectures [8]. Another good reference is Bertoin's book [1].

**Definition 1.3.1** A stochastic process $(X_t, t \in \mathbb{R}_+)$ with values in $\mathbb{R}^d$ is called a Lévy process, if:

(i) $X_t$ is a càdlàg process;
(ii) For any collection $0 \leq t_1 < t_2 \cdots < t_k < \infty$, the family of random variables
$$Y_i \equiv X_{t_i} - X_{t_{i-1}}, \quad i = 1, \dots, k$$
is independent;
(iii) The law of $X_{t+h} - X_t$ is independent of $t$.

The theory of Lévy processes is intimately linked to the theory of infinitely divisible laws, and we will provide some background information on this.

**Definition 1.3.2** A probability measure on $\mathbb{R}^d$ is called *infinitely divisible* if, for each $n$, there exists a probability measure, $\mu_n$, on $\mathbb{R}^d$, such that, if $V_i$ are independent random variables with law $\mu_n$, then the law of

$$\sum_{i=1}^{n} V_i$$

is $\mu$.

The connection with Lévy processes is apparent, since clearly the law of $X_t$ is infinitely divisible, being the law of the sum of iid random variables $Y_i \equiv X_{it/n} - X_{(i-1)t/n}$. Note also that the Gaussian distribution is infinitely divisible, and that Brownian motion is the corresponding Lévy process.

The following famous theorem gives a complete characterization of infinitely divisible laws. We will state it without proof, but give the proof in a special case.

**Theorem 1.3.5** *For each $b \in \mathbb{R}^d$, and non-negative definite matrix $M$, and each measure, $\nu$, on $\mathbb{R}^d \backslash \{0\}$, that satisfies*

$$\int \min(|x|^2, 1)\nu(dx) < \infty, \tag{1.8}$$

*the function*

$$\phi(\theta) \equiv \exp(\psi(\theta)),$$

*where*

$$\psi(\theta) \equiv i(b,\theta) - \frac{1}{2}(\theta, M\theta) + \int (e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\leq 1})\nu(dx), \tag{1.9}$$

*is the characteristic function of an infinitely divisible distribution. Moreover, the characteristic function of any infinitely divisible law can be written in this form with uniquely determined $(b, M, \nu)$.*

Note that it is easy to see that any law of the form given above is infinitely divisible. Namely, for any $n \in \mathbb{N}$, consider the function

$$\psi_n(\theta) \equiv \frac{1}{n}\psi(\theta).$$

Then $\phi_n$ corresponds to a Lévy triple $(b/n, M/n, \nu/n)$, and if $X_i$ are iid with characteristic function $\exp(\psi_n(\theta))$, then $\sum_{i=1}^n X_i$ has the characteristic function $\phi$.

In the case of distributions that take values on the positive reals only, one has the following alternative result. Its proof will be easier since here the characterisation involves the Laplace rather than the Fourier transform.

**Theorem 1.3.6** *Let $F$ be a distribution function on $\mathbb{R}_+$. Then $F$ is the distribution function of an infinitely divisible law, iff, for*
*geq0,*

$$\int_0^\infty e^{-\lambda x} F(dx) = \exp\left[-c\lambda - \int_0^\infty \left(1-e^{-x}\right)\mu(dx)\right], \tag{1.10}$$

*where $c \in \mathbb{R}$ and $\mu$ is a measure on $(0,\infty)$ such that*

$$\int_0^\infty (x \wedge 1)\mu(dx) < \infty. \tag{1.11}$$

*Proof.* The fact that the right-hand side of (1.10) represent the Laplace transform of an infinitely divisible law follows by inspection. The converse direction is more interesting. The starting observation is that the infinite divisibility implies that for any $n \in \mathbb{N}$, there exists distribution functions with support on $\mathbb{R}_+$ such that

$$F_n^*(\lambda) \equiv \int_0^\infty e^{-\lambda x} F_n(dx) = \left[F^*(\lambda)\right]^{1/n}. \tag{1.12}$$

Clearly $F_n^*(\lambda) \uparrow 1$, uniformly on compact subsets of $\mathbb{R}_+$. Taking logarithms, we get first that

$$\ln F^*(\lambda) = n \ln F_n^*(\lambda) = n \ln\left(1-(1-F_n^*(\lambda))\right) \leq -n(1-F_n^*(\lambda)). \tag{1.13}$$

We want to proof that for $n$ large, the last inequality is essentially an equality. To see this, note that the convergence of $F_n^*$ mentioned above means that for any $\delta > 0$ and $K < \infty$, there exists $n_0 < \infty$, such that for all $\lambda \leq K$ and $n \geq n_0$,

$$(1 - F_n(\lambda)) \leq \delta. \tag{1.14}$$

On the other hand, for any $\varepsilon > 0$, there exists $\delta > 0$ such that for all $0 \leq x \leq \delta$,

$$\ln(1+x) > (1+\varepsilon)x. \tag{1.15}$$

Hence, for all $\varepsilon > 0, K < \infty$, there exists $n_0 < \infty$, such that for all $\lambda \leq K$ and $n \geq n_0$,

$$\ln\left(1-(1-F_n^*(\lambda))\right) > -(1+\varepsilon)\left(1-(1-F_n^*(\lambda))\right). \tag{1.16}$$

Thus,

$$n(1-F_n^*(\lambda)) \to -\ln F^*(\lambda), \text{as } n \uparrow \infty, \tag{1.17}$$

uniformly on compact intervals. Now we can write

$$n(1-F_n^*(\lambda)) = n\int\left(1-e^{-\lambda x}\right)F_n(dx) = \int \frac{1-e^{-\lambda x}}{1-e^{-x}} n\left(1-e^{-x}\right)F_n(dx). \tag{1.18}$$

Now $m_n(dx) \equiv n\,(1-e^{-x})\,F_n(dx)$ is a measure on $(0,\infty)$ with total mass $n(1-F_n^*(1))$, which by the observations above converges to the finite value $-\ln F^*(1)$. Hence there exist subsequences (which) along which $m_n$ converges to some finite measure $m$ on $[0,+\infty]$. Then

$$n(1-F_n^*(\lambda)) \to m(\{0\})\lambda + \int \frac{1-e^{-\lambda x}}{1-e^{-x}} m(dx) + m(\{+\infty\}), \tag{1.19}$$

which thus must be $-\ln F^*(\lambda)$. The first two terms are what we want (with $m(0)=c$ and $m(dx)=(1-e^{-x})\mu(dx)$, while setting $\lambda=0$ shows that in fact

$$\ln F^*(0) = \ln 1 = 0 = m(\{+\infty\}).$$

This proves the theorem. □

The description of infinitely divisible laws in terms of the (Lévy) triplets $(b, M, \nu)$ is called the Lévy-Khintchine representation, $\nu$ is called the Lévy measure, and $\psi$ the characteristic exponent.

We now use the Lévy-Khintchine representation to study Lévy processes. Since $X_t = \sum_{i=1}^t Y_i$ where $Y_i$ has the same law as $X_1$ (assume $t \in \mathbb{N}$ for a moment), we should expect that

$$\mathbb{E}\exp\left(i(\theta, X_t)\right) = \exp\left(t\psi(\theta)\right) \tag{1.20}$$

where $\psi$ is the characteristic exponent of the distribution of $X_1$. In fact, for any infinitely divisible law, (1.20) provides a characteristic function of a process with independent and stationary increments. Let $\mu_t$ be the law of $X_t$. Just like in the case of Brownian motion, we can thus define a Markov transition kernel for the process $X$ via

$$(P_t f)(x) \equiv \int f(x+y)\mu_t(dy), \tag{1.21}$$

for bounded continuous functions, $f$, vanishing at infinity. We will see later that properties of this transition kernel guarantee that $X$ can be constructed as a càdlàg process, and hence a Lévy process.

An important example of Lévy processes can be constructed from Poisson counting processes. Let $N_t$ be a Poisson counting process, and let $Y_i, i \in \mathbb{N}$ be iid real random variables with distribution function $F$.

Then define

$$X_t \equiv \sum_{i=1}^{N_t} Y_i.$$

Clearly $X$ has càdlàg paths and independent increments (both the increments of $N_t$ and the accumulated $Y's$ are independent). Moreover, it is easy to compute the characteristic function of $X_{t+s} - X_t$:

$$\begin{aligned}
\mathbb{E}e^{i(\theta, X_{t+s}-X_t)} &= \sum_{n=0}^{\infty} \frac{s^n e^{-s}}{n!} \left( \int e^{i(\theta,x)} F(dx) \right)^n \qquad (1.22)\\
&= \exp\left( s \int \left( e^{i(\theta,x)} - 1 \right) F(dx) \right) \\
&= \exp\Bigg( si \left( \theta, \int_{|x|\le 1} xF(dx) \right) \\
&\quad + s \int \left( e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\le 1} \right) F(dx) \Bigg)
\end{aligned}$$

Thus $X$ is a Lévy process, called a *compound Poisson process* with Lévy triple $(\int_{\|x\|\le 1} xF(dx), 0, F)$, where the Lévy measure is finite.

Compound Poisson processes are of course pure jump processes, i.e. the only points of change are discontinuities. We will, as an application, show that a non-trivial Lévy measure always makes a Lévy process discontinuous, i.e. produces jumps. This is the content of Lévy's theorem:

**Theorem 1.3.7** *If $X$ is a Lévy process with continuous paths, then it's Lévy triple is of the form $(b, M, 0)$, i.e.*

$$X_t = MB_t + bt,$$

*where $B_t$ is Brownian motion.*

*Proof.* Let $X_t$ be a Lévy process with triple $(b, M, \nu)$. Fix $\varepsilon \in (0,1)$ and construct an independent Lévy process with characteristic exponent

$$\psi_\varepsilon(\theta) \equiv i(b,\theta) - \frac{1}{2}(\theta, M\theta) + \int_{|x|\le\varepsilon} (e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\le 1})\nu(dx).$$

Finally set $\psi^\varepsilon(\theta) \equiv \psi(\theta) - \psi_\varepsilon(\theta)$, i.e.

$$\psi^\varepsilon(\theta) = \int_{|x|>\varepsilon} (e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\le 1})\nu(dx).$$

Due to the integrability assumption of Lévy measures, $\int_{|x|>\varepsilon} \nu(dx) < \infty$, and therefore, the process $Y^\varepsilon$ with characteristic exponent $\psi_\varepsilon$ is a compound Poisson process, and as such has only finitely many jumps on any compact interval. If $X^\varepsilon$ is the process with exponent $\psi^\varepsilon$, independent of $Y^\varepsilon$, then $X^\varepsilon + Y^\varepsilon$ have the same law as $X$. Now $X^\varepsilon$ has only countably many jumps, that occur at times independent of the process $Y^\varepsilon$. But this means that, with probability one, all the jumps of $Y^\varepsilon$ occur at times when there is no jump of $X^\varepsilon$, and whence $X$ jumps whenever $Y^\varepsilon$ jumps. But this means that $X$ cannot be continuous, unless the process $Y^\varepsilon$ never jumps, which is only the case if $\nu = 0$. This proves the theorem. □

A slightly different look at the construction of compound Poissom processes will provide us with the means to construct general Lévy processes with pure jump part. For notational simplicity we consider only the case of Lévy processes with values in $\mathbb{R}$. To this end, let $\nu$ be any measure on $\mathbb{R}$ that satisfies the integrability condition (1.8). For $\varepsilon > 0$, set $\nu_\varepsilon(dx) \equiv \nu(dx)\mathbb{1}_{|x|>\varepsilon}$. Then $\nu^\varepsilon$ is a finite measure. Define the measures on $\lambda_e(dx, dt) \equiv \nu^\varepsilon(dx)dt$ be a measure on $\mathbb{R}^2$. Then we can associate to $\lambda_\varepsilon$ a Poisson process, $\mathcal{P}_\varepsilon$, on $\mathbb{R}^2$ with intensity measure $\lambda_\varepsilon$. Clearly, for any $\varepsilon > 0$, and any $t < \infty$, $\nu^\varepsilon((0, t] \times \mathbb{R}) < \infty$. Thus we can define the functions

$$X^\varepsilon(t) \equiv \int_0^t \int x\mathcal{P}^\varepsilon(ds, dx). \tag{1.23}$$

Note that this is nothing but a random finite sum, and in fact, up to a time change, a compound Poisson process (with $Y$ distributed according to the normalization of the measure $\nu^\varepsilon$). Now we may ask whether the limit $\varepsilon \downarrow 0$ of these processes exists as a Lévy process. To do this, we would like to argue that

$$\int_0^t \int x\mathcal{P}(ds, dx) = \int_0^t \int x\mathcal{P}^\varepsilon(ds, dx) + \int_0^t \int_{|x|<\varepsilon} x\mathcal{P}(ds, dx)$$

and that the second integral tends to zero as $\varepsilon \downarrow 0$. A small problem with this is that we cannot be sure under our conditions on $\nu$ that

$$\mathbb{E} \int_0^t \int_{|x|<\varepsilon} x\mathcal{P}(ds, dx) = \int_0^t \int_{|x|<\varepsilon} x\lambda(ds, dx) = t \int_{|x|<\varepsilon} x\nu(dx)$$

is finite. To remedy this problem, we modify the definition of our target process and set

$$X(t) \equiv ct + \int_0^t \int x \left(\mathcal{P}(ds, dx) - \mathbb{1}_{|x|\leq 1}\nu(dx)\right). \tag{1.24}$$

This can indeed be decomposed as (for $0 < \varepsilon < 1$)

$$X(t) = ct + \int_0^t \int_{|x|>\varepsilon} x \left(\mathcal{P}(ds,dx) - \mathbb{1}_{|x|\leq 1}\nu(dx)\right) \tag{1.25}$$
$$+ \int_0^t \int_{|x|\leq\varepsilon} x \left(\mathcal{P}(ds,dx) - \nu(dx)\right).$$

The first line is well defined. The second line satisfies

$$\mathbb{E}\int_0^t \int_{|x|\leq\varepsilon} x \left(\mathcal{P}(ds,dx) - \nu(dx)\right) = 0, \tag{1.26}$$

and

$$\mathbb{E}\left(\int_0^t \int_{|x|\leq\varepsilon} x \left(\mathcal{P}(ds,dx) - \nu(dx)\right)\right)^2 \tag{1.27}$$
$$= \int_0^t \int_{|x|\leq\varepsilon} x^2 \lambda(ds,dx) = t\int_{|x|\leq\varepsilon} x^2\nu(dx)$$

The last expression is finite, and hence it follows that the second line in (1.25) represents a square integrable martingale, for any $0 < \varepsilon \leq 1$. The last expression tends to zero as $\varepsilon \downarrow= 0$, and hence the second line in (1.25) tends to zero in probability as $\varepsilon \downarrow 0$ (This follows from Lebesgue's dominated convergence theorem).

Since $\varepsilon$ is arbitray, we see that $X(t)$ is a finite random variable (with possibly infinite variance), and that $X(t)$ is the limit of the càdlàg processes given by the first line of (1.25). To conclude that $X(t)$ is a Lévy process we still need to show that, using a maximum inequality, the convergence of the second line to zero holds for maxima on compact sets, and that uniform limits of càdlàg functions are càdlàg functions. To make this waterproof, we will need to have closer look at the issue of weak convergence. We come back to this later.

The decomposition of a Lévy process given above with $\varepsilon = 1$ is called the Lévy-Itô decomposition.

**Markov jump processes.** Another class of Markov processes with continuous time can be constructed "explicitly" from Markov processes with discrete time. They are called Markov jump processes. The idea is simple: take a discrete time Markov process, say $Y_n$, and make it into a continuous time process by randomizing the waiting times between each move in such a way as to make the resulting process Markovian.

Let us be more precise. Let $Y_n$, $Y_n \in \mathcal{S}$, $n \in \mathbb{N}$, be some discrete time Markov process with transition kernel $P$ and initial distribution $\mu$. Let

$m(x) : \mathcal{S} \to \mathbb{R}_+$ be a uniformly bounded, measurable function. Let $e_{i,x}$, $i \in \mathbb{N}, x \in \mathcal{S}$, be a family of independent exponential random variables with mean $m(x)$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as $Y_n$, and let $Y_n$ and the $e_x$ be mutually independent. Then define the process

$$S(n) \equiv \sum_{i=0}^{n-1} e_{i,Y_i}. \tag{1.28}$$

$S(n)$ is called a *clock process*. It is supposed to represent the time at which the $n$-th jump is to take place. We define the inverse function

$$S^{-1}(t) \equiv \sup \{n : S(n) \leq t\}. \tag{1.29}$$

Then set

$$X(t) \equiv Y_{S^{-1}(t)}. \tag{1.30}$$

**Theorem 1.3.8** *The process $X(t)$ defined through (1.30) is a continuous time Markov process with càdlàg paths.*

*Proof.* Exercise. □

## 1.4 Doob's regularity theorem

We will now show that the setting of càdlàg functions is in fact suitable for the theory of martingales.

**Theorem 1.4.9** *Let $(Y_t, t \in \mathbb{R}_+)$ be a supermartingale defined on a filtered space $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t, t \in \mathbb{R}_+))$. Define the set*

$$G \equiv \{\omega \in \Omega : \text{the map}\, \mathbb{Q}_+ \ni q \to Y_q(\omega) \in \mathbb{R} \text{ is regularisable}\}. \tag{1.31}$$

*Then $G \in \mathcal{G}$ and $\mathbb{P}(G) = 1$. The process $X$ defined by*

$$X_t(\omega) \equiv \begin{cases} \lim_{q \downarrow t} Y_q(\omega), & \text{if } \omega \in G, \\ 0, & \text{else} \end{cases} \tag{1.32}$$

*is a càdlàg process.*

*Proof.* The proof makes use of our observations in Theorem 1.1.1. There are only countably many triples $(N, a, b)$ with $N \in \mathbb{N}$, $a < b \in \mathbb{Q}$. Thus in view of Theorem 1.1.1, we must show that with probability one,

$$\sup_{q \in \mathbb{Q} \cap [0,N]} |Y_q| < \infty, \tag{1.33}$$

and

$$U_N([a,b];Y|_{\mathbb{Q}}) < \infty, \tag{1.34}$$

where $Y|_{\mathbb{Q}}$ denotes the restriction of $Y$ to the rational numbers.

To do this, we will use discrete time approximations of $Y$. Let $D(m) \subset \mathbb{Q} \cap [0,N]$ be an increasing sequence of finite subsets of $\mathbb{Q}$ converging to $\mathbb{Q} \cap [0,N]$ as $m \uparrow \infty$. Then

$$\begin{aligned} \mathbb{P}\left[\sup_{q\in\mathbb{Q}\cap[0,N]} |Y_q| > 3c\right] &= \lim_{m\uparrow\infty} \mathbb{P}\left[\sup_{q\in D(m)} |Y_q| > 3c\right] \\ &\leq c^{-1}\left(4\mathbb{E}|Y_0| + 3\mathbb{E}|Y_N|\right), \end{aligned} \tag{1.35}$$

by Lemma 4.4.15 in [2]. Taking $c \uparrow \infty$ (1.33) follows. Note that we used the uniformity of the maximum inequality in the number of steps!

Similarly, using the upcrossing estimate of Theorem 4.2.2 in [2], we get that

$$\mathbb{E}\left[U_N([a,b];Y|_{\mathbb{Q}}\right] = \lim_{m\uparrow\infty} \mathbb{E}\left[U_N([a,b];Y|_{D(m)})\right] < \infty \leq \frac{\mathbb{E}|Y_N| + |a|}{b-a}, \tag{1.36}$$

uniformly in $m$, and so (1.34) also follows.

Now Theorem 1.1.1 implies the asserted result. □

We may think that Theorem 1.4.9 solves all problems related to continuous time martingales. Simply start with any supermartingale and then pass to the càdlàg regularization. However, a problem of measurability arises. This can be seen in the most trivial example of a process with a single jump. Let $Y_t$ be defined for any $\omega \in \Omega$ as

$$Y_t(\omega) = \begin{cases} 0, & \text{if } t \leq 1, \\ q(\omega), & \text{if } t > 1, \end{cases} \tag{1.37}$$

where $\mathbb{E}q = 0$. Let $\mathcal{G}_t$ be the natural filtration associated to this process. Clearly, for $t \leq 1$, $\mathcal{G}_t = \{\emptyset, \Omega\}$. $Y_t$ is a martingale with respect to this filtration. The càdlàg version of this process is

$$X_t(\omega) = \begin{cases} 0, & \text{if } t < 1, \\ q(\omega), & \text{if } t \geq 1, \end{cases} \tag{1.38}$$

Now first, $X_t$ is not adapted to the filtration $\mathcal{G}_t$, since $X_1$ is not measurable with respect to $\mathcal{G}_1$. This problem can also not be remedied by a simple modification on sets of measure zero, since $\mathbb{P}[X_1 = Y_1] < 1$. In particular, $X_t$ is not a martingale with respect to the filtration $\mathcal{G}_t$, since

$$\mathbb{E}[X_{1+\varepsilon}|\mathcal{G}_1] = 0 \neq X_1.$$

We see that the right-continuous regularization of $Y$ at the point of the jump anticipates information from the future. If we want to develop our theory on càdlàg processes, we must take this into account and introduce a richer filtration that contains this information.

**Definition 1.4.1** Let $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t, t \in \mathbb{R}_+))$ be a filtered space. Define, for any $t \in \mathbb{R}_+$,

$$\mathcal{G}_{t+} \equiv \bigcap_{s>t} \mathcal{G}_s = \bigcap_{\mathbb{Q} \ni q > t} \mathcal{G}_q \tag{1.39}$$

and let

$$\mathcal{N}(\mathcal{G}_\infty) \equiv \{G \in \mathcal{G}_\infty : \mathbb{P}[G] \in \{0,1\}\}. \tag{1.40}$$

Then the *partial augmentation*, $(\mathcal{H}_t, t \in \mathbb{R}_+)$, of the filtration $\mathcal{G}_t$ is defined as

$$\mathcal{H}_t \equiv \sigma(\mathcal{G}_{t+}, \mathcal{N}(\mathcal{G}_\infty)). \tag{1.41}$$

The following lemma, which is obvious from the construction of càdlàg versions, justifies this definition.

**Lemma 1.4.10** *If $Y_t$ is a supermartingale with respect to the filtration $\mathcal{G}_t$, and $X_t$ is its càdlàg version defined in Theorem 1.4.9, then $X_t$ is adapted to the partially augmented filtration $\mathcal{H}_t$.*

The natural question is whether in this setting $X_t$ is a supermartingale. The next theorem answers this question and is to be seen as the completion of Theorem 1.4.9

**Theorem 1.4.11** *With the assumptions and notations of Lemma 1.4.10, the process $X_t$ is a supermartingale with respect to the filtrations $\mathcal{H}_t$. Moreover, $X$ is a modification of $Y$ if and only if $Y$ is right-continuous in the sense that, for every $t \in \mathbb{R}_+$,*

$$\lim_{s \downarrow t} \mathbb{E}|Y_t - Y_s| = 0. \tag{1.42}$$

*Proof.* This is now pretty straight-forward. Fix $s > t$, and take a decreasing sequence, $s > q(n) \in \mathbb{Q}$, of rational points converging to $t$. Then

$$\mathbb{E}[Y_s|\mathcal{G}_{q(n)}] \leq Y_{q(n)}.$$

By the Lévy-Doob downward theorem (Theorem 4.2.9 in [2]),

$$\mathbb{E}[Y_s|\mathcal{G}_{t+}] = \lim_{n\uparrow\infty} \mathbb{E}[Y_s|\mathcal{G}_{q(n)}] \leq \lim_{q\downarrow t} Y_q = X_t.$$

Thus

$$\mathbb{E}[Y_s|\mathcal{H}_t] \le X_t.$$

Next take $u \ge t$ and $q(n) \downarrow u$. Then

$$\mathbb{E}[Y_{q(n)}|\mathcal{H}_t] \le X_t.$$

On the other hand, Lemma 1.2.4 and Theorem 1.4.9, $Y_{q(n)} \to X_u$ in $\mathcal{L}^1$, so

$$\mathbb{E}[X_u|\mathcal{H}_t] = \lim_{n\uparrow\infty} \mathbb{E}[Y_{q(n)}|\mathcal{H}_t] \le X_t.$$

Hence $X$ is a supermartingale with respect to $\mathcal{H}_t$.

The last statement is obvious since

$$\lim_{s\downarrow t} \mathbb{E}|Y_t - Y_s| = \lim_{s\downarrow t} \mathbb{E}|Y_t - X_t + X_t - Y_s| = \mathbb{E}|Y_t - X_t|.$$

□

With the partial augmentation we have found the proper setting for martingale theory. Henceforth we will work on filtered spaces that are already partially augmented, that is our standard setting (called the *usual setting* in [13]) is as as follows:

**Definition 1.4.2** A filtered càdlàg space is a quadruple $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}))$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{F}_t$ is a filtration of $\mathcal{F}$ that satisfies the following properties:

(i) $\mathcal{F}$ is $\mathbb{P}$-complete (contains sets of outer-$\mathbb{P}$ measure zero).
(ii) $\mathcal{F}_0$ contains all sets of $\mathbb{P}$-measure 0.
(iii) $\mathcal{F}_t = \mathcal{F}_{t+}$, i.e. $\mathcal{F}_t$ is right-continuous.

If $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t, t \in \mathbb{R}_+))$ is a filtered space, then the the minimal enlargement of this space, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$ that satisfies the conditions (i),(ii),(iii) is called the right-continuous regularization of this space.

On these spaces everything is now nice.

The following lemma details how a right-continuous regularization is achieved.

**Lemma 1.4.12** *If* $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t, t \in \mathbb{R}_+))$ *is filtered space, and* $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$ *its right-continuous regularization, then*

*(i)* $\mathcal{F}$ *is the* $\mathbb{P}$*-completion of* $\mathcal{G}$ *(i.e. the smallest* $\sigma$*-algebra containing* $\mathcal{G}$ *and all sets of* $\mathbb{P}$*-outer measure zero;*

*(ii) If $\mathcal{N}$ denotes the set of all $\mathbb{P}$-null sets in $\mathcal{F}$, then*

$$\mathcal{F}_t \equiv \bigcap_{u>t} \sigma(\mathcal{G}_u, \mathcal{N}) = \sigma(\mathcal{G}_{t+}, \mathcal{N}); \tag{1.43}$$

*(iii) If $F \in \mathcal{F}_t$, then there exists $G \in \mathcal{G}_{t+}$ such that*

$$F \Delta G \in \mathcal{N}, \tag{1.44}$$

*where $F\Delta G$ denotes the symmetric difference of the sets $F$ and $G$.*

*Proof.* Exercise. □

**Proposition 1.4.13** *The process $X$ constructed in Theorem 1.4.9 is a supermartingale with respect to the filtration $\mathcal{F}_t$.*

*Proof.* Since by (1.44) $\mathcal{F}_t$ and $\mathcal{H}_t$ differ only by sets of measure zero, $\mathbb{E}(X_{t+s}|\mathcal{F}_t)$ and $\mathbb{E}(X_{t+s}|\mathcal{H}_t)$ differ only on null sets and thus are versions of the same conditional expectation. □

We can now give a version of Doob's regularity theorem for processes defined on càdlàg spaces.

**Theorem 1.4.14** *Let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$ be a filtered càdlàg space. Let $Y$ be an adapted supermartingale. Then $Y$ has a càdlàg modification, $Z$, if and only if the map $t \to \mathbb{E}Y_t$ is right-continuous, in which case $Z$ is a càdlàg supermartingale.*

*Proof.* Since $Y$ is a supermartingale, for any $u \geq t$, $\mathbb{E}(Y_u)|\mathcal{F}_t) \leq Y_t$, a.s.. Construct the process $X$ as in Theorem 1.4.9 Then

$$\mathbb{E}(X_t|\mathcal{F}_t) = \mathbb{E}\left(\lim_{u\downarrow t} Y_u|\mathcal{F}_t\right) = \lim_{u\downarrow t} \mathbb{E}\left(Y_u|\mathcal{F}_t\right) \leq Y_t, \text{ a.s..} \tag{1.45}$$

since $Y_u \downarrow Y_t$ in $\mathcal{L}^1$. Since $X_t$ is adapted to $\mathcal{F}_t$, this implies $X_t \leq Y_t$, a.s..

If now $\mathbb{E}(Y_t)$ is right-continuous, then $\lim_{u\downarrow t} \mathbb{E}Y_u = \mathbb{E}Y_t$, while from the $\mathcal{L}^1$-convergence of $Y_u$ to $X_t$, we get $\mathbb{E}X_t = \lim_{u\downarrow t} \mathbb{E}Y_u = \mathbb{E}Y_t$. Hence $\mathbb{E}X_t = \mathbb{E}Y_t$, and so, since already $X_t \leq Y_t$, a.s., $X_t = Y_t$, a.s., i.e. $X_t$ is the càdlàg modification of $Y$. If, on the other hand, $\mathbb{E}Y_t$ fails to be right-continuous at some point $t$, then it follows that $X_t < Y_t$ with positive probability, and so the càdlàg process $X_t$ is not a modification of $Y$. □

### 1.5 Convergence theorems and martingale inequalities

Key results on discrete time martingale theory were Doob's forward and backward convergence theorems and the maximum inequalities. We will now consider the corresponding results in continuous time. This will not be very hard.

**Theorem 1.5.15** *Let $X$ be a càdlàg supermartingale with respect to a filtered space $(\Omega, \mathcal{G}, \mathbb{P}, \mathcal{G}_t)$ and assume that* $\sup_t \mathbb{E}|X_t| < \infty$. *Then*

$$\lim_{t\uparrow\infty} X_t \equiv X_\infty \in \mathbb{R}, \tag{1.46}$$

*exists almost surely.*

*Proof.* A càdlàg function is determined by its values on the rational numbers. Thus $X_t$ will converge if and only if $X_q$, $q \in \mathbb{Q}$ does. Therefore, the proof of our theorem can be reduced to proving the same fact for the restriction of $X$ to the rationals, and all arguments of the discrete time case simply carry over. □

Similarly, one obtains the corresponding uniform integrability results.

**Theorem 1.5.16** *Let $X$ be as in the previous theorem. Then*

*(i) if $X$ is uniformly integrable, then $X_t \to X_\infty$ in $\mathcal{L}^1$, and for any $t$,* $\mathbb{E}(X_\infty|\mathcal{G}_t) \leq X_t$, a.s., *with equality in the martingale case;*
*(ii) If $X$ is a martingale (or a supermartingale that is bounded from above) and $X_t \to X_\infty$ in $\mathcal{L}^1$, then $X$ is uniformly integrable.*

*Proof.* The proof of the first statement uses Theorem 1.4.15 from [1] (Vitali's theorem) which implies that uniform integrability and a.s. convergence implies convergence in $\mathcal{L}^1$ along any discrete subsequence $t_n$. If $X$ is a martingale that converges in $\mathcal{L}^1$, then $X_t = \mathbb{E}(X_\infty|\mathcal{G}_t)$ a.s., and so $X_t$ is a family of conditional expectations, and hence uniformly integrable by Theorem 4.2.6 of [1]. If $X$ is a supermartingale and bounded from above by a constant $C < \infty$, then $C \geq X_t \geq \mathbb{E}(X_\infty|\mathcal{G}_t)$ where the lower bound is uniformly integrable. This implies that $X_t$ is itself uniformly integrable. □

**Remark 1.5.1** In general, (ii) does not hold for supermartingales without further assumptions. This is different from the discrete time case, where Vitali's theorem yields uniform integrability of $\mathcal{L}^1$-convergent supermartingales.

Finally we have an analog of the downward theorem, with a slightly different twist:

**Theorem 1.5.17** *Suppose we have a càdlàg supermartingale as before but on the parameter space* $(0,\infty)$. *Assume that* $\sup_{t>0} \mathbb{E}X_t < \infty$. *Then*

$$X_{0+} \equiv \lim_{t\downarrow 0} X_t$$

*exists a.s. and in* $\mathcal{L}^1$. *Moreover,* $\mathbb{E}(X_t|\mathcal{G}_{0+}) \leq X_{0+}$, a.s..

Again the proof is virtually the same as in the discrete case and will not be given.

In a similar way the maximum inequalities for càdlàg submartingales can be inferred from the discrete ones.

**Theorem 1.5.18** *Let* $Z$ *be a non-negative càdlàg submartingale on a filtered space. Then, for any* $c > 0$ *and* $t \geq 0$,

$$\mathbb{P}\left(\sup_{s\leq t} Z_s \geq c\right) \leq c^{-1}\mathbb{E}\left(Z_t \mathbb{1}_{\sup_{s\leq t} Z_s \geq c}\right) \leq c^{-1}\mathbb{E}Z_t. \tag{1.47}$$

*Proof.* The proof contains some basic ideas how to control suprema over uncountable sets and is thus instructive. Consider an increasing sequence, $D(m)$, of finite subsets of $[0,t]$ containing each 0 and $t$ such that $D \equiv \cup_m D(m)$ is dense in $[0,1]$. Then, since $Z$ is càdlàg

$$\sup_{s\in[0,t]} Z_s(\omega) = \sup_m \sup_{s\in D(m)} Z_s(\omega). \tag{1.48}$$

Thus,

$$\left\{\omega : \sup_{s\in[0,t]} Z_s(\omega) \geq c\right\} = \lim_m \left\{\omega : \sup_{s\in D(m)} Z_s(\omega) \geq c\right\}.$$

Now use the discrete time submartingale inequality to see that

$$\begin{aligned}\mathbb{P}\left(\sup_{s\in[0,t]} Z_s \geq c\right) &= \lim_m \mathbb{P}\left(\sup_{s\in D(m)} Z_s \geq c\right)\\ &\leq \lim_m c^{-1}\mathbb{E}\left(Z_t \mathbb{1}_{\sup_{s\in D(m)} Z_s \geq c}\right)\\ &= c^{-1}\mathbb{E}\left(Z_t \mathbb{1}_{\sup_{s\in[0,t]} Z_s \geq c}\right)\end{aligned}$$

□

Finally we state the continuous time analog of Doob's $\mathcal{L}^p$ inequality.

**Theorem 1.5.19** *Let $1/p + 1/q = 1$, $p > 1$. Let $Z$ be a non-negative càdlàg sub-martingale on a filtered space $(\Omega, \mathcal{G}.\mathbb{P}, \mathcal{G}_t)$ such that $\mathbb{E}Z_t^p < \infty$, uniformly in $t \in \mathbb{R}$. Let $Z^* \equiv \sup_{t\geq 0} Z_t$. Then*

$$\|Z^*\|_p \leq q \sup_{t\in\mathbb{R}_+} \|Z_t\|_p. \tag{1.49}$$

*Then $Z_\infty \equiv \lim_{t\uparrow\infty} Z_t$ exists a.s. and in $\mathcal{L}^p$, and*

$$\|\mathbb{Z}_\infty\|_p = \sup_{t\in\mathbb{R}_+} \|Z_t\|_p = \lim_{t\uparrow\infty} \|Z_t\|_p. \tag{1.50}$$

*In the case when $Z$ is a martingale, then $Z_t = \mathbb{E}(Z_\infty|\mathcal{G}_t)$, a.s. .*

*Proof.* Compare to Theorem 4.3.13 in [2] and adopt the proof to the continuous setting. □

## 1.6 Brownian motion revisited

We have shown in [2] Theorem 6.2.1 that Brownian motion exists through an explicite construction. We now want to show an alternative that starts with the "pre-Brownian" motion that we had defined without reference to continuity of paths.

Now we want to take that process, show first that it can be regularized to define a càdlàg martingale and show than that it has almost surely continuous paths.

We consider the Gaussian stochastic process $Y_t$ defined on a filtered space $(\Omega, \mathcal{G}, \mathbb{P}, \mathcal{G}_t)$ with covariance $\mathbb{E}Y_tY_s = s \wedge t$; we have seen that $Y_t - Y_s$, for $t > s$, is independent of the $\sigma$-algebra $\mathcal{G}_s$ and $\mathbb{E}(Y_t - Y_s|\mathcal{G}_s) = 0$ so that $Y_s$ is a martingale with respect to the filtration $\mathcal{G}_t$. Since $\mathbb{E}|Y_t - Y_s| \leq \sqrt{\mathbb{E}(Y_t - Y_s)^2} = \sqrt{t-s}$ tends to zero as $t \downarrow s$, the assumption of Theorem 1.4.11 are verified and there exists a càdlàg modification, $X$, of $Y$, that is a càdlàg martingale relative to the usual augmentation $\mathcal{F}_t$.

It is not entirely trivial that this modification will have the desired independence properties of Brownian motion, but the following lemma shows why it does.

**Lemma 1.6.20** *With the notation above,*

*(i) For $t \geq 0$, the $\sigma$-algebra $\mathcal{U}_t \equiv \sigma(Y_{t+u} - Y_t, u \in \mathbb{R}_+\}$, is independent of $\mathcal{G}_{t+}$.*

*(ii) For $t \geq 0$, $\mathcal{G}_{t+} \subset \sigma(\mathcal{G}_t, \mathcal{N}(\mathcal{G}_\infty))$, where $\mathcal{N}(\mathcal{G}_\infty)$ denotes the $\mathbb{P}$-null sets in $\mathcal{G}_\infty$.*

*Proof.* First it is clear that the càdlàg modification of $Y$, satisfies that $X_{t+u+\varepsilon} - X_{t+\varepsilon}$ is independent of $\mathcal{G}_{t+\varepsilon/2}$ and hence of $\mathcal{G}_{t+}$. Thus, for $G \in \mathcal{G}_{t+}$,

$$\mathbb{E}\left(f(X_{t+u+\varepsilon} - X_{t+\varepsilon})\mathbb{1}_G\right) = \mathbb{P}(G)\mathbb{E}\left(f(X_{t+u+\varepsilon} - X_{t+\varepsilon})\right),$$

for any bounded continuous function. Since $X$ is right-continuous, bounded convergence shows that

$$\mathbb{E}\left(f(X_{t+u} - X_t)\mathbb{1}_G\right) = \mathbb{P}(G)\mathbb{E}\left(f(X_{t+u} - X_t)\right),$$

Then the monotone class theorem shows that $X_{t+u} - X_t$ is independent of $\mathcal{G}_{t+}$. Since $Y$ is a modification of $X$, the same holds for $Y_{t+u} - Y_t$.

Next, let $\eta$ be $\mathcal{G}_{t+}$-measurable and let $\xi = \eta - \mathbb{E}(\eta|\mathcal{G}_t)$ almost surely. We want to show that $\xi = 0$ a.s. . We know that $\xi$ is independent of $\mathcal{U}_t$, so for any $G_t \in \mathcal{G}_t$ and $A_t \in \mathcal{U}_t$,

$$\mathbb{E}(\xi \mathbb{1}_{G_t}\mathbb{1}_{A_t}) = \mathbb{P}(A_t)\mathbb{E}(\eta\mathbb{1}_{G_t}) - \mathbb{P}(A_t)\mathbb{P}(G_t)\mathbb{E}(\eta|G_t) = 0, \text{ a.s.} \tag{1.51}$$

as desired. Now events of the form $A_t \cap G_t$ form a $\pi$-system that generates the $\sigma$-algebra $\mathcal{G}_\infty = \sigma(\mathcal{U}_t, \mathcal{G}_t)$. Thus (1.51) shows $\xi = 0$ a.s.. □

By definition of the augmentation $\mathcal{F}_t$, the statements of the lemma can also be read as

(i) For $t \geq 0$, the $\sigma$-algebra $\sigma(X_{t+u} - X_t), u \in \mathbb{R}_+\}$, is independent of $\mathcal{F}_t$.
(ii) For $t \geq 0$, $\mathcal{F}_t = \sigma(\mathcal{G}_t, \mathcal{N}(\mathcal{G}_\infty))$, where $\mathcal{N}$ denotes the $\mathbb{P}$-null sets in $\mathcal{F}$.

Therefore, $X_t$ as a càdlàg martingale satisfies the properties required of Brownian motion, except so far the continuity of paths. We will now show that this also holds.

**Theorem 1.6.21** *$\mathbb{P}$-almost all paths of $X$ are continuous.*

*Proof.* The process $X^4$ is a càdlàg submartingale, since by Jensen's inequality

$$\mathbb{E}(X_t^4|\mathcal{F}_s) = \mathbb{E}\left((X_t - X_s + X_s)^4|\mathcal{F}_s\right) \leq \left(\mathbb{E}(X_t - X_s + X_s|\mathcal{F}_s)\right)^4 = X_s^4.$$

Hence

$$\mathbb{P}\left(\sup_{s\leq\delta}|X_s| > \varepsilon\right) = \mathbb{P}\left(\sup_{s\leq\delta}|X_s^4| > \varepsilon^4\right) \leq \varepsilon^{-4}\mathbb{E}X_\delta^4 = 3\varepsilon^{-4}\delta^2.$$

Put

$$D_n \equiv \{k2^{-n} : 0 \leq k < 2^n\},$$

and $\delta_n \equiv 2^{-n}$. Then

$$\mathbb{P}\left(\sup_{r\in D_n}\sup_{s\leq\delta_n}|X_{r+s}-X_r|>1/n\right)\leq 2^n\mathbb{P}\left(\sup_{s\leq\delta_n}|X_s-X_0|>1/n\right)$$
$$\leq 32^n\delta_n^2 n^4 = 3n^4 2^{-n}.$$

The right-hand side is summable over $n$, and so the first Borel-Cantelli lemma implies that on a set of probability one, for all except finitely many values of $n$,

$$\sup_{r\in D_n}\sup_{s\leq\delta_n}|X_{r+s}-X_r|\leq 1/n$$

and so

$$\sup_{r\in[0,1]}\sup_{s\leq\delta_n}|X_{r+s}-X_r|\leq 3/n$$

which implies uniform continuity of all paths on a set of measure one. Then it suffices to modify the process on a set of measure zero to obtain Brownian motion. □

## 1.7 Stopping times

The notions around stopping times that we will introduce in this section will be very important in the sequel, in particular also in the theory of Markov processes. We have to be quite a bit more careful now in the continuous time setting, event though we would like to have everything resemble the discrete time setting.

We consider a filtered space $(\Omega,\mathcal{G}:\mathbb{P},(\mathcal{G}_t,t\in\mathbb{R}_+))$.

**Definition 1.7.1** A map $T:\Omega\to[0,\infty]$ is called a $\mathcal{G}_t$-stopping time if

$$\{T\leq t\}\equiv\{\omega\in\Omega: T(\omega)\leq t\}\in\mathcal{G}_t,\forall t\leq\infty. \tag{1.52}$$

If $T$ is a stopping time, then the *pre-$T$-$\sigma$-algebra*, $\mathcal{G}_T$, is the set of all $\Lambda\in\mathcal{G}$ such that

$$\Lambda\cap\{T\leq t\}\in\mathcal{G}_t,\forall t\leq\infty. \tag{1.53}$$

With this definition we have all the usual elementary properties of pre-$T$-$\sigma$-algebras:

**Lemma 1.7.22** *Let $S,T$ be stopping times. Then:*

*(i) If $S\leq T$, then $\mathcal{G}_S\subset\mathcal{G}_T$.*
*(ii) $\mathcal{G}_{T\wedge S}=\mathcal{G}_T\cap\mathcal{G}_S$.*
*(iii) If $F\in\mathcal{G}_{S\vee T}$, then $F\cap\{S\leq T\}\in\mathcal{G}_T$.*

*(iv)* $\mathcal{G}_{S\vee T} = \sigma(\mathcal{G}_T, \mathcal{G}_S)$.

*Proof.* Exercise. □

It will be useful to talk also about stopping time with respect to the filtrations $\mathcal{G}_{t+}$.

**Definition 1.7.2** A map $T : \Omega \to [0, \infty]$ is called a $\mathcal{G}_{t+}$-stopping time if

$$\{T < t\} \equiv \{\omega \in \Omega : T(\omega) < t\} \in \mathcal{G}_t, \forall t \leq \infty. \tag{1.54}$$

If $T$ is a $\mathcal{G}_{t+}$-stopping time, then the *pre-$T$-$\sigma$-algebra*, $\mathcal{G}_{T+}$, is the set of all $\Lambda \in \mathcal{G}$ such that

$$\Lambda \cap \{T < t\} \in \mathcal{G}_t, \forall t \leq \infty. \tag{1.55}$$

**Lemma 1.7.23** *Let $S_n$ be a sequence of $\mathcal{G}_t$-stopping times. Then:*

*(i) if $S_n \uparrow S$, then $S$ is a $\mathcal{G}_t$ stopping time;*
*(ii) if $S_n \downarrow S$, then $S$ is a $\mathcal{G}_{t+}$-stopping time and $\mathcal{G}_{S+} = \bigcap_{n\in\mathbb{N}} \mathcal{G}_{S_n+}$.*

*Proof.* Consider case (i). Since $S_n$ is increasing, the sequence of sets $\{S_n \leq t\} \in \mathcal{G}_t$ is decreasing, and its limit is also in $\mathcal{G}_t$. In case (ii), since if $S_n \downarrow S$, $\{S < t\}$ contains all sets $\{S_n < t\}$. On the other hand, for any $\varepsilon > 0$, there exists $n_0 < \infty$, such that $\{S \leq t - \varepsilon\} \subset \{S_n < t\}$ for all $n \geq n_0$. Hence the event $\{S < t\}$ is contained in $\bigcup_n \{S_n \leq t\}$, and by the previous observation, $\{S < t\} = \bigcup_n \{S_n \leq t\} \in \mathcal{G}_t$. □

**Definition 1.7.3** A process $X_t, t \in \mathbb{R}_+$ is called *$\mathcal{G}_t$-progressive* if, for every $t \geq 0$, the restriction of the map $(s, \omega) \to X_s(\omega)$ to $[0, t] \times \Omega$ is $\mathcal{B}([0, t] \times \mathcal{G}_t$-measurable.

The notion of a progressive process is stronger than that of an adapted process. The importance of the notion of progressiveness arises from the fact that $T$-stopped progressive processes are measurable with respect to the respective pre-$T$ $\sigma$-algebra.

The good news is that in the usual càdlàg world we need not worry:

**Lemma 1.7.24** *An adapted càdlàg process with values in a metrisable space, $(S, \mathcal{B}(S))$, is progressive.*

*Proof.* The whole idea is to approximate the process by a piecewise constant one, to use that this is progressive, and then to pass to the limit. To do this, fix $t$ and set, for $s < t$, (we will always understand $X(s) = X_s$)

$$X^n(s, \omega) \equiv X((k+1)2^{-n}t, \omega), \quad \text{if } k2^{-n}t \leq s < [k+1]2^{-n}t.$$

For $n$ fixed, checking measurability of the map $X^n$ involves the inspection of only finitely many time points, i.e.

$$\begin{aligned}(X^n)^{-1}(B) &= \{(\omega, s) \in \Omega \times [0,t] : X^n(s,\omega) \in B\} \\ &= \{(\omega, s) \in \Omega \times [0,t] : X^n(k(s)2^{-n}t, \omega) \in B\}\end{aligned}$$

where $k(s) = \max\{k \in \mathbb{N} : k2^{-n}t \leq s\}$. The latter set is clearly measurable.

Finally, $X^n$ converges pointwise to $X$ on $[0,t]$, and so $X$ shares the same measurability properties. □

***Exercise:*** *Show why the right-continuity of paths is important. Can you find an example of an adapted process that is not progressive?*

**Lemma 1.7.25** *If $X$ is progressive with respect to the filtration $\mathcal{G}_t$ and $T$ is a $\mathcal{G}_t$-stopping time, then $X_T$ is $\mathcal{G}_T$ measurable.*

*Proof.* For $t \geq 0$ let $\widehat{\Omega}_t \equiv \{\omega : T(\omega) \leq t\}$. Define $\widehat{\mathcal{G}}_t$ to be the sub-$\sigma$-algebra of $\mathcal{G}_t$ such that any set $A \in \widehat{\mathcal{G}}_t$ is in $\widehat{\Omega}_t$. Let $\rho : \widehat{\Omega}_t \to [0,t] \times \widehat{\Omega}_t$ be defined by

$$\rho(\omega) \equiv (T(\omega), \omega).$$

Define further the map $\widehat{X}_t : [0,t] \times \widehat{\Omega}_t \to S$ by

$$\widehat{X}_t(s, \omega) \equiv X_s(\omega).$$

Note that the map $\widehat{X}_t$ is measurable with respect to $\mathcal{B}([0,t]) \times \mathcal{G}_t$ due to the progressiveness of $X$. $\rho$ is measurable with respect to $\mathcal{G}_t$ by the definition of stopping times and the obvious measurability of the identity map. Hence $\widehat{X}_t \circ \rho$ as map from $\widehat{\Omega}_t \to S$ is $\mathcal{G}_t$- measurable.

Then we can write, for $\omega \in \widehat{\Omega}_t$, $X_T(\omega) = \widehat{X}_t \circ \rho(\omega)$, and hence, for any Borel set $\Gamma$

$$\begin{aligned}\{\omega \in \Omega : X_T(\omega) \in \Gamma\} \cap \{T \leq t\} &= \{\omega \in \widehat{\Omega}_t : X_T(\omega) \in \Gamma\} \\ &= (\widehat{X}_t \circ \rho)^{-1}(\Gamma) \in \widehat{\mathcal{G}}_t \subset \mathcal{G},\end{aligned}$$

which proves the measurability of $X_T$. □

## 1.8 Entrance and hitting times

Already in the case of discrete time Markov processes we have seen that the notion of hitting times of certain sets provides particularly important examples of stopping times. We will here extend this discussion to the continuous time case. It is quite important to distinguish two notions of hitting and first entrance time. They differ in the way the position of the process at time 0 is treated.

**Definition 1.8.1** Let $X$ be a stochastic process with values in a measurable space $(E, \mathcal{E})$. Let $\Gamma \in \mathcal{E}$. We call

$$\tau_\Gamma(\omega) \equiv \inf\{t > 0 : X_t(\omega) \in \Gamma\} \tag{1.56}$$

the *first hitting time* of the set $\Gamma$; we call

$$\Delta_\Gamma(\omega) \equiv \inf\{t \geq 0 : X_t(\omega) \in \Gamma\} \tag{1.57}$$

the *first entrance time* of the set $\Gamma$. In both cases we infimum is understood to yield $+\infty$ if the process never enters $\Gamma$.

Recall that in the discrete time case we have only worked with $\tau_\Gamma$, which is in fact the more important notion.

We will now investigate cases when these times are stopping times.

**Lemma 1.8.26** *Consider the case when $E$ is a metric space and let $F$ be a closed set. Let $X$ be a continuous adapted process. Then $\Delta_F$ is a $\mathcal{G}_t$-stopping time and $\tau_F$ is a $\mathcal{G}_{t+}$-stopping time.*

*Proof.* Let $\rho$ denote the metric on $E$. Then the map $x \to \rho(x, F)$ is continuous, and hence the map $\omega \to \rho(X_q(\omega), x)$ is $\mathcal{G}_q$ measurable, for $q \in \mathbb{Q}_+$. Since the paths $X_t(\omega)$ are continuous, $\Delta_F(\omega) \leq t$ if and only if

$$\inf_{q \in \mathbb{Q} \cap [0,t]} \{\rho(X_q(\omega), F)\} = 0.$$

and so $\Delta_F$ is measurable w.r.t. $\mathcal{G}_t$. For $\tau_F$ the situation is slightly different at time zero. Let us define, for $r > 0$, $\Delta_F^r \equiv \inf\{t \geq r : X_t \in F\}$. Obviously, from the previous result, $D_F^r$ is a $\mathcal{G}_t$-stopping time. On the other hand, $\{\tau_F > 0\}$ if and only if there exists $\delta > 0$, such that, for all $\mathbb{Q} \ni r > 0$, $\Delta_F^r > \delta$. But clearly, the event

$$A_\delta \equiv \cap_{\mathbb{Q} \ni r > 0} \{\Delta_F^r > \delta\}$$

is $\mathcal{G}_\delta$-measurable, and so the event

$$\{\tau_F = 0\} = \{\tau_F > 0\}^c = \cap_{\delta > 0} A_\delta^c$$

is $\mathcal{G}_{0+}$-measurable and so $\tau_F$ is a $\mathcal{G}_{t+}$-stopping time. $\square$

To see where the difference in the two times comes from, consider the process starting at the boundary of $F$. Then $\Delta_F = 0$ can be deduced from just that knowledge. On the other hand, $\tau_F$ may or may not be zero: it could be that the process leaves $F$ and only returns after some time $t$, or it may stay a little while in $F$, in which case $\tau_F = 0$; to distinguish the two cases, we must look a little bit into the future!

## 1.9 Optional stopping and optional sampling

We have seen the theory of discrete time Markov processes that martingale properties of processes stopped at stopping times are important. We want to recover such results for càdlàg processes.

In the sequel we will work on a filtered càdlàg space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$ on which all processes will be defined and adapted.

Our aim is the following *optional sampling theorem*:

**Theorem 1.9.27** *Let $X$ be a càdlàg submartingale and let $T, S$ be $\mathcal{F}_t$-stopping times. Then for each $M < \infty$,*

$$\mathbb{E}\left(X(T \wedge M) | \mathcal{F}_S\right) \geq X(S \wedge T \wedge M), \text{ a.s..} \tag{1.58}$$

*If, in addition,*

*(i)* *$T$ is finite a.s.,*
*(ii)* *$\mathbb{E}|X(T)| < \infty$, and*
*(iii)* $\lim_{M \uparrow \infty} \mathbb{E}\left(X(M) \mathbb{1}_{T > M}\right) = 0$,

*then*

$$\mathbb{E}\left(X(T) | \mathcal{F}_S\right) \geq X(S \wedge T), \text{ a.s..} \tag{1.59}$$

*Equality holds in the case of martingales.*

*Proof.* In order to prove Theorem 1.9.27 we frst prove a result for stopping times taking finitely many values.

**Lemma 1.9.28** *Let $S, T$ be $\mathcal{F}_t$ stopping times that take only values in the set $\{t_1, \dots, t_m\}$, $0 \leq t_1 < \dots < t_m \leq \infty$. If $X$ is a $\mathcal{F}_t$-submartingale, then*

$$\mathbb{E}\left(X(T) | \mathcal{F}_S\right) \geq X(S \wedge T), \text{ a.s..} \tag{1.60}$$

*Proof.* We need to prove that for any $A \in \mathcal{F}_S$,

$$\mathbb{E}\left(\mathbb{1}_A X(T)\right) \geq \mathbb{E}\left(\mathbb{1}_A X(T \wedge S)\right). \tag{1.61}$$

Now we can decompose $A = \cup_{i=1}^m A \cap \{S = t_i\}$. Hence we just have to prove (1.61) with $A$ replaced by $A \cap \{S = t_i\}$, for any $i = 1, \dots m$. Now, since $A \in \mathcal{F}_S$, we have that $A \cap \{S = t_i\} \in \mathcal{F}_{t_i}$. We will first show that

$$\mathbb{E}\left(X(T)|\mathcal{F}_{t_i}\right) \geq X(T \wedge t_i). \tag{1.62}$$

To do this, note that

$$\begin{aligned} \mathbb{E}\left(X(T \wedge t_{k+1})|\mathcal{F}_{t_k}\right) &= \mathbb{E}\left(X(t_{k+1})\mathbb{1}_{T>t_k} + X(T)\mathbb{1}_{T\leq t_k}|\mathcal{F}_{t_k}\right) \\ &= \mathbb{E}\left(X(t_{k+1})|\mathcal{F}_{t_k}\right)\mathbb{1}_{T>t_k} + X(T)\mathbb{1}_{T\leq t_k} \\ &\leq X(t_{k+1})\mathbb{1}_{T>t_k} + X(T)\mathbb{1}_{T\leq t_k} \\ &= X(t_k \wedge T), \text{ a.s..} \end{aligned} \tag{1.63}$$

Since $S = S \wedge t_m$, this gives (1.62) for $i = m - 1$. Then we can iterate (1.63) to get (1.62) for general $i$.
Using (1.61), we can now deduce that

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{A\cap\{S=t_i\}}X(T)\right) &= \mathbb{E}\left(\mathbb{1}_{A\cap\{S=t_i\}}\mathbb{E}(X(T)|\mathcal{F}_{t_i})\right) \\ &\geq \mathbb{E}\left(\mathbb{1}_A X(T \wedge t_i)\right) \\ &= \mathbb{E}\left(\mathbb{1}_A X(T \wedge S)\right) \end{aligned} \tag{1.64}$$

as desired. This concludes the proof of the lemma. □

We now continue the proof of the theorem through approximation arguments. Let $S^n = (k+1)2^{-n}$, if $S \in [k2^{-n}, (k+1)2^{-n})$, and $T^{(n)} = \infty$, if $T = \infty$; define $T^{(n)}$ in the same way. Fix $\alpha \in \mathbb{R}$ and $M > 0$. Then the preceeding lemma implies that

$$\mathbb{E}\left(X(T^{(n)} \wedge M) \vee \alpha|\mathcal{F}_{S^{(n)}}\right) \geq X(T^{(n)} \wedge S^{(n)} \wedge M) \vee \alpha, \text{ a.s..} \tag{1.65}$$

Since $\mathcal{F}_S \subset \mathcal{F}_{S^{(n)}}$, it follows that

$$\mathbb{E}\left(X(T^{(n)} \wedge M) \vee \alpha|\mathcal{F}_S\right) \geq \mathbb{E}\left(X(T^{(n)} \wedge S^{(n)} \wedge M) \vee \alpha|\mathcal{F}_S\right), \text{ a.s..} \tag{1.66}$$

Again from using Lemma 1.9.28, we get that

$$\alpha \leq X(T^{(n)} \wedge M) \vee \alpha \leq \mathbb{E}\left(X(M) \vee \alpha|\mathcal{F}_{T^{(n)}}\right), \text{ a.s.,}$$

and therefore $X(T^{(n)} \wedge M) \vee \alpha$ is uniformly integrable. Similarly $X(T^{(n)} \wedge S^{(n)} \wedge M) \vee \alpha$ is uniformly integrable. Therefore we can pass to the limit $n \uparrow \infty$ in (1.66) and obtain, using that $X$ is right-continuous,

$$\mathbb{E}\left(X(T \wedge M) \vee \alpha|\mathcal{F}_S\right) \geq \mathbb{E}\left(X(T \wedge S \wedge M) \vee \alpha|\mathcal{F}_S\right), \text{ a.s..} \tag{1.67}$$

Since this relation holds for all $\alpha$, we may let $\alpha \downarrow -\infty$ to get (1.58).

Using the additional assumptions on $T$; we can pass to the limit $M \uparrow \infty$ and get (1.59) in this case: First, the a.s. finiteness of $T$ implies that

$$\lim_{M\uparrow\infty} X(T\wedge S\wedge M) = X(T\wedge S), \text{ a.s.},$$

Do deal with the left-hand side, write

$$\begin{aligned}\mathbb{E}\left(X(T\wedge M)|\mathcal{F}_S\right) &= \mathbb{E}\left(X(T)|\mathcal{F}_S\right)\\ &\quad + \mathbb{E}\left(X(M)\mathbb{1}_{T>M}|\mathcal{F}_S\right) - \mathbb{E}\left(X(T)\mathbb{1}_{T>M}|\mathcal{F}_S\right)\end{aligned}$$

The first term in the second line converges to zero by Assumption (iii), since

$$\left|\mathbb{E}\left(X(M)\mathbb{1}_{T>M}|\mathcal{F}_S\right)\right| \le \mathbb{E}\left(|X(M)|\mathbb{1}_{T>M}|\mathcal{F}_S\right)$$

and

$$\mathbb{E}\mathbb{E}\left(|X(M)|\mathbb{1}_{T>M}|\mathcal{F}_S\right) = \mathbb{E}\left(|X(M)|\mathbb{1}_{T>M}\right) \downarrow 0.$$

The mean of the absolute value of the second term is bounded by

$$\mathbb{E}\left(|X(T)|\mathbb{1}_{T>M}\right),$$

which tends to zero by dominated convergence due to Assumtions (i) and (ii). □

A special case of the preceeding theorem implies the following corollary:

**Corollary 1.9.29** *Let $X$ be a càdlàg (super, sub )martingale, and let $T$ be a stopping time. Then $X^T \equiv X_{T\wedge t}$ is a (super, sub) martingale.*

In the case of uniformly integrable supermartingales we get Doob's optional sampling theorem:

**Theorem 1.9.30** *Let $X$ be a uniformly integrable or a non-negative càdlàg supermartingale. Let $S$ and $T$ be stopping times with $S \le T$. Then $X_T \in \mathcal{L}^1$ and*

$$\mathbb{E}\left(X_\infty|\mathcal{F}_T\right) \le X_T, \text{ a.s.} \tag{1.68}$$

*and*

$$\mathbb{E}\left(X_T|\mathcal{F}_S)\right) \le X_S, \text{ a.s.}, \tag{1.69}$$

*with equality in the uniformly integrable martingale case.*

*Proof.* The proof is along the same lines of approximation with discrete supermartingales as in the preceding theorem and uses the analogous results in discrete time (see [13], Thms (59.1,59.5)). □

# 2

# Weak convergence

In this short section we collect some necessary material for understanding the convergence of sequences of stochastic processes with path properties. This will allow us to put the analysis of the Donsker theorem into a general framework.

## 2.1 Some topology

We consider the general setup on a compact Hausdorff space, $J$. We denote by $C(J)$ the Banach space of bounded, continuous real-valued functions equipped with the supremum norm. We denote by $\mathcal{M}_1(J)$ the space of probability measures on $J$. We denote by $C(J)^*$ the space of bounded linear functionals $C(J) \to \mathbb{R}$ on $C(J)$.

We need two basic facts from functional analysis:

**Theorem 2.1.1** *[Stone-Weierstrass theorem] Let $A$ be a sub-algebra of $C(J)$ that contains constant functions and separates points of $J$, i.e. for any $x \in J$ there exits $f, g \in A$ such that $f(x) \neq g(x)$. Then $A$ is dense in $C(J)$.*

**Theorem 2.1.2** *[Riesz representation theorem] Let $\phi$ be a linear increasing functional $\phi : C(J) \to \mathbb{R}$ with $\Phi(1) = 1$. Then there exists a unique inner regular probability measure, $\mu \in \mathcal{M}_1(J)$, such that*

$$\phi(f) = \mu(f) = \int_J f d\mu. \tag{2.1}$$

Recall (see [2], page 12) that a measure is *inner regular*, if for any Borel set, $B$, $\mu(B) = \sup\{\mu(K), K \subset B, \text{compact}\}$. We have shown there already, that if $J$ is a compact metrisable space, then any probability measure on it is inner regular.

The weak-$*$ topology on the space $C(J)^*$ is obtained by choosing sets of the form

$$B_{f_1,\dots,f_n,\varepsilon}(\phi_0) \equiv \{\phi \in C(J)^* : \forall_{1\leq i\leq n} |\phi(f_i) - \phi^0(f_i)| < \varepsilon\} \tag{2.2}$$

with $n \in \mathbb{N}, \varepsilon > 0, f_i \in C(J)$ as a basis of neighborhoods. The ensuing space is a Hausdorff space.

When speaking of convergence on topological spaces, it is useful to extend the notion of convergence of sequences to that of *nets*.

**Definition 2.1.1** A *directed set*, $D$, is a partially ordered set all of whose finite subsets have an upper bound in $D$. A *net* is a family $(x_\alpha, \alpha \in D)$ indexed by a directed set.

If $(x_\alpha, \alpha \in D)$ is a net in a topological space, $E$, then $x_a \to x$ if, for every open neighborhood, $G$, of $x$, there exists $a_0 \in D$ such that for all $\alpha \geq \alpha_0$, $x_\alpha \in G$.

**Lemma 2.1.3** *A net $\phi_\alpha$ in $C(J)^*$ converges in the weak-* topology to some element, $\phi$, if and only if, for all $f \in C(J)$, $\phi_\alpha(f) \to \phi(f)$.*

*Proof.* Let us prove first the "if" part. Then for any $f$, and any $\varepsilon$, there exists $\alpha_f$, such that for all $\alpha \geq \alpha_f$, $\phi_\alpha(f) - \phi(f)| < \varepsilon$. Now take any neighborhood $B_{f_1,\dots,f_n,\varepsilon}(\phi)$. Then, let $\alpha_0 \equiv \max_{i=1}^n \alpha_{f_i}$, and it follows that $\phi_\alpha \in B_{f_1,\dots,f_n,\varepsilon}(\phi)$, for $\alpha \geq \alpha_0$, hence $\phi_a \to \phi$. For the converse, we have that for any $n \in \mathbb{N}$, any collection $f_1, \dots, f_n$, and any $\varepsilon > 0$, there exists $\alpha_0$ such that, if $\phi_{\alpha_0} \in B_{f_1,\dots,f_n,\varepsilon}(\phi)$, then for all $\alpha \geq \alpha_0$, $\phi_\alpha \in B_{f_1,\dots,f_n,\varepsilon}(\phi)$. Thus to show that for any given $f$, $\phi_\alpha(f) \to \phi(f)$ we just have to use this fact with $B_{f,\varepsilon}(\phi)$. □

One of the most important facts about the weak-$*$ topology is Alaoglu's theorem. The space $C(J)^*$ is in fact a Banach space equipped with the norm $\|\phi\| \equiv \sup_{f\in C(J)} \frac{\phi(f)}{\|f\|_\infty}$

**Theorem 2.1.4** *The unit ball*

$$\{\phi \in C(J)^* : \|\phi\| \leq 1\} \tag{2.3}$$

*is compact in the weak-$*$ topology.*

(for a proof, see any textbook on functional analysis, e.g. Dunford and Schwartz [5]).

The importance for us is that when combined with the Riesz representation theorem, it yields:

**Corollary 2.1.5** *The set of inner regular probability measures on a compact Hausdorff space is compact in the weak-$*$ topology.*

*Proof.* By the Riesz representation theorem, each inner regular probability measure corresponds to a unique increasing functional, $\phi \in C(J)^*$ with $\phi(1) = 1$. Since the function $f \equiv 1$ is the largest function such that $\|f\|_\infty \leq 1$, it follows that $\|\phi\| \leq \phi(1) = 1$. Hence this set is a subset of the unit ball. Moreover, the set of increasing (in the sense of non-decreasing) linear functionals mapping 1 to 1 is closed, and hence, as a closed subset of a compact set, compact. $\square$

**Corollary 2.1.6** *The set of probability measures on a compact metrisable space is compact in the weak-$*$ topology.*

*Proof.* By Theorem 1.2.6 in [2], any probability measure on a compact metrisable space is inner regular, hence the restriction to inner regular measures in Corollary 2.1.5 can be dropped in this case. $\square$

As a matter of fact, in the compact metrisable case we get even more.

**Theorem 2.1.7** *Let $J$ be a compact metrisable space. Then $C(J)$ is separable, and $\mathcal{M}_1(J)$ equipped with the weak-$*$ topology is compact metrisable.*

*Proof.* We may take $J$ to be metric with metric $\rho$. Since $J$ is separable (any compact metric space is separable), there is a countable dense set of points, $x_n$, $n \in \mathbb{N}$. Define the functions

$$h_n(x) \equiv \rho(x, x_n).$$

The functions $h_n$ separate points in $J$, i.e. if $x \neq y$, then there exists $n$ such that $h_n(x) \neq h_n(y)$. Now let $A$ be the set of all functions of the form

$$q\mathbb{1} + \sum_{n_1,\dots,n_r;k_1,\dots,k_r} q(n_1,\dots,n_r;k_1,\dots,k_r) h_{n_1}^{k_1} \dots h_{n_r}^{k_r}$$

where all $q$'s are rational. Then the closure of $A$ is an algebra containing all constant functions and separating points in $J$. The Stone-Weierstrass theorem asserts therefore that the countable set $A$ is dense in $C(J)$, so $C(J)$ is separable.

Now let $f_n$, $n \in \mathbb{N}$, be a countable dense subset of $C(J)$. Consider the map $\Phi : \mathcal{M}_1(J) \to V \equiv \times_{n\in\mathbb{N}}[-\|f_n\|_\infty, \|f_n\|_\infty]$, given by

$$\Phi(\mu) = (\mu(f_1), \mu(f_2), \dots).$$

This map is one to one. Namely, assume that $\mu \neq \nu$, but $\Phi(\mu) = \Phi(\nu)$. Then on the one hand, there must exists $f \in C(J)$ such that $\mu(f) \neq \nu(f)$, while for all $n$, $\mu(f_n) = \nu(f_n)$. But there are sequences $f_i \in A$ such that $f_i \to f$. Thus $\lim_i \mu(f_i) = \lim_i \nu(f_i)$, and by dominated convergence, both limits equal $\mu(f)$, resp. $\nu(f)$, which must be equal contrary to the assumption. Moreover, the set $A$ determines convergence, i.e. a net $\mu_\alpha$ converges to $\mu$ (in the weak-$*$ topology , if $\mu_\alpha(f_n) \to \mu(f_n)$, for all $f_n \in A$. But the product space $V$ is compact and metrisable (by Tychonoff's theorem), and from the above, $\mathcal{M}_1(J)$ is homeomorphic to a compact subset of this space. Thus it is compact and metrisable. □

Let us remark that a metric on $\mathcal{M}_1(J)$ can be defined by

$$\hat{\rho}(\mu, \nu) \equiv \sum_{n=1}^{\infty} 2^{-n} \left(1 - e^{-|\mu(f_n)-\nu(f_n)|}\right). \tag{2.4}$$

## 2.2 Polish and Lousin spaces

When dealing with stochastic processes, an obviously important space is that of continuous, real valued functions on $\mathbb{R}_+$. We will call

$$W \equiv C([0, \infty), \mathbb{R}). \tag{2.5}$$

This space is not compact, so we have to go slightly beyond the previous setting.

**Lemma 2.2.8** *The space $W$ equipped with the topology of uniform convergence on compact sets is a Polish space. The $\sigma$-algebra, $\mathcal{A}$, of cylinders generated by the projections $\pi_t : W \to \mathbb{R}$, $\pi_t(w) = w(t)$, is the Borel-$\sigma$-algebra on $W$.*

*Proof.* We can metrise the topology on $W$ by the metric

$$\rho(w_1, w_2) \equiv \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_n(w_1, w_2)}{1 + \rho_n(w_1, w_2)},$$

where

$$\rho_n(w_1, w_2) \equiv \sup_{0\leq t\leq n} |w_1(t) - w_2(t)|.$$

Then it inherits its properties form the metric space $C([0, n), \mathbb{R})$ equipped with the uniform topology.

Now the maps $\pi_t$ are continuous, and hence $\mathcal{A} \subset \mathcal{B}(W)$. On the other hand, for continuous functions, $w_i$,

$$\rho_n(w_1, w_2) = \sup_{q \in \mathbb{Q} \cap [0,n]} |w_1(q) - w_2(q)|,$$

so that $\rho_n$ and hence $\rho$ are $\mathcal{A}$-measurable. Now let $F$ be a closed subset of $W$. Take a countable dense subset of $F$, say $w_n, n \in \mathbb{N}$. Then

$$F = \{w \in W : \inf_n \rho(w, w_n) = 0\},$$

which (since all is countable) implies that $F \in \mathcal{A}$, and thus $\mathcal{A} = \mathcal{B}(W)$. □

This (and the fact that quite similarly the corresponding spaces of càdlàg functions are Polish) implies that we can most of the time assume that we will be working on Polish probability spaces. In the construction of stochastic processes we have actually been working on Lousin spaces (and used the fact that these are homeomorphic to a Borel subset of a compact metric space). The next theorem nicely clarifies that Polish spaces are even better.

**Theorem 2.2.9** *A topological space is Polish, if and only if it is homeomorphic to a $G_\delta$ subset (i.e. a countable intersection of open subsets) of a compact metric space. In particular, every Polish space is a Lousin space.*

*Proof.* We really only care about the "only if" part and only give its proof. Let $S$ be our Polish space. We will actually show that it can be embedded in a $G_\delta$ subset of the compact metrisable space $J \equiv [0,1]^{\mathbb{N}}$. Let $\rho$ be a metric on $S$, and set $\hat{\rho} = \frac{\rho}{1+\rho}$. This is an equivalent metric that is bounded by 1. Chose a countable dense subset $x_n$, $n \in \mathbb{N}$, of $S$ and define

$$\alpha(x) \equiv (\hat{\rho}(x, x_1), \hat{\rho}(x, x_2), \dots).$$

Let us show that $\alpha$ is a homeomorphism from $S$ to its image, $\alpha(S) \subset [0,1]^{\mathbb{N}}$. For this we must show that a sequence of elements $x(n)$ converges to $x$, if and only if

$$\hat{\rho}(x(n), x_k) \to \hat{\rho}(x, x_k),$$

for all $k$. The only if direction foillow from the continuity of the map $\hat{\rho}(\cdot, x_k)$. To show the other direction, note that by the triangle inequality

$$\hat{\rho}(x(n), x) \le \hat{\rho}(x(n), x_k) + \hat{\rho}(x_k, x).$$

Therefore, for all $k$,

$$\limsup \hat{\rho}(x(n), x) \leq 2\hat{\rho}(x_k, x). \tag{2.6}$$

Now take a sequence of $x_k$ that converges to $x$. Then (2.6) implies that $\limsup \hat{\rho}(x(n), x) \leq 0$, and so $x(n) \to$, as desired.

Next, let $d$ be a metric on $J$. By continuity of the inverse map $\alpha^{-1}$ on the image of $S$, for any $n \in \mathbb{N}$ we can find $1/2n \geq \delta > 0$, such that the pre-image of the ball $B_d(\alpha(x), \delta) \cap \alpha(S)$ has diameter smaller than $1/n$ (with respect to the metric $\hat{\rho}$).

Now think of $\alpha(S)$ as a subset of $J$. Let $\bar{\alpha}(S)$ be its closure. For $n$ given, let $U_n$ be the union of all points $x \in \bar{\alpha}(S)$ such that it has a neighborhood, $N_{n,x}$ in $J$ such that $\alpha^{-1}(N_{n,x} \cap \alpha(S))$ has $\hat{\rho}$-diameter at most $1/n$. Note that by what we just showed, all points in $\alpha(S)$ belong to $U_n$. Now we show that $U_n$ is open in $\bar{\alpha}(S)$: if $x \in U_n$, and $y \in \bar{\alpha}(S)$ is close enough to $x$, then $y \in N_{n,x}$, and the set $N_{n,x}$ may serve as $N_{n,y}$, so that $y \in U_n$. Thus $U_n$ is open.

Now let $x \in \bigcap_n U_n$. Choose for any $n$ a point $x_n \in \alpha(S) \cap \bigcap_{k \leq n} N_{k,x}$. Clearly $d(x, x_n) \leq 1/n$ and hence $x_n \to x$. Moreover, for any $r \geq n$, both $x_r \in N_{n,x}$ and $x_n \in N_{n,x}$, so that $\hat{\rho}(\alpha^{-1}(x_r), \alpha^{-1}(x_n)) \leq 1/n$. Thus $\alpha^{-1}(x_n)$ is a Cauchy sequence in complete metric space, and so $\alpha^{-1}(x_n) \to y \in S$. Thus, since $\alpha$ is a homeomorphism, $x_n \to \alpha(y)$ in $J$, and clearly $\alpha(y) = x$, implying that $\alpha(S) = \bigcap_n U_n$. Finally, since $U_n$ is open in $\bar{\alpha}(S)$, there are open sets $V_n$ such that $U_n = \bar{\alpha}(S) \cap V_n$. Hence

$$\alpha(S) = \bar{\alpha}(S) \cap \left( \bigcap_n V_n \right).$$

Remember that we want to show that $\alpha(S)$ is a countable intersection of open sets: all that remains to show that is that $\bar{\alpha}(S)$ is such a set, but this is obvious in a metric space:

$$\bar{\alpha}(S) = \bigcap_n \{y \in J : d(y, \alpha(S)) < 1/n\}.$$

□

On the space of probability measures on Lousin spaces we introduce a the weak-$*$ topology with respect to the set of bounded continuous functions (the boundedness having been trivial in the compact setting). Convergence in this topology is usually called *weak convergence*, which is bad, since it is not what weak convergence would be in functional analysis. But that is how it is, anyway.

Let us state this as a definition:

**Definition 2.2.1** Let $S$ be a Lousin space. Let $C_b(S)$ be the space of bounded, continuous functions on $S$, and let $\mathcal{M}_1(S)$ be the space of probability measures on $S$. Then a net, $\mu_\alpha \in \mathcal{M}_1(S)$ converges *weakly* to $\mu \in \mathcal{M}_1(S)$, if and only if, for all $f \in C_b(S)$,

$$\mu_a(f) \to \mu(f). \tag{2.7}$$

Weak convergence is related to convergence in probability.

**Lemma 2.2.10** *Assume that $X_n$ is a sequence of random variables with values in a Polish space such that $X_n \to X$ in probability, where $X$ is a random variable on the same probability space. Let $\mu_n, \mu$ denote their distributions. Then $\mu_n \to \mu$ weakly.*

*Proof.* Let us first show that convergence in probability implies convergence of $\mu_n(f)$ if $f$ be a bounded uniformly continuous function. Then there there exists $C < \infty$ such that $|f(x)| \leq C$ and for any $\delta > 0$ there exists $\varepsilon = \varepsilon(\delta)$ such that $\rho(x-y) \leq \varepsilon$ implies $|f(x) - f(y)| \leq \delta$. Clearly

$$\begin{aligned} |\mu_n(f) - \mu(f)| &= |\mathbb{E}(f(X_n) - f(X))| \\ &\leq \left|\mathbb{E}\left[(f(X_n) - f(X))\mathbb{1}_{\rho(X_n - X) \leq \varepsilon}\right]\right| \\ &\quad + \left|\mathbb{E}\left[(f(X_n) - f(X))\mathbb{1}_{\rho(X_n - X) > \varepsilon}\right]\right| \\ &\leq \delta + C\mathbb{P}\left(\rho(X_n - X) > \varepsilon\right) \end{aligned} \tag{2.8}$$

Since the second term on the right tends to zero as $n \uparrow \infty$ for any $\varepsilon > 0$, for any $\delta > 0$,

$$\limsup_{n\uparrow\infty} |\mu_n(f) - \mu(f)| \leq \delta,$$

hence

$$\lim_{n\uparrow\infty} |\mu_n(f) - \mu(f)| = 0,$$

as claimed.

To conclude the prove, we must only show that convergence of $\mu_n(f)$ to $\mu(f)$ for all absolutely continuous functions implies that the same holds for all bounded continuous functions. To this end we use that if $f$ is a bounded continuous function, then there exists s sequence of uniformly continuous functions, $f_k$, such that $\|f_k - f\|_\infty \to 0$. One then has the decomposition

$$|\mu_n(f) - \mu(f)| \leq \mu_n(|f - f_k|) + |\mu_n(f_k) - \mu(f_k)| + \mu(|f_k - f|).$$

by unifrom convergence of $f_k$ to $f$, the first term is smaller than $\varepsilon/3$, provided only $k$ is large enough; the second bracket is smaller than $\varepsilon/3$ if $n \geq n_0(k)$; the last bracket is smaller than $\varepsilon/3$, of $k$ is large enough, independent of $n$. Hence choosing $k \geq k_0$ and $n \geq n_0(k)$, we see that for any $\varepsilon > 0$, there exists $n_0$, s.t. for $n \geq n_0$, $|\mu_n(f) - \mu(f)| \leq \varepsilon$. □

The following characterization of weak convergence is important, but the proof is somewhat technical and will be skipped (try as an exercise).

**Proposition 2.2.11** *Let $\mu_\alpha$ be a net of elements of $\mathcal{M}_1(S)$ where $S$ is a Lousin space. Then the following conditions are equivalent:*

*(i)* $\mu_\alpha \to \mu$ *weakly;*
*(ii) for every closed* $F \subset S$, $\limsup \mu_\alpha(F) \leq \mu(F)$*;*
*(iii) for every open* $G \subset S$, $\liminf \mu_\alpha(G) \geq \mu(G)$*;*

*Thus, if $B \in \mathcal{B}(S)$ with $\mu(\partial B) = 0$, then, if $\mu_\alpha \to \mu$, then $\mu_\alpha(B) \to \mu(B)$.*

We will use this proposition to prove the fundamental result that the weak topology on $\mathcal{M}_1(S)$ is metrisable if $S$ is Lousin. This is very convenient, and in particular will allow us to never use nets anymore!

**Theorem 2.2.12** *Let $S$ be a Lousin space and let $J$ be the compact metrisable space such that $S$ is homeomorphic to one of its Borel subsets, $B$. Let $\hat{\mu}$ be the extension of (the natural image of[1]) $\mu$ on $B$ on $J$ such that $\hat{\mu}(J \backslash B) = 0$. The map $\mu \to \hat{\mu}$ is a homeomorphisms from $\mathcal{M}_1(S)$ to the set $\{\nu \in \mathcal{M}_1(J) : \nu(B) = 1\}$ in the weak topologies. Therefore, the weak topology on $\mathcal{M}_1(S)$ is metrisable.*

*Proof.* We must show that, if $\mu_\alpha$ is a net in $\mathcal{M}_1(S)$ and $\mu \in \mathcal{M}_1(S)$, then the conditions

(i) $\mu_\alpha(f) \to \mu(f), \forall f \in C_b(S)$, and
(ii) $\hat{\mu}_\alpha(f) \to \hat{\mu}(f), \forall f \in C(J)$

are equivalent. Assume that (i) holds. Let $f \in C(J)$ and set $f_B = f \mathbb{1}_B$. Clearly $f_B$ is bounded on $B$, and if $\phi : S \to B$ is our homeomorphism, then $g \equiv f_B \circ \phi$ is a bounded function on $S$, and $\mu_n(g) = \hat{\mu}_n(f_B) = \hat{\mu}_n(f)$. Thus (i) implies (ii).

Now assume that (ii) holds. Let $F \subset S$ be a closed. Then there exists a closed subset, $Y$, of $J$ such that $F = \phi^{-1}(B \cap Y)$. By Proposition 2.2.11,

[1] That is, if $A \in \mathcal{B}(J)$, then $\hat{\mu}(A) \equiv \mu(\phi^{-1}(A \cap B))$

$$\limsup \mu_\alpha(F) = \limsup \hat{\mu}_\alpha(B \cap Y) = \limsup \hat{\mu}_\alpha(Y)$$
$$\leq \hat{\mu}(Y) = \hat{\mu}(B \cap Y) = \mu(F).$$

Hence again by Proposition 2.2.11, (i) holds.

Now that we have shown that the space $\mathcal{M}_1(S)$ is homeomorphic to a subspace of the compact metrisable space $\mathcal{M}_1(J)$ (because of Theorem 2.5), $\mathcal{M}_1(S)$ is metrisable. □

We now introduce the very important concept of *tightness*. The point here is the following. We already know, from the Kolmogorov-Daniell theorem, that the finite dimensional marginals of a process determine its law. It is frequently possible, for a sequence of processes, to prove convergence of of the finite dimensional marginals. However, to have path properties, we want to construct the process on a more suitable space of, say, continuous or càdlàg paths. The question is whether the sequence converges weakly to a probability measure on on this space. For this purpose it is useful to have a compactness criterion for set of probability measures (e.g. for the sequence under consideration). This is provided by the famous *Prohorov theorem* .

We need to recall the definition of conditional compactness.

**Definition 2.2.2** Let $S$ be a topological space. A subset, $J \subset S$, is called *conditionally compact* if its closure in the weak topology is compact. $J$ is called *conditionally sequentially compact*, if its closure is sequentially compact. If $S$ is a metrisable space, then any conditionally compact set is conditionally sequentially compact.

**Remark 2.2.1** The terms *conditionally compact* and *relatively compact* are used interchangebly by different authors with the same meaning.

The usefulness of this notion for us lies in the following. Assume that we are given a sequence of probability measures, $\mu_n$, on some space, $S$. If the set $\{\mu_n, n \in \mathbb{N}\}$, is conditionally sequentially compact in the weak topology, then there exist limit points, $\mu \in \mathcal{M}_1(S)$, and subsequences, $n_k$), such that $\mu_{n_k} \to \mu$, in the weak topology. E.g., if we take as our space $S$ the space of càdlàg paths, if our sequence of meaures is tight, the limit points will be probability measures on càdlàg paths.

**Definition 2.2.3** A subset, $H \subset \mathcal{M}_1(S)$ is called *tight*, if and only if there exists, for any $\varepsilon > 0$, a compact set $K_\varepsilon \subset S$, such that, for all $\mu \in H$,

$$\mu(K_\varepsilon) > 1 - \varepsilon. \tag{2.9}$$

**Theorem 2.2.13 (Prohorov)** *If $S$ is a Lousin space, then a subset $H \subset \mathcal{M}_1(S)$ is* conditionally compact*, if it is tight.*

*If $S$ is a Polish space then any conditionally compact subset of $\mathcal{M}_1(S)$ is tight.*

*Moreover, since the spaces $\mathcal{M}_1(S)$ are metrisable under both hypothesis, conditionally compact may be replaced by sequentially conditionally compact in both statements.*

*Proof.* We prove the first (and most important statement). Let again $J$ be the compact metrisable space, and let $\phi$ be a homeomorphiosm $\phi : \Sigma \to B \subset J$, for some Borel set $B$. We know that $\mathcal{M}_1(J)$ is compact metrisable, so that every subset of it is conditionally compact. Since compactness and sequential compactness are equivalent in our setting, we know that any sequence, $\hat{\mu}_n \in \mathcal{M}_1(J)$ has limit points in $\mathcal{M}_1(J)$. Now let $H = \{\mu_n, n \in \mathbb{N}\} \subset \mathcal{M}_1(S)$ be tight. Let $\hat{\mu}_N \equiv \mu_n \circ \phi^{-1}$. Let $\hat{\mu}$ be a limit point of the sequence $\hat{\mu}_n$. We want to show that $\hat{\mu}$ is the image of a probability measure on $S$, and thus $\mu \equiv \hat{\mu} \circ \phi$ exists and is a limt point of the sequence $\mu_n$. For this we need to show that $\hat{\mu}(B) = 1$. Now let $K_\varepsilon$ be the compact set in $S$ such that $\mu_n(K_\varepsilon) > 1 - \varepsilon$. Then, by Proposition 2.2.11,

$$\hat{\mu}(\phi(K_\varepsilon)) \geq \limsup_n \hat{\mu}_n(\phi(K_\varepsilon)) = \limsup_n \mu_n(K_\varepsilon) \geq 1 - \varepsilon,$$

for all $\varepsilon > 0$, and so $\hat{\mu}(B) = 1$, as desired.

The proof of the less important converse will be skipped. □

We will consider an application of the Prohorov theorem in the case when $S$ is the space, $W$, of continuous paths defined in (2.5).

This is based on the *Arzelà–Ascoli theorem* that characterizes conditionally compact set in $W$.

**Theorem 2.2.14** *A subset, $\Gamma \subset W$ is conditionally compact if and only if the following hold:*

*(i)* $\sup\{|w(0)| : w \in \Gamma\} < \infty$*;*
*(ii)* $\forall_{N \in \mathbb{N}} \lim_{\delta \downarrow 0} \sup_{w \in \Gamma} \Delta(\delta, N, w) = 0$*, where*

$$\Delta(\delta, N.w) \equiv \sup\left\{|w(t) - w(s)| : t, s \in [0, N], |t - s| < \delta\right\}. \tag{2.10}$$

For the proof, see texts on functional analysis, e.g. [5].

This allows us to formulate the following tightness-criterion.

**Theorem 2.2.15** *A subset, $H \subset \mathcal{M}_1(W)$, is conditionally compact (equiv. tight), if and only if:*

*(i)* $\lim_{c\uparrow\infty} \sup_{\mu\in H} \mu\left(|w(0)| > c\right) = 0$*;*

*(ii) for all $n \in \mathbb{N}$ and all $\varepsilon > 0$, $\lim_{\delta\downarrow 0} \sup_{\mu\in H} \mu\left(\Delta(\delta, N.w) > \varepsilon\right) = 0$, where $\Delta$ is defined in (2.10)*

*Proof.* We give only the prove of the relevant "if" direction. We should find a compact subset of $W$ of measure aritrarily close to one for all measures in $H$. Clearly, we can do this by giving a conditionally compact set, $\Gamma_\varepsilon$, of measure $\mu(\Gamma_\varepsilon) > 1-\varepsilon$, since then its closure is a compact set of at least the same measure. Now assume that (i) and (ii) hold. Then take, for given $\varepsilon$, $C$ such that the set

$$A \equiv \{w \in W : |w(0)| \leq C\}$$

satisfies, for all $\mu \in H$, $\mu(A) \leq 1 - \varepsilon/2$. By (ii) we can chose $\delta(n, N)$ such that the sets

$$A_{n,N} \equiv \{w \in W : \Delta(\delta, N, w) \leq 1/n\}$$

satisfy, for all $\mu \in H$, $\mu(A_{n,N}) \geq 1 - \varepsilon 2^{-(n+N+2)}$. Then the set

$$\Gamma \equiv A \cap \bigcap_{n,N\in\mathbb{N}} A_{n,N}$$

satisfies $\mu(\Gamma) > 1 - \varepsilon$, for all $\mu \in H$.

This proves this part of the theorem. □

Finally we come to the most important result of this chapter.

**Lemma 2.2.16** *Let $\mu_n, \mu$ be probability measures in $W$. Then $\mu_n$ converges weakly to $\mu$, if and only if*

*(i) the finite dimensional distributions of $\mu_n$ converge to those of $\mu$;*

*(ii) the family $\{\mu_n, n \in \mathbb{N}\}$ is tight.*

*Proof.* Let us first show the "if" direction. From tightness and Prohorov's theorem it follows that the family $\{\mu_n, n \in \mathbb{N}\}$ is conditionally sequentially compact, so that there are subsequences, $n(k)$, along which $\mu_{n(k)}$ converges weakly to some measure $\mu$. Assume that there is another subsequence, $m(k)$, such that $\mu_{m(k)}$ converges weakly to a measure $\nu$.

But then also the finite dimensional distibutions of $\mu_{n(k)}$, respectively, $\mu_{m(k)}$, converge to those of $\mu$, repectively $\nu$. But by (i), the finite dimensional marginals of $\mu_n$ converge, so that $\mu$ and $\nu$ have the same finite dimensional marginals, and hence, are the same measures. Since this holds for any limit point, it follows that $\mu_n \to \mu$, weakly.

The "only if" direction: first, the projection to finite dimensional marginals is a continuous map, hence weak convergence implies that of the marginals. Second, Prohorov's theorem in the case of the Polish space $W$ implies that the existence of sequential limits, hence sequential conditional compactness, hence conditional compactness implies tightness. □

**Exercise.** As an application of this theorem, you are invited to prove Donsker's theorem (Theorem 6.3.3 in [2]) without using the Skorokhod embedding that was used in in the last section of [2]. Note that we already have: (i) convergence of the finite dimensional distributions (Exercise in [2]) and the existence of BM on $W$. Thus all you need to prove tightness of the sequences $S_n(t)$. Note that here it pays to chose the linealy interpolated version (6.3) in [2].

Finally, we give a useful characterisation of weak convergence, known as Skorokhod's theorem, that may appear somewhat surprising at first sight. It is, however, extremely useful.

**Theorem 2.2.17** *Let $S$ be a Lousin space and assume the $\mu_n, \mu$ are probability measures on $S$. Assume that $\mu_n \to \mu$ weakly. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables $X_n$ with law $\mu_n$, and $X$ with law $\mu$, such that $X_n \to X$ $\mathbb{P}$-almost surely.*

*Proof.* The proof is quite simple in the case when $S = \mathbb{R}$. In that case, weak convergence is equivalent to convergence of the distribution function, $F_n(x) = \mu([-\infty, x])$ at all continuity points of the limit, $F$. In that case we chose the probability space $\Omega = [0,1]$, $\mathbb{P}$ the uniform measure on $[0,1]$ and define the random variables $X_n(x) = F_n^{-1}(x)$. Then clearly

$$\mathbb{P}\left(X_n \leq z\right) = \mathbb{P}\left(x \leq F_n(z)\right) = F_n(z)$$

so that indeed $X_n$ has the desired law. On the other hand, $F_n(x)$ converges to $F(x)$ at all continuity points of $F$, and one can check that the same is true for $F_n^{-1}$, implying almost sure convergence of $X_n$.

In the general case, the prove is quite involved and probably not very enlightening.... □

Skorohod's theorem is very useful if one wants to prove convergence of functionals of probability distributions.

## 2.3 The càdlàg space $D_E[0,\infty)$

In the general theory of Markov processes it will be important that we can treat the space of càdlàg functions with values in a metric space as a Polish space much like the space of continuous functions. The material from this section is taken from [6] where omitted proofs and further details can be found.

### 2.3.1 A Skorokhod metric

We will now construct a metric on càdlàg space which will turn this space into a complete metric space. This was first don by Skorokhod. In fact, there are various different metrics one may put on this space which will give rise to different convergence properties. This is mostly related to the question whether each jump in the limiting function is associated to one, several, or no jumps in approximating functions. A detailed discussion of these issues can be found in [15]. Here we consider only one case.

**Definition 2.3.1** Let $\Lambda$ denote the set of all strictly increasing maps $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, such that $\lambda$ is Lipshitz continuous and

$$\gamma(\lambda) \equiv \sup_{0\leq t<s} \left| \ln \frac{\lambda(s)-\lambda(t)}{s-t} \right| < \infty. \tag{2.11}$$

For $x, y \in D_E[0,\infty)$, $u \in \mathbb{R}_+$, and $\lambda \in \Lambda$, set

$$d(x,y,\lambda,u) \equiv \sup_{\geq 0} \rho\left(x(t\wedge u), y(\lambda(t)\wedge u)\right). \tag{2.12}$$

Finally, the *Skorohod metric* on $D_E[0,\infty)$ is given as

$$d(x,y) \equiv \inf_{\lambda\in\Lambda} \left( \gamma(\lambda) \vee \int_0^\infty e^{-u} d(x,y,u,\lambda) du \right). \tag{2.13}$$

To get the idea behind this definition, note that with $\lambda$ the identity, this is just the metric on the space of continuous functions. The rôle of the $\lambda$ is to make the distance of two functions that look much the same except that they jump at two points very close to each other by sizable amount. E.g., we clearly want the functions

$$x_n(t) = \mathbb{1}_{[1/n,\infty]}(t)$$

to converge to the function

$$x_\infty(t) = 1_{[0,\infty]}(t).$$

This is wrong under the sup-norm, since $\sup_t \|x_n(t) - x_\infty(t)\| = 1$, but it will be true under the metric $d$ (Exercise!).

**Lemma 2.3.18** *$d$ as defined a above is a metric on $D_E[0,\infty)$.*

*Proof.* We first show that $d(x, y) = 0$ implies $y = x$. Note that for $d(x, y) = 0$, it must be true that there exists a sequence $\lambda_n$ such that $\gamma(\lambda_0) \downarrow 0$ and $\lim_{n\uparrow\infty} d(x, y, \lambda_n, u) = 0$; one easily checks that then

$$\lim_{n\uparrow\infty} \sup_{0\le t\le T} |\lambda_n(t) - t| = 0,$$

and hence $x(t) = y(t)$ at all continuity points of $x$. But since $x$ and $y$ are càdlàg , this implies $x = y$.

Symmetry follow from the fact that $d(x, y, \lambda, u) = d(y, x, \lambda^{-1}, u)$ and that $\gamma(\lambda) = \gamma(\lambda^{-1})$.

Finally we need to prove the triangle inequality. A simple calculation shows that

$$d(x, z, \lambda_2 \circ \lambda_1, u) \le d(x, y, \lambda_1, u) + d(y, z, \lambda_2, u).$$

Finally $\gamma(\lambda_1 \circ \lambda_2) \le \gamma(\lambda_1) + \gamma(\lambda_2)$, and putting this together one derives $d(x, z) \le d(x, y) + d(y, z)$. □

**Exercise:** Fill in the details of the proof of the triangle inequality.

The next theorem completes our task.

**Theorem 2.3.19** *If $E$ is separable, then $D_E[0,\infty)$ is separable, and if $E$ is complete, then $D_E[0,\infty)$ is complete.*

*Proof.* The proof of the first statement is similar to the proof of the separability of $C(J)$ (Theorem 2.1.7) and is left to the reader. To prove completeness, we only need to show that every Cauchy sequence converges. Thus let $x_n \in D_E[0,\infty)$ be Cauchy. Then, for any constant $C > 1$, and any $k \in \mathbb{N}$, there exist values $n_k$, such that for all $n, m \ge n_k$, $d(x_n, x_m) \le C^{-k}$. Then we can select sequences $u_k$, and $\lambda_k$, such that

$$\gamma(\lambda_k) \vee d(x_{n_k}, x_{n_{k+1}}, \lambda_k, u_k) \le 2^{-k}.$$

Then, in particular,

$$\mu_k \equiv \lim_{m\uparrow\infty} \lambda_{k+m} \circ \lambda_{k+m-1} \circ \cdots \circ \lambda_{k+1} \circ \lambda_k$$

exists and satisfies

$$\gamma(\mu_k) \leq \sum_{m=k^\infty} \gamma(\lambda_m) \leq 2^{-k+1}.$$

Now

$$\begin{aligned} &\sup_{t\geq 0} \rho\left(x_{n_k}(\mu_k^{-1}(t)\wedge u_k), x_{n_{k+1}}(\mu_{k+1}^{-1}(t)\wedge u_k)\right) \\ &= \sup_{t\geq 0} \rho\left(x_{n_k}(\mu_k^{-1}(t)\wedge u_k), x_{n_{k+1}}(\lambda_k(\mu_{k+1}^{-1}(t))\wedge u_k)\right) \\ &= \sup_{t\geq 0} \rho\left(x_{n_k}(t\wedge u_k), x_{n_{k+1}}(\lambda_k^{-1}(t)\wedge u_k)\right) \\ &\leq 2^{-k}. \end{aligned}$$

Therefore, by the completeness of $E$, the sequence of functions $z_k \equiv x_{n_k}(\mu_k^{-1}(t))$ converges uniformly on compact intervals to a function $z$. Each $z_k$ being càdlàg , so $z$ is also càdlàg . Since $\gamma(\mu_k) \to 0$, it follows that

$$\lim_{k\uparrow\infty} \sup_{0\leq t\leq T} \rho(x_{n_k}(\mu_k^{-1}(t)), z(t)) = 0,$$

for all $T$, and hence $d(x_{n_k}, z) \to 0$. Since a Cauchy sequence that contains a convergent subsequence converges, the proof is complete. □

To use Prohorov's theorem for proving convergence of probability measures on the space $D_E[0,\infty)$, we need first a characterisation of compact sets.

The first lemma states that the clusure of the space of step functions that are uniformly bounded and where the distance between steps is uniformly bounded from below is compact:

**Lemma 2.3.20** *Let $\Gamma \subset E$ be compact and $\delta > 0$ be fixed. Let $A(\Gamma,\delta)$ denote the set of step functions, $x$, in $D_E[0,\infty)$ such that*

*(i)* $x(t) \in \Gamma$, *for all* $\tau \in [0,\infty)$, *and*
*(ii)* $s_k(x) - s_{k-1}(x) > d$, *for all* $k \in \mathbb{N}$,

*where*

$$s_k(x) \equiv \inf\{t > s_{k-1}(x) : x(t) \neq x(t-)\}.$$

*Then the closure of $A(\Gamma,\delta)$ is compact.*

We leave the prove as an exercise.

The analog of the modulus of continuity in the Arzelà-Ascoli theorem on càdlàg space is the following: For $x \in D_E[0,\infty)$, $\delta > 0$, and $T < \infty$, set

$$w(x, \delta, T) \equiv \inf_{t_i} \max_i \sup_{s,t \in [t_{i-1}, t_i)} \rho(x(s), x(t)), \tag{2.14}$$

where the first infimum is over all collections $0 = t_0 < t_1 < \cdots < t_{n-1} < T < t_n$, with $t_i - t_{i-1} < \delta$, for all $i$.

The following theorem is the analog of the Arzelà-Ascoli theorem:

**Theorem 2.3.21** *Let $E$ be a complete metric space. Then the closure of a set $A \subset D_E([0, \infty)$ is compact, if and only if,*

*(i) For every rational $t \geq 0$, there exists a compact set $\Gamma_t \subset E$, such that for all $x \in A$; $x(t) \in \Gamma_t$.*
*(ii) For each $T < \infty$,*

$$\lim_{\delta \downarrow 0} \sup_{x \in A} w(x, \delta, T) = 0. \tag{2.15}$$

A proof of this result can be found, e.g. in [6].

Based on this theorem, we now get the crucial tightness criterion:

**Theorem 2.3.22** *Let $E$ be complete and separable, and let $X_\alpha$ be a family of processes with càdlàg paths. Then the family of probability laws, $\mu_\alpha$, of $X_\alpha$, is conditionally compact, if and only if the following holds:*

*(i) For every $\eta > 0$ and rational $t \geq 0$, there exists a compact set, $\Gamma_{\eta,t} \subset E$, such that*

$$\inf_\alpha \mu_\alpha \left(x(t) \in \Gamma_{\eta,t}\right) \geq 1 - \eta, \tag{2.16}$$

*and*
*(ii) For every $\eta > 0$ and $T < \infty$, there exists $\delta > 0$, such that*

$$\sup_\alpha \mu_\alpha \left(w(x, \delta, T) \geq \eta\right) \leq \eta. \tag{2.17}$$

An application of the preceeding theorem to the case of Lévy processes allows us to prove that the processes constructed in Section 1 from Poisson point processes do indeed have càdlàg paths with probability one, i.e. they have a modification that are Lévy processes.

**Exercise.** Consider the family of processes defined by the first line of (1.24). Show that the corresponding family of laws on $D_{\mathbb{R}}[0, \infty)$ is tight. *Hint:* Introduce a further cutoff, $\varepsilon_0$, to break this process into one with small jumps and one with few jumps. Use a maximum inequality for the small jump part, and the fact that the large jump part is a compound Poisson process.

# 3

# Markov processes

In this chapter we return to the most important class of stochastic processes, *Markov processes*. In Chapter 5 of [2] we have seen a lot of aspects of Markov processes in the case of discrete time. We would expect to have many similar results in continuous time, but on the technical level, we will encounter many analytical problems that were absent in the discrete time setting. The need for studying continuous time processes is motivated in part from the fact that they arise a natural limits of discrete time processes. We have already seen this in the case of Brownian motion, but the same holds for certain classes of Lévy processes. We will also see that they lend themselves in may respects to simpler, or more elegant computations and are therefore used in many areas of applications, e.g. mathematical finance. In the remainder of this section, $S$ denotes at least a Lousin space, and in fact you may assume $S$ to be Polish. In this section we will restrict our attention to *time-homogeneous* Markov process.
**Notation:** In this section $S$ will usually denote a metric space. Then $B(S,\mathbb{R}) \equiv B(S)$ will be the space of real valued, bounded, measurable functions on $S$; $C(S,\mathbb{R}) \equiv C(S)$ will be the space of continuous functions, $C_b(S,\mathbb{R}) \equiv C_b(S)$ the space of bounded continuous functions, and $C_0(S,\mathbb{R}) \equiv C_0(S)$ the space of bounded continuous functions that vanish at infinity. Clearly $C_0(\Sigma) \subset C_b(S) \subset C(S) \subset B(S)$.

### 3.1 Semi-groups, resolvents, generators

The main building block for a time homogeneous Markov process is the so called transition kernel, $P : \mathbb{R}_+ \times S \times \mathcal{B} \to [0,1]$.

### 3.1.1 Transition functions and semi-groups

We will denote in the sequel by $B(S) \equiv B(S, \mathbb{R})$ the space on bounded real valued functions on a space $S$.

**Definition 3.1.1** A *Markov transition function*, $P_t$ is a family of kernels $P_t : S \times \mathcal{B}(S) \to [0, 1]$ with the following properties:

(i) For each $t \geq 0$ and $x \in S$, $P_t(x, \cdot)$ is a measure on $(S, \mathcal{B})$ with $\mathbb{P}_t(x, S) \leq 1$.
(ii) For each $A \in \mathcal{B}$, and $t \in \mathbb{R}_+$, $P_t(\cdot, A)$ is a $\mathcal{B}$-measurable function on $S$.
(iii) For any $t, s \geq 0$,
$$P_{s+t}(x, A)) = \int P_t(y, A) P_s(x, dy). \tag{3.1}$$

**Definition 3.1.2** Then, a stochastic process $X$ with state space $S$ and index set $\mathbb{R}$ is a continuous time homogeneous Markov process with law $\mathcal{P}$ on a filtered space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$ with transition function $P_t$, if it is adapted to $\mathcal{F}_t$ and, for all bounded $\mathcal{B}$-measurable functions $f$, $t, s \in \mathbb{R}_+$,
$$\mathbb{E}\left[f(X_{t+s})|\mathcal{F}_s\right](\omega) = (P_t f)(X_s(\omega)), \text{ a.s.}. \tag{3.2}$$

It will be very convenient to think of the transition kernels as bounded linear operators on the space of bounded measurable functions on $S$, $B(S, \mathbb{R})$, acting as
$$(P_t f)(x) \equiv \int_S P_t(x, dy) f(y). \tag{3.3}$$

The Chapman-Kolmogorov equations (iii) then take the simple form $P_s P_t = P_{t+s}$. $P_t$ can then be seen as a *semi-group* of bounded linear operators. Note that we also have the dual action of $P_t$ on the space of probability measures via
$$(\mu P_t)(A) \equiv \int_S \mu(dx) P_t(x, A). \tag{3.4}$$

Of course we then have the duality relation
$$(\mu P_t)(f) = \int_S \mu(dx)(P_t f)(x) = \mu\left(P_t f\right),$$
for $f \in B(S, \mathbb{R})$.

**Remark 3.1.1** The conditions $\mathbb{P}_t(x, S) \leq 1$ may look surprising, since you would expect $\mathbb{P}_t(x, S) = 1$; the latter is in fact the standard case, and is sometimes called an “honest” transition function. However, one will want to deal with the case when probability is lost, i.e. when the process can "die". In fact, there are several scenarios where this is useful. First, if our state space is not compact, we may want to allow for our processes to *explode*, resp. go to infinity *in finite time*. Such phenomena happen in deterministic dynamical systems, and it would be too restrictive to to exclude this option for Markov chains, which we think of as *stochastic dynamical systems*. Another situation concerns open state spaces with boundaries where we want to stop the process upon arrival at the boundary. Finally, we might want to consider processes that *die* with certain rates out of pure spite.

In all these situations, it is useful to consider a compactification of the state space by adjoining a so-called *coffin state*, usually denoted by $\partial$. This state will always be considered absorbing. A dishonest transition function then becomes honest if considered extended to the space $S \cup \partial$. These extensions will sometimes be called $P_t^\partial$. To be precise, we will set

(i) $P_t^\partial(x, A) \equiv P_t(x, A)$, for $x \in S, A \in \mathcal{B}(S)$,
(ii) $P_t^\partial(\partial, \partial) = 1$,
(iii) $P_t^\partial(x, \partial) = 1 - P_t(x, S)$.

We will usually not distinguish the semi-group and its honest extension when talking about $S^\partial$-valued processes.

It is not hard to see, by somewhat tedious writing, that the transition functions (and an initial distribution) allow to express finite dimensional marginals of the law of the Markov process. This also allows to construct a process on the level of the Daniell-Kolmogorov theorem. The really interesting questions in continuous time, however, require path properties. Given a semi-group, can we construct a Markov process with càdlàg paths? Does the strong Markov property hold? We will see that this will involve analytic regularity properties of the semi-groups.

Another issue is that semi-groups are somewhat complicated and in almost no cases (except some Gaussian processes, like Brownian motion) can they be written down explicitly. In the case of discrete time we have seen the rôle played by the generator (respectively one-step transition probabilities). The corresponding object, the infinitesimal generator of the semi-group, will be seen to play an even more important rôle here. In fact, our goal in this section is to show how and when we can characterize and construct a Markov process by specifying a generator. This is fundamental for applications, since we are more likely to be able to describe the law of the instantaneous change of the state of the system, then its behavior at all times. This is very similar to the theory of differential equations: there, too, the modeling input is the prescription of the instantaneous change of state, described by specifying some derivatives, and the task of the theory is to compute the evolution at later times.

Eq. (3.1) allows us to think of Markov kernels as operators on the Banach space of bounded measurable functions.

**Definition 3.1.3** A family, $P_t$ of bounded linear operators on $B(S, \mathbb{R})$ is called *sub-Markov semi-group* , if for all $t \geq 0$,

(i) $P_t : B(S, \mathbb{R}) \to B(S, \mathbb{R})$;
(ii) if $0 \leq f \leq 1$, then $0 \leq P_t f \leq 1$;
(iii) for all $s > 0$, $P_{t+s} = P_t P_s$;
(iv) if $f_n \downarrow 0$, then $P_t f_n \downarrow 0$.

A sub-Markov semigroup is called *normal* if $P_0 = 1$. It is called *honest*, if, for all $t \geq 0$, $P_t 1 = 1$.

**Exercise.** Verify that the transition functions of Brownian motion (Eq. (6.18) in [2]) define a honest normal semi-group.

In the sequel we assume that $P_t$ is *measurable* in the sense that the map $(x, t) \to P_t(x, A)$, for any $A \in \mathcal{B}$, is $\mathcal{B}(S) \times \mathcal{B}(\mathbb{R}_+)$-measurable.

Let us now assume that $P_t$ is a family of Markov transition kernels. Then we may define, for $\lambda > 0$, the *resolvent*, $R_\lambda$, by

$$(R_\lambda f)(x) \equiv \int_0^\infty e^{-\lambda t} (P_t f)(x) dt = \int_S R_\lambda(x, dy) f(y), \tag{3.5}$$

where the *resolvent kernel*, $R_\lambda(x, dy)$, is defined as

$$R_\lambda(x, A) \equiv \int_0^\infty e^{-\lambda t} P_t(x, A) dt. \tag{3.6}$$

The following properties of a *sub-Markovian resolvent* are easily established:

(i) For all $\lambda > 0$, $R_\lambda$ is a bounded operator from $B(S, \mathbb{R})$ to $B(S, \mathbb{R})$;
(ii) if $0 \leq f \leq 1$ then $0 \leq R_\lambda f \leq \lambda^{-1}$;
(iii) for $\lambda, \mu > 0$,

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu; \tag{3.7}$$

(iv) if $f_n \downarrow 0$, then $R_\lambda f_n \downarrow 0$.

Moreover, if $P_t$ is honest, then $R_\lambda 1 = \lambda^{-1}$, for all $\lambda > 0$.

Eq. (3.7) is called the *resolvent identity.* To prove it, use the identity

$$\int e^{-\lambda s} e^{-\mu t} f(s+t) ds dt = \int \frac{e^{-\lambda u} - e^{-\mu u}}{\mu - \lambda} f(u) du.$$

Our immediate aim will be to construct the generator of the semi-group. Let us see how this goes formally. We search an operator, $G$, such that $P_t = \exp(tG)$, where exp is the usual exponential map, defined e.g. through its Taylor expansion. Then, formally, we see that

$$R_\lambda = \int_0^\infty e^{-\lambda t} e^{Gt} dt = \frac{1}{\lambda - G}. \tag{3.8}$$

This should make sense, because $e^{Gt}$ is bounded, so that the integral converges at infinity. Finally, we can recover $G$ from $R_\lambda$: set

$$G_\lambda \equiv \lambda(\lambda R_\lambda - 1) = \frac{G}{1 - G/\lambda};$$

formally, at least $G_\lambda \to G$, if $\lambda \uparrow \infty$.

While the above discussion makes sense only for bounded $G$, we can define, for $\lambda > 0$, $\exp(tG_\lambda)$, since $G_\lambda$ is bounded, and we will see that (under certain circumstances, $\exp(tG_\lambda) \to P_t$, as $\lambda \uparrow \infty$.

### 3.1.2 Strongly continuous contraction semi-groups

These manipulations become rigorous in the context of so called *strongly continuous contraction semi-groups* and constitute the famous Hille-Yosida theorem.

**Definition 3.1.4** Let $B_0$ be a Banach space. A family, $P_t : B_0 \to B_0$, of bounded linear operators is called a *strongly continuous contraction semigroup* if the following conditions are verified:

(i) for all $f \in B_0$, $\lim_{t\downarrow 0} \|P_t f - f\| = 0$:
(ii) $\|P_t\| \leq 1$, for all $t \geq 0$;
(iii) $P_t P_s = P_{t+s}$, for all $t, s \geq 0$.

Here $\|\cdot\|$ denotes the operator norm corresponding to the norm on $B_0$.

**Lemma 3.1.1** *If $P_t$ is a strongly continuous contraction semigroup, then, for any $f \in B_0$, the map $t \to P_t f$ is continuous.*

*Proof.* Let $t \geq s \geq 0$. We need to show that $P_t f - P_s f$ tends to zero in norm as $t - s \downarrow 0$. But

$$\|P_t f - P_s f\| = \|P_s(P_{t-s} f - f)\| \leq \|P_{t-s} f - f\|,$$

which tends to zero by property (i). Note that we needed all three defining properties!. □

Note that continuity allows to define the resolvent through a (limit of) Riemann integrals,

$$R_\lambda f \equiv \lim_{T\uparrow\infty} \int_0^T e^{-\lambda t} P_t f.$$

The inherited properties if such an $R_\lambda$ are now used to *define* a strongly continuous contraction resolvent.

**Definition 3.1.5** Let $B$ be a Banach space, and let $R_\lambda$, $\lambda > 0$, be a family of bounded linear operators on $B$. Then $R_\lambda$ is called a *contraction resolvent*, if

(i) $\lambda \|R_\lambda\| \leq 1$, for all $\lambda > 0$;
(ii) the resolvent identify (3.7) holds.

A contraction resolvent is called *strongly continuous*, if in addition

(iii) $\lim_{\lambda\uparrow\infty} \|\lambda R_\lambda f - f\| = 0$.

**Exercise.** Verify that the resolvent of a strongly continuous contraction semi-group is a strongly continuous contraction resolvent.

**Lemma 3.1.2** *Let $R_\lambda$ be a contraction resolvent on $B_0$. Then the the range of $R_\lambda$ is independent of $\lambda$, and that the closure of its range coincides with the space of functions, $h$, such that $\lambda R_\lambda h \to h$, as $\lambda \uparrow \infty$.*

*Proof.* Both observations follow from the resolvent identity. Let$\mu, \lambda > 0$, then $R_\mu = R_\lambda(1 + (\lambda - \mu)R_\mu$. Thus, if $g$ is in the range of $R_\mu$, then it is also in the range of $R_\lambda$: if $g = R_\mu f$, then $g = R_\lambda h$, where $h = (1 + (\lambda - \mu)R_\mu)f$! Denote the common range of the $R_\lambda$ by $\mathcal{R}$.

Moreover, if $h \in \mathcal{R}$, then $h = R_\mu g$, and so

$$(\lambda R_\lambda - 1)h = (\lambda R_\lambda - 1)R_\mu g = \frac{\mu}{\lambda - \mu} R_\mu g - \frac{\lambda R_\lambda}{\lambda - \mu} g$$

Since $\lambda R_\lambda$ is bounded, it follows that the right-hand side tends to zero, as $\lambda \uparrow \infty$. Also, if $h$ is in the closure of $\mathcal{R}$, the there exist $h_n \in \mathcal{R}$, such that $h_n \to h$; then

$$\|\lambda R_\lambda h - h\| \leq \|\lambda R_\lambda h_n - h_n\| + \|h_n - h\| + \|\lambda R_\lambda (f - f_n)\|,$$

and since $\lambda R_\lambda$ is a contraction, the right hand side can be made as small as desired by letting $n$ and $\lambda$ tend to infinity. Finally, it is clear that if $h = \lim_{\lambda \uparrow \infty} R_\lambda h$, then $h$ must be in the closure of $\mathcal{R}$. □

As a consequence, the restriction of a contraction resolvent to the closure of its range is strongly continuous. Moreover, for a strongly continuous contraction resolvent, the closure of its range is equal to $B_0$, and so the range of $R_\lambda$ is dense in $B_0$.

We now come to the definition of an infinitesimal generator.

**Definition 3.1.6** Let $B_0$ be a Banach space and let $P_t$, $t \in \mathbb{R}_+$ be a strongly continuous contraction semigroup. We say that $f$ is in the domain of $G$, $\mathcal{D}(G)$, if there exists a function $g \in B_0$, such that

$$\lim_{t \downarrow 0} \|t^{-1}(P_t f - f) - g\| = 0. \tag{3.9}$$

For such $f$ we set $Gf = g$ if $g$ is the function that satisfies (3.9).

**Remark 3.1.2** Note that we define the domain of $G$ at the same time as $G$. In general, $G$ will be an unbounded (e.g. a differential) operator whose domain is strictly smaller than $B_0$. Some authors (e.g. [6]) describe the generator of a Markov process as a collections of the pairs of functions $(f, g)$ satisfying (3.9).

The crucial fact is that the resolvent is related to the generator in the way anticipated in (3.8).

**Lemma 3.1.3** *Let $P_t$ be a strongly continuous contraction semigroup on $B_0$. Then the operators $R_\lambda$ and $(\lambda - G)$ are inverses.*

*Proof.* Let $g \in B_0$ and let $f = R_\lambda g$. We want to show that $(\lambda - G) f = g$, i.e. that (3.9) holds for pairs of functions $f$ and $\lambda f - g$ where $f$ is in the range of $\mathcal{R}_\lambda$. But

$$\lambda f - t^{-1}(P_t f - f) = t^{-1}(f(1 + \lambda t) - P_t f)$$

As $t \downarrow 0$, we may replace $(1 + \lambda t)$ by $e^{\lambda t}$ and write

$$\lim_{t \downarrow 0} \lambda f - t^{-1}(P_t f - f) = \lim_{t \downarrow 0} e^{\lambda t} t^{-1}(R_\lambda g - e^{-\lambda t} P_t R_\lambda g)$$

Now

$$e^{-\lambda t}P_tR_\lambda g = \int_0^\infty e^{-\lambda(t+s}P_{t+s}gds = \int_t^\infty e^{-\lambda s}P_sgds,$$

and so

$$t^{-1}(R_\lambda g - e^{-\lambda t}P_tR_\lambda g) = t^{-1}\int_0^t e^{-\lambda s}P_sgds.$$

By continuity of $P_t$, the latter expression converges to $g$, as $t \downarrow 0$, so we have shown that $(\lambda - G)R_\lambda g = g$, and that $R_\lambda g \in \mathcal{D}(G)$.

Next we take $f \in \mathcal{D}(G)$. Then $\varepsilon^{-1}(P_{t+\varepsilon}f - P_tf) = P_t(\varepsilon^{-1}(P_\varepsilon f - f) \to P_tGf$. Thus,

$$\frac{d}{dt}P_tf = P_tGf.$$

Integrating this relation gives that

$$P_tf - f = \int_0^t P_sGfds.$$

Multiplying with $e^{-\lambda t}$ and integrating gives

$$R_\lambda f - \lambda^{-1}f = \lambda^{-1}R_\lambda Gf,$$

which shows that for $f \in \mathcal{D}(G)$, $R_\lambda(\lambda - G)f = f$, and in particular $f \in \mathcal{R}$. Thus $\mathcal{D}(G) = \mathcal{R}$. This concludes the proof of the lemma. □

### 3.1.3 The Hille-Yosida theorem

We now prove the fundamental theorem of Hille and Yosida that allows us to construct a semi-group from the resolvent.

**Theorem 3.1.4** *Let $R_\lambda$ be a strongly continuous contraction resolvent on a Banach space $B_0$. Then there exists a* unique *strongly continuous contraction semi-group, $P_t$, $t \in \mathbb{R}$, on $B_0$, such that, for all $\lambda > 0$ and all $f \in B_0$,*

$$\int_0^\infty e^{-\lambda t}P_tfdt = R_\lambda f. \tag{3.10}$$

*Moreover, if*

$$G_\lambda \equiv \lambda(\lambda R_\lambda - 1) \tag{3.11}$$

*and*

$$P_{t,\lambda} \equiv \exp\left(tG_\lambda\right), \tag{3.12}$$

*then*

$$P_t f = \lim_{\lambda\uparrow\infty} P_{t,\lambda} f. \tag{3.13}$$

*Proof.* When proving the Hille-Yosida theorem we must take care not to assume the existence of a semi-group. So we want to rely essentially on the resolvent identity.

We have seen before that the range, $\mathcal{R}$, of $R_\lambda$ is independent of $\lambda$ and dense in $B_0$, due to the assumption of strong continuity. Now we want to show that $R_\lambda$ is a bijection. Note that we cannot use Lemma 3.1.3 here because in its prove we used the existence of $P_t$. Namely, let $h \in B_0$ such that $R_\lambda h = 0$. Then,by the resolvent identity,

$$R_\mu h = (1 - (\lambda - \mu)R_\mu)R_\lambda h = 0,$$

for every $\mu$. But by strong continuity, $\lim_{\mu\uparrow\infty} \mu R_\mu h = h$, so we must have that $h = 0$.

Therefore, there exists an inverse, $R_\lambda^{-1}$, of $R_\lambda$, with domain equal to $\mathcal{R}$, such that for all $h \in B_0$, $R_\lambda^{-1} R_\lambda h = h$, and for $g \in \mathcal{R}$, $R_\lambda R_\lambda^{-1} g = g$. Moreover, by the resolvent identity,

$$R_\lambda R_\mu^{-1} = (R_\mu + (\mu - \lambda)R_\lambda R_\mu)R_\mu^{-1} = 1 + (\mu - \lambda)R_\lambda.$$

Thus

$$R_\mu^{-1} - (\mu - \lambda) = R_\lambda^{-1}, \tag{3.14}$$

which we may rewrite as

$$R_\lambda^{-1} - \lambda = R_\mu^{-1} - \mu \equiv -G \tag{3.15}$$

in other words, there exists an operator $G$ with domain $\mathcal{D}(G) = \mathcal{R}$, such that, for all $\lambda$,

$$\frac{1}{\lambda - G} = R_\lambda. \tag{3.16}$$

We now show the following lemma:

**Lemma 3.1.5** *Let $G_\lambda$ be defined in (3.11). Then, $f \in \mathcal{D}(G)$ if and only if*

$$\lim_{\lambda\uparrow\infty} G_\lambda f \equiv g$$

*exists. Then $Gf = g$.*

*Proof.* Let first $f \in \mathcal{D}(G)$. Then

$$G_\lambda f = \lambda(\lambda R_\lambda - 1)f = \lambda R_\lambda(\lambda - R_\lambda^{-1})f = \lambda R_\lambda G f,$$

and by strong continuity, $\lim_{\lambda\uparrow\infty} \lambda R_\lambda G f = G f$, as claimed.
Assume now that $\lim_{\lambda\uparrow\infty} G_\lambda f = g$. The by the resolvent identity,

$$R_\mu G_\lambda f = \lambda\left(\frac{\mu R_\mu - \lambda R_\lambda}{\lambda - \mu}\right) f = \frac{\lambda\mu}{\lambda-\mu} R_\mu f - \frac{\lambda}{\lambda-\mu}\lambda R_\lambda f.$$

As $\lambda \uparrow \infty$, the right-hand side clearly tends to $\mu R_\mu f - f$, while the left hand side, by assumption, tends to $R_\mu g$. Hence,

$$f = \mu R_\mu f - R_\mu g = R_\mu(\mu f - g).$$

Therefore, $f \in \mathcal{R}$, and

$$Gf = (\mu - R_\mu^{-1}) R_\mu(\mu f - g) = \mu f - R_\mu^{-1} R_\mu(\mu f - g) = \mu f - \mu f + g = g.$$

□

We now continue the proof of the theorem. Note that $G_\lambda$ is bounded, and so by the standard properties of the exponential map, we have the following three facts:

(i) $P_{t,\lambda} P_{s.\lambda} = P_{t+s,\lambda}$.
(ii) $\lim_{t\downarrow 0} t^{-1}(P_{t,\lambda} - 1) = G_\lambda$.
(iii) $P_{t,\lambda} - 1 = \int_0^t P_{s,\lambda} G_\lambda ds$.

Moreover, since $\|\lambda R_\lambda\| \leq 1$, from the definition of $P_{t,\lambda}$ it follows that

$$\|P_{t,\lambda}\| \leq e^{-\lambda t} e^{t\lambda \|\lambda R_\lambda|} \leq 1.$$

Now the resolvent identity implies that the operators $R_\lambda$ and $R_\mu$ commute for all $\lambda, \mu > 0$, and so all derived operators commute. Thus we have the telescopic expansion

$$\begin{aligned} P_{t,\lambda} - P_{t,\mu} &= P_{t,\lambda} P_{0,\mu} - P_{0,\lambda} P_{t,\mu} \\ &= \sum_{k=1}^n \left( P_{kt/n,\lambda} P_{(n-k)t/n,\mu} - P_{(k-1)t/n,\lambda} P_{(n-k+1)t/n,\mu} \right) \\ &= \sum_{k=1}^n P_{(k-1)t/n,\lambda} P_{(n-k)t/n,\mu} \left( P_{t/n,\lambda} - P_{t/n,\mu} \right). \end{aligned} \tag{3.17}$$

By the bound on $\|P_{t,\lambda}\|$, it follows that for any $f \in B_0$,

$$\begin{aligned} \|P_{t,\lambda} f - P_{t,\mu} f\| &\leq n \|P_{t/n,\lambda} f - P_{t/n,\mu} f\| \\ &= n \left\| \left(P_{t/n,\lambda} - 1\right) f - \left(P_{t/n,\mu} - 1\right) f \right\|. \end{aligned}$$

Passing to the limit $n \uparrow \infty$, and using (ii), we conclude that

$$\|P_{t,\lambda} f - P_{t,\mu} f\| \leq t \|G_\lambda f - G_\mu f\|. \tag{3.18}$$

This implies the existence of $\lim_{\lambda \uparrow \infty} P_{t,\lambda} f \equiv P_t f$ whenever $\lim_{\lambda \uparrow \infty} G_\lambda f$ exists, hence by Lemma 3.1.5 for all $f \in \mathcal{D}(G)$. Moreover, the convergence is uniform in $t$ on compact sets, so the map $t \to P_t f$ is continuous. Since $\mathcal{D}(G) = \mathcal{R}$ is dense in $B_0$, and $\mathbb{P}_t^\lambda$ are uniformly bounded in norm, these results in fact extends to all functions $f \in B_0$.

It remains to show that (3.10) holds. To do so, note that

$$\int_0^\infty e^{-\lambda t} P_{t,\mu} f dt = \int_0^\infty e^{-t(\lambda - G_\mu)} f dt = \frac{1}{\lambda - G_\mu} f$$

As $\mu$ tends to infinity, the left-hand side converges to $\int_0^\infty e^{-\lambda t} P_t f$, and, using the resolvent identity, the right hand side is shown to tend to $R_\lambda f$. This concludes the prove of the theorem. □

The Hille-Yosida theorem clarifies how a strongly continuous contraction semi-group can be recovered from a resolvent. To summarize where we stand, the theorem asserts that if we have a strongly continuous contraction resolvent family, $R_\lambda$, then there exists a unique operator, $G$, such that $R_\lambda = (\lambda - G)^{-1}$, that is the generator of a unique strongly continuous contraction semi-group, $P_\lambda$.

One might rightly ask if we can *start* from a generator: of course, the answer is yes: if we have linear operator, $G$, with $\mathcal{D}(G) \subset B_0$, this will generate a strongly continuous contraction semi-group, if the operators $(\lambda - G)^{-1}$ exist for all $\lambda > 0$ and form a strongly continuous contraction resolvent family.

One may not be quite happy with this answer, which leaves a lot to verify. It would seem nicer to have a characterization of when this is true in terms of direct properties of the operator $G$.

In the next theorem (sometimes also called the Hille-Yosida theorem, see [6]), formulates such conditions.

**Theorem 3.1.6** *A linear operator, $G$, on a Banach space, $B_0$, is the generator of a strongly continuous contraction semi-group, if and only if the following hold:*

*(i) The domain of $G$, $\mathcal{D}(G)$, is dense in $B_0$.*

*(ii) $G$ is* dissipative, *i.e. for all $\lambda > 0$ and all $f \in \mathcal{D}(G)$,*

$$\|(\lambda - G) f\| \geq \lambda \|f\|. \tag{3.19}$$

*(iii) There exists a $\lambda > 0$ such that* $\mathbf{range}(\lambda - G) = B_0$.

*Proof.* By theorem 3.1.4, we just have to show that the family $(\lambda-G)^{-1}$ is a strongly continuous contraction resolvent, if and only if (i)–(iii) hold. In fact, we have seen that properties (i)–(iii) are satisfied by the generator associated to a strongly continuous contraction resolvent: (i) was shown at the beginning of the proof of Thm. 3.1.4, (ii) is a consequence of the bound $\|\lambda R_\lambda\| \leq 1$: Note that

$$1 \geq \sup_{f \in B_0} \frac{\|\lambda R_\lambda f\|}{\|f\|} \geq \sup_{g \in \mathcal{D}(G)} \frac{\|\lambda R_\lambda (\lambda - G) g\|}{\|(\lambda - G)g\|} = \sup_{g \in \mathcal{D}(G)} \frac{\lambda \|g\|}{\|(\lambda - G)g\|}.$$

Finally, since for any function $f \in B_0$,

$$(\lambda - G) R_\lambda f = f,$$

any such $f$ is in the range of $(\lambda - G)$.

It remains to show that these conditions are sufficient, i.e. that under them, if $R_\lambda \equiv (\lambda - G)^{-1}$ is a strongly continuous contraction resolvent.

We need to recall a few notions from operator theory.

**Definition 3.1.7** A linear operator, $G$, on a Banach space, $B_0$, is called *closed*, if and only if its *graph*, the set

$$\Gamma(G) \equiv \{(f, Gf) : f \in \mathcal{D}(G)\} \subset B_0 \times B_0 \tag{3.20}$$

is closed in $B_0 \times B_0$. Equivalently, $G$ is closed if for any sequence $f_n \in \mathcal{D}(G)$ such that $f_n \to f$ and $Gf_n \to g$, $f \in \mathcal{D}(G)$ and $g = Gf$.

**Definition 3.1.8** If $G$ is a closed operator on $B_0$, then a number $\lambda \in \mathbb{C}$ is an element of the *resolvent set*, $\rho(G)$, of $G$, if and only if

(i) $(\lambda - G)$ is one-to-one;
(ii) $\mathbf{range}(\lambda - G) = B_0$,
(iii) $R_\lambda \equiv (\lambda - G)^{-1}$ is a bounded linear operator on $B_0$.

It comes as no surprise that whenever $\lambda, \mu \in \rho(G)$, then the resolvents $R_\lambda, R_\mu$ satisfy the resolvent identity. (**Exercise:** Prove this!).

Another important fact is that if for some $\lambda \in \mathbb{C}$, $\lambda \in \rho(G)$, then there exists a neighborhood of $\lambda$ that is contained in $\rho(G)$. Namely, if $|\lambda - \mu| < 1/\|R_\lambda\|$, then the series

$$\widehat{R}_\mu \equiv \sum_{n=0}^{\infty} (\lambda - \mu)^n R_\lambda^{n+1}$$

converges and defines a bounded operator. Moreover, for $g \in \mathcal{D}(G)$, a simple computation shows that

$$\widehat{R}_\mu (\mu - G) g = g,$$

and for any $f \in B_0$,

$$(\mu - G)\widehat{R}_\mu f = f.$$

Hence $\widehat{R}_\mu = (\mu - G)^{-1}$, $\mathbf{range}(\mu - G) = B_0$, and so $\mu \in \rho(G)$. Thus, $\rho(G)$ is an open set.

We will first show that (i) and (ii) imply that $G$ is closed.

**Lemma 3.1.7** *Let $G$ be a dissipative operator and let $\lambda > 0$ be fixed. Then $G$ is closed if and only if* $\mathbf{range}(\lambda - G)$ *is closed.*

*Proof.* Let us first show that the range of $(\lambda - G)$ is closed if $G$ is closed. Take $f_n \in \mathcal{D}(G)$ and assume that $(\lambda - G)f_n \to h$. Since $G$ is dissipative, $\|(\lambda - G)(f_n - f_{n+k})\| \geq \lambda\|f_n - f_{n+k}\|$, so $f_n$ is a Cauchy sequence, and by closedness, there exists a limit $f = \lim_n f_n \in \mathcal{D}(G)$. Thus $\lim_n Gf_n = \lambda f - h$. But also, $\lim_n Gf_n = Gf$, so $h = (\lambda - G)f$, i.e. $\lim_n(\lambda - G)f_n$ is in the range of $(\lambda - G)$ and so the range is closed. On the other hand, if $\mathbf{range}(\lambda - G)$ is closed, then take some $\mathcal{D}(G) \ni f_n \to f$ and $Gf_n \to g$. Then $(\lambda - G)f_n \to \lambda f - g$ in the range of $(\lambda - G)$. Thus there exists $f_0 \in \mathcal{D}(G)$, such that

$$(\lambda - G)f_0 = \lambda f - g.$$

But since $G$ is dissipative, if $(\lambda - G)f_n \to (\lambda - G)f_0$, then $f_n \to f_0$, so $f_0 = f$. Hence $(\lambda - G)f = \lambda f - g$, or $Gf = g$. Hence $f$ is in the domain and $g$ in the range of $G$, so $G$ is closed. □

It follows that if the range of $(\lambda - G)$ is closed for some $\lambda > 0$, then it is closed for all $\lambda > 0$.

The next lemma establishes that the resolvent set of a closed dissipative operator contains $(0, \infty)$, if some point in $(0, \infty)$ is in the resolvent set.

**Lemma 3.1.8** *If $G$ is a closed dissipative operator on $B_0$, then the set $\rho^+(G) \equiv \rho(G) \cap (0, \infty)$ is either empty or equal to $(0, \infty)$.*

*Proof.* We will show that $(0, \infty)$ is open and closed in $(0, \infty)$. First, since $\rho(G)$ is open, its intersection with $(0, \infty)$ is relatively open. Let now $\lambda_n \in \rho^+(G)$ and $\lambda_n \to \lambda \in (0, \infty)$. For any $g \in B_0$, and any $n$ we can define $g_n = (\lambda - G)R_{\lambda_n} g$. Then

$$\begin{aligned} \|g_n - g\| &= \|(\lambda - G)R_{\lambda_n} g - (\lambda_n - G)R_{\lambda_n} g\| = \|(\lambda - \lambda_n)R_{\lambda_n} g\| \\ &\leq \lambda_n^{-1}(\lambda - \lambda_n)\|g\| \end{aligned}$$

which tends to zero as $n \uparrow \infty$. Note that the inequality used the dissipativity of $G$. Therefore, the range of $(\lambda - G)$ is dense in $B_0$; but from the preceding lemma we know that the range of $(\lambda - G)$ is closed. Hence **range**$(\lambda - G) = B_0$. But since $G$ is dissipative, if $\|f - g\| > 0$, then $\|(\lambda - G)f - (\lambda - G)g\| > 0$, and so $(\lambda - G)$ is one-to one. Finally, for any $g \in B_0$, $f = (\lambda - G)^{-1}g$ is in $\mathcal{D}(G)$. Then dissipativity shows that

$$\|g\| = \|(\lambda - G)f\| \geq \lambda\|f\| = \lambda\|(\lambda - G)^{-1}g\|,$$

so that $(\lambda - G)^{-1}$ is bounded by $\lambda^{-1}$ on $B_0$. Thus $\lambda \in \rho^+(G)$, and hence $\rho^+(G)$ is closed. □

We now continue with the proof of the theorem. We know from (ii) and (iii) and Lemma 3.1.7 that $G$ is closed and **range**$(\lambda - G) = B_0$ for all $\lambda > 0$. Moreover, just as in the proof of Lemma 3.1.8, dissipativity implies then that $\rho^+(G) = (0, \infty)$. Also as in that proof, we get the bound $\lambda\|R_\lambda\| \leq 1$. As we have already explained, the resolvent identity holds for all $\lambda > 0$, so $R_\lambda$ is a contraction resolvent family.

All what remains to prove is the strong continuity. Let first $f \in \mathcal{D}(G)$. Then we can write

$$\|\lambda R_\lambda f - f\| = \lambda\|R_\lambda(f - \lambda^{-1}(\lambda - G)f)\| \leq \lambda^{-1}\|Gf\|.$$

Since $f \in \mathcal{D}(G)$, $Gf \in B_0$, and $\|Gf\| < \infty$, so the right hand side tends to zero as $\lambda \uparrow \infty$.

Thus $\lambda R_\lambda f \to f$ for all $f$ in $\mathcal{D}(G)$. For general $f$, since $\mathcal{D}(G)$ is dense in $B_0$, take a sequence $f_n \in \mathcal{D}(G)$ such that $f_n \to f$. Then,

$$\|\lambda R_\lambda f - f\| \leq \|\lambda R_\lambda(f - f_n)\| + \|\lambda R_\lambda f_n - f_n\| + \|f - f_n\|$$

and so

$$\limsup_{\lambda\uparrow\infty} \|\lambda R_\lambda f - f\| \leq 2\|f - f_n\|.$$

Since the right-hand side can be made as small as desired by taking $n \uparrow \infty$, it follows that $\|\lambda R_\lambda f - f\| \to 0$, as claimed. Thus $R_\lambda \equiv (\lambda - G)^{-1}$ is a strongly continuous contraction resolvent family, and the theorem is proven. □

One may find the the conditions (i)–(iii) of Theorem 3.1.6 are just as difficult to verify then those of Theorem 3.1.4. In particular, it does not seem easy to check whether an operator is dissipative.

The following lemma, however, can be very helpful.

**Lemma 3.1.9** *Let $S$ be a complete metric space. A linear operator, $G$, on $C_0(S)$ is dissipative, if for any $f \in \mathcal{D}(G)$, if $y \in S$ is such that $f(y) = \max_{x\in S} f(x)$, then $Gf(y) \leq 0$.*

*Proof.* Since $f \in C_0(S)$ vanishes at infinity, there exists $y$ such that $|f(y)| = \|f\|$. Assume without loss of generality that $f(y) \geq$, so that $f(y)$ is a maximum. For $\lambda \geq 0$, let $g \equiv f - \lambda^{-1}Gf$. Then

$$\max_x f(x) = f(y) \leq f(y) - \lambda^{-1}Gf(y) = g(y) \leq \max_x g(x).$$

Since the same holds for the function $-f$, we also get that

$$\min_x f(x) \geq \min_x g(x),$$

and hence $G$ is dissipative. □

**Examples** We can verify the conditions of Theorem 3.1.6 in some simple examples.

- Let $S = [0,1]$, $G = \frac{1}{2}\frac{d^2}{dx^2}$, $\mathcal{D}(G) = \{f \in C^2([0,1] : f'(0) = f'(1) = 0\}$. Since here $S$ is compact, clearly any continuous function takes on its minimum at some point $y \in [0,1]$. If $y \in (0,1)$, then clearly $\frac{1}{2}\frac{d^2}{dx^2}f(y) = 0$; if $y = 0$, for 0 to be a minimum, since $f'(0) = 0$, the second derivative must be non-negative; the same is true if $y = 1$. Thus $G$ is dissipative.

  The fact the $\mathcal{D}(G)$ is dense is clear from the definition. To show that the range of $\lambda - G$ is $C([0,1])$, we must show that the equation

  $$\lambda f - \frac{1}{2}f'' = g \tag{3.21}$$

  with boundary conditions $f'(0) = f'(1) = 0$ has a solution for all $g \in B([0,1])$. Such a solution can be written down explicitely. In fact, (we just consider the case $\lambda = 1$, which is enough)

  $$f(x) = -2e^{\sqrt{2}x}\int_0^x e^{-\sqrt{2}t}\int_0^t g(s)dsdt + K\sinh(\sqrt{2}x) \tag{3.22}$$

  with

  $$K\sinh\sqrt{2} \equiv -2e^{\sqrt{2}}\int_0^1 e^{-\sqrt{2}t}\int_0^t g(s)dsdt$$

  is easily verified to solve this problem uniquely.

(ii) The same operator as above, but replace $[0,1]$ with $\mathbb{R}$ and $\mathcal{D}(G) = C_b^2(\mathbb{R})$. We first show that the range of $R_\lambda$ is contained in $C_b^2(\mathbb{R})$. Let $f$ be given by $f = R_\lambda g$ with $g \in B(\mathbb{R})$. $R_\lambda$ is the resolvent corresponding to the Gaussian transition kernel

$$P_t(x, dy) \equiv \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} dy.$$

Thus

$$f(x) \equiv (R_\lambda g)(x) = \int_0^\infty e^{-\lambda t} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} g(y) dy.$$

Now one can show that

$$\int_0^\infty e^{-\lambda t} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} = \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{2\lambda}|x-y|},$$

and so

$$f(x) = \int \frac{1}{\sqrt{2\lambda}} e^{-\sqrt{2\lambda}|x-y|} g(y) dy.$$

Hence

$$\begin{aligned} f'(x) &= \int e^{-\sqrt{2\lambda}|x-y|} \operatorname{sign} g(y) dy \qquad (3.23) \\ &= -\int_{-\infty}^x e^{-\sqrt{2\lambda}|x-y|} g(y) dy + \int_x^\infty e^{-\sqrt{2\lambda}|x-y|} g(y). \end{aligned}$$

Thus, differentiating once more,

$$\begin{aligned} f''(x) = -2g(x) &+ \sqrt{2\lambda} \int_{-\infty}^x e^{-\sqrt{2\lambda}|x-y|} g(y) dy \qquad (3.24) \\ &+ \sqrt{2\lambda} \int_x^\infty e^{-\sqrt{2\lambda}|x-y|} g(y) \\ &= -2g(x) + 2\lambda f(x). \end{aligned}$$

Hence $f \in B(\mathbb{R})$ as claimed. Moreover, $f$ solves (3.21) and thus $(\Delta/2 - \lambda)$ is the inverse of $R_\lambda$. Since this operator maps $C_b^2(\mathbb{R})$ into $B(\mathbb{R})$, we see that $C_b^2(\mathbb{R}) \subset \mathcal{D}(G)$. Hence $C_b^2(\mathbb{R}) = \mathcal{D}(G)$, $\Delta$ is closed and is the generator of our semigroup.

(iii) If we replace in the previous example $\mathbb{R}$ with $\mathbb{R}^d$, then the the result will not carry over. In fact, $\Delta$ is not a closed operator in $\mathbb{R}^d$ if $d \geq 2$. This may appear disappointing, because it says that $\frac{1}{2}\Delta$ is *not* the generator of Brownian motion in $d \geq 2$. Rather, the generator of BM will be the closure of $\frac{1}{2}\Delta$. We will come back to this issue in a systematic way when we discuss the martingale problem approach to Markov processes.

### 3.2 Feller-Dynkin processes

We will now turn to a special class of Markov semi-groups that will be seen to have very nice properties. Our setting is that the state space is a locally compact Hausdorff space with countable basis (but think of $\mathbb{R}^d$ if you like). The point is that we do not assume compactness. We will, however, consider the one-point compactification of such a space obtained by adding a "coffin state", $\partial$, ("infinity") to it. Then $S^\partial \equiv S \cup \partial$ is a compact metrisable space.

We will now place ourselves in the setting where the Hille-Yosida theorem work, and make a specific choice for the underlying Banach space, namely we will work on the the space $C_0(S)$ of continuous functions vanishing at infinity. This will actually place a restriction of the semi-groups to preserve this space. This (and similar properties) is known as the *Feller property.*

**Definition 3.2.1** A *Feller-Dynkin semigroup* is a strongly continuous sub-Markov semigroup, $P_t$, acting on the space $C_0(S)$, in particular for all $t \geq 0$,

$$P_t : C_0(S) \to C_0(S). \tag{3.25}$$

It is an analytic fact that follows from the Riesz representation theorem, that to any strongly continuous contraction semigroup corresponds a sub-Markov kernel, $P_t(x, dy)$, such that $(P_t f)(x) = \int_S P_t(x, dy) f(y)$, for all $f \in C_0(S)$.

To see this recall that the Riesz representation theorem asserts that for any linear map, $L$, from the space of continuous functions $C(S)$ there corresponds a unique measure, $\mu$, such that

$$Lf = \int_S f(y)\mu(dy).$$

If moreover $L1 = 1$, this measure will be a probability measure.

Thus for any $x \in S$, there exists a probability measure $P_t(x, dy)$, such that for any continuous function $f$

$$(P_t f)(x) = \int f(y) P_t(x, dy).$$

Since $P_t f$ is measurable, we also get that $\int f(y) P_t(x, dy)$ is measurable. Finally, using the monotone class theorem, one shows that $P_t(x, A)$ is measurable for any Borel set $A$, and hence $P_t(x, dy)$ is a probability kernel, and in fact a sub-Markov kernel.

Note that, since we are in a setting where the Hille-Yosida theorem applies and that there exists a generator, $G$, exists on a domain $\mathcal{D}(G) \subset C_0(S)$. Note that then we have for $f \in \mathcal{D}(G)$ the formula

$$Gf(x) \equiv \lim_{t \downarrow 0} t^{-1} \left( \int_S P_t(x, dy) f(y) - f(x) \right) \tag{3.26}$$

Therefore, if $f$ attains its maximum at a point $x$, then

$$\int_S P_t(x, dy) f(y) \leq f(x),$$

and so $Gf(x) \leq 0$, if $f(x) \geq 0$ (this condition is not needed if $P_t$ is honest).

Dynkin's maximum principle states that this property characterizes the domain of the generator. Let us explain what we mean by this.

**Definition 3.2.2** Let $G, C$ be two linear operators with domains $\mathcal{D}(G), \mathcal{D}(C)$, respectively. We say that $C$ is an *extension* of $G$, if

(i) $\mathcal{D}(G) \subset \mathcal{D}(C)$, and
(ii) For all $f$ in $\mathcal{D}(G)$, $Gf = Cf$.

**Lemma 3.2.10** *Let $G$ be a generator of a Feller-Dynkin semigroup and let $C$ be an extension of $G$. Assume that if $f \in \mathcal{D}(C)$ and $f$ attains its maximum in $x$ with $f(x) \geq 0$, then $Cf(x) \leq 0$. Then $G = C$.*

*Proof.* Note first that $C = G$ if $Cf = f$ implies $f = 0$. To see this, let $g \equiv f - Cf$ und $h = R_1 g$. But $R_1 g \in \mathcal{D}(G)$ and thus

$$h - Ch = h - Gh = g = f - Cf.$$

Hence $f - h = C(f - h)$, and so $f = h$. In particular $f \in \mathcal{D}(G)$.

Now let $f \in \mathcal{D}(C)$ and $Cf = f$. We see that if $f$ attains its maximum at $x$ with $f(x) \geq 0$, then under the hypothesis of the lemma, $Cf(x) \leq 0$. Since $Cf = f$, this means that $f(x) = Cf(x) = 0$. Thus $\max_y f(y) = 0$. Applying the same argument to $-f$, it follows that $\min_y f(y) = 0$. □

The now turn to the central result of this section, the existence theorem for Feller-Dynkin processes.

**Theorem 3.2.11** *Let $P_t$ be a Feller-Dynkin semigroup on $C_0(S)$. Then there exists a strong Markov process with values in $S^\partial$ and càdlàg paths and transition kernel $P_t$.*

**Remark 3.2.1** Note that the unique existence of the Markov process on the level of finite dimensional distributions does not require the Feller property.

*Proof.* First, the Daniell-Kolmogorov theorem guarantees the existence of a unique process on the product space $(S^\partial)^{\mathbb{R}_+}$, provided the finite dimensional marginals satisfy the compatibility conditions. This is easily verified just as in the discrete time case using the Chapman-Kolmogorov equations.

We now want to show that the paths of this process are regularisable, and finally that regularization entrains just a modification. For this we need to get martingales into the game.

**Lemma 3.2.12** *Let $g \in C_0(S)$ and $g \geq 0$. Set $h = R_1 g$. Then*

$$0 \leq e^{-t} P_t h \leq h. \tag{3.27}$$

*If $Y$ is the corresponding Markov process, $e^{-t}h(Y_t)$ is a supermartingale.*

*Proof.* Let us first prove (3.27). The lower bound is clear since $P_t$ and hence $R_\lambda$ map positive function to positive functions. Next

$$\begin{aligned} e^{-s} P_s h = e^{-s} P_s R_1 g = e^{-s} P_s \int_0^\infty e^{-u} P_u g du \qquad (3.28)\\ = \int_s^\infty e^{-u} P_u g du \leq R_1 g = h. \end{aligned}$$

Now $e^{-t}h(Y_t)$ is a supermartingale since

$$\mathbb{E}[e^{-s-t} h(Y_{t+s} | \mathcal{G}_t] = e^{-s-t} P_s h(Y_t) \leq e^{-t} h(Y_t),$$

where of course we used (3.27) in the last step. □

As a consequence of the previous lemma, the functions $e^{-q}h(Y_q)$ are regularisable, i.e. $\lim_{q \downarrow t} e^{-q} h(Y_q)$ exists for all $t$ almost surely.

Now we can take a countable dense subset, $g_1, g_2, \dots$, of elements of $C_0(S)$, and set $h_i = R_1 g_i$. The set $\mathcal{H} = \{h_i\}_{i \in \mathbb{N}}$ separates points in $S^\partial$, while almost surely, $e^{-q} h_i(Y_q)$ is regularisable for all $i \in \mathbb{N}$. But then $X_t \equiv \lim_{q \downarrow t} Y_q$ exists for all $t$, almost surely and is a càdlàg process.

Finally we establish that $X$ is a modification of $Y$. To do this, let $f, g \in C_0(S)$. Then

$$\mathbb{E}[f(Y_t)g(X_t)] = \lim_{q \downarrow t} \mathbb{E}[f(Y_t)g(Y_q)] = \lim_{q \downarrow t} \mathbb{E}[f(Y_t) P_{t-q} g(Y_t)] = \mathbb{E}[f(Y_t)g(Y_t)]$$

where the first inequality used the definition of $X_t$ and the third the strong continuity of $P_t$. By an application of the monotone class theorem, this implies that $\mathbb{E}[f(Y_t, X_t)] = \mathbb{E}[f(Y_t, Y_t)]$ for any bounded measurable function on $S^\partial \times S^\partial$, and hence in particular $\mathbb{P}[X_t = Y_t] = 1$. □

The previous theorem allows us to henceforth consider Feller-Dynkin Markov processes defined on the space of càdlàg functions with values in $S^\partial$ (with the additional property that, if $X_t = \partial$ or $X_{t-} = \partial$, then $X_s = \partial$ for all $s \geq t$). We will henceforth think of our Markov processes as defined on that space (with the usual right-continuous filtration).

### 3.3 The strong Markov property

Of course our Feller-Dynkin processes have the Markov property. In particular, if $\zeta$ is a $\mathcal{F}_t$ measurable function and $f \in C_0(S)$, then

$$\mathbb{E}[\zeta f(X_{t+s})] = \mathbb{E}[\zeta P_s f(X_t)]. \tag{3.29}$$

Of course we want more to be true, namely as in the case of discrete time Markov chains, we want to be able to split past and future at stopping times. To formulate this, we denote as usual by $\theta_t$ the shift acting on $\Omega$, via

$$X(\theta_t \omega)_s \equiv (\theta_t X)(\omega)_s \equiv X(\omega)_{s+t}. \tag{3.30}$$

We then have the following *strong Markov property*:

**Theorem 3.3.13** *Let $T$ be a $\mathcal{F}_{t+}$stopping time, and let $\mathbb{P}$ be the law of a Feller-Dynkin Markov process, $X$. Then, for all bounded random variables $\eta$, if $T$ is a stopping time, then*

$$\mathbb{E}[\theta_T \eta | \mathcal{F}_{T+}] = \mathbb{E}_{X_T}[\eta], \tag{3.31}$$

*or equivalently, for all $\mathcal{F}_{T_+}$-measurable bounded random variables $\xi$,*

$$\mathbb{E}[\xi \theta_T \eta] = \mathbb{E}[\xi \mathbb{E}_{X_T}[\eta]], \tag{3.32}$$

*Proof.* We again use the dyadic approximation of the stopping time $T$ defined as

$$T^{(n)}(\omega) \equiv \begin{cases} k2^{-n}, & \text{if}\, (k-1)2^{-n} \leq T(\omega) < k2^{-n}, k \in \mathbb{N} \\ +\infty, & \text{if}\, T(\omega) = +\infty. \end{cases}$$

For $\Lambda \in \mathcal{F}_{T+}$ we set

$$\Lambda_{n,k} \equiv \{\omega \in \Omega : T^{(n)}(\omega) = 2^{-n}k\} \cap \Lambda \in \mathcal{F}_{k2^{-n}}.$$

Let $f$ be a continuous function on $S$. Then

$$\mathbb{E}\left[f(X_{T^{(n)}+s}\mathbb{1}_\Lambda\right] = \sum_{k\in\mathbb{N}\cup\{+\infty\}} \mathbb{E}\left[f(X_{k2^{-n}+s}\mathbb{1}_{\Lambda_{n,k}}\right] \tag{3.33}$$
$$= \sum_{k\in\mathbb{N}\cup\{+\infty\}} \mathbb{E}\left[P_s f(X_{k2^{-n}}\mathbb{1}_{\Lambda_{n,k}}\right]$$
$$= \mathbb{E}\left[P_s f(X_{T^{(n)}})\mathbb{1}_\Lambda\right]$$

Now let $n$ tend to infinity: by right-continuity of the paths,

$$X_{T^{(n)}+s} \to X_{T+s},$$

for any $s \geq 0$. Since $f$ is continuous, it also follows that

$$f(X_{T^{(n)}+s}) \to f(X_{T+s}),$$

and *since, by the Feller property,* $P_s f$ is also continuous, it holds that

$$P_s f(X_{T^{(n)}}) \to P_s f(X_T)$$

Note that finally working with Feller semi groups has payed off!

Now, by dominated convergence,

$$\mathbb{E}\left[f(X_{T+s})\mathbb{1}_\Lambda\right] = \mathbb{E}\left[P_s f(X_T\mathbb{1}_\Lambda\right]$$

To conclude the proof we must only generalize this result to more general functions, but this is done as usual via the monotone class theorem and presents no particular difficulties (e.g. we first see that $\mathbb{1}_\Lambda$ can be replaced by any bounded $\mathcal{F}_{T+}$-measurable function; next through explicit computation one shows that instead of $f(X_{T+s})$ we can put $\prod_{i=1}^n f_i(X_{T+s_i})$, and then we can again use the monotone class theorem to conclude for the general case. □

## 3.4 The martingale problem

In the context of discrete time Markov chains we have encountered a characterization of Markov processes in terms of the so-called *martingale problem*. While this proved quite handy, there was nothing really profoundly important about its use. This will change in the continuous time setting. In fact, the martingale problem characterizations of Markov processes, originally proposed by Stroock and Varadhan, turns out to be the "proper" way to deal with the theory in many respects.

Let us return to the issues around the Hille-Yosida theorem. In principle, that theorem gives us precise criteria to recognize when a given linear operator generates a strongly continuous contraction semigroup and hence a Markov process. However, if one looks at the conditions carefully, one will soon realize that in many situations it will be essentially impractical to verify them. The point is that the domain of a generator is usually far too big to allow us to describe the action of the generator on all of its elements. E.g., in Brownian motion we want to think of the generator as the Laplacian, but, except in $d = 1$, this is not the case. We really can describe the generator only on twice differentiable functions, but this is not the domain of the full generator, but only a dense subset.

Let us discuss this issue from the functional analytic point of view first. We have already defined the notion of the (linear) extension of a linear operator.

First, we call the *closure*, $\overline{G}$, of a linear operator, $G$, the minimal extension of $G$ that is closed. An operator that has a closed linear extension is called *closable*.

**Lemma 3.4.14** *A dissipative linear operator, $G$, on $B_0$ whose domain, $\mathcal{D}(G)$, is dense in $B_0$ is closable, and the closure of* **range**$(\lambda - G)$ *is equal to* **range**$(\lambda - \overline{G})$ *for all $\lambda > 0$.*

*Proof.* Let $f_n \in \mathcal{D}(G)$ be a sequence such that $f_n \to f$, and $Gf_n \to g$. We would like to associate with any such $f$ the value $g$ and then define $Gf = g$ for all achievable $f$ that would then be the desired closed extension of $G$. So all we need to show that if $f'_n \to f$ and $Gf'_n \to g'$, then $g' = g$. Thus, in fact all we need to show is that if $f_n \to 0$, and $Gf_n \to g$, then $g = 0$. To do this, consider a sequence of functions $g_n \in \mathcal{D}(G)$ such that $g_n \to g$. This exists because $\mathcal{D}(G)$ is dense in $B_0$. Using the dissipativity of $G$, we get then

$$\|(\lambda - G)g_n - \lambda g\| = \lim_{k\uparrow\infty} \|(\lambda - G)(g_n + \lambda f_k)\| \geq \lim_{k\uparrow\infty} \lambda \|g_n + \lambda f_k\| = \lambda \|g_n\|.$$

Note that in the first inequality we used that $0 = \lim_k f_k$ and $g = \lim_k Gf_k$. Dividing by $\lambda$ and taking the limit $\lambda \uparrow \infty$ implies that

$$\|g_n\| \leq \|g_n - g\|.$$

Since $g_n - g \to 0$, this implies $g_n \to 0$.

The identification of the closure of the range with the range of the closure follows from the observation made earlier that a range of a dissipative operator is closed if and only if it is closed. □

As a consequence of this lemma, if a dissipative linear operator on $B_0$, $G$, is closable, and if the range of $\lambda - G$ is dense in $B_0$, then its closure is the generator of a strongly continuous contraction semigroup on $B_0$.

These observations motivate the definition of a *core* of a linear operator.

**Definition 3.4.1** Let $G$ be a linear operator on a Banach space $B_0$. A subspace $D \subset \mathcal{D}(G)$ is called a *core* for $G$, if the closure of the restriction of $G$ to $D$ is equal to $G$.

**Lemma 3.4.15** *Let $G$ be the generator of a strongly continuous contraction semigroup on $B_0$. Then a subspace $D \subset \mathcal{D}(G)$ is a core for $G$, if and only if $D$ is dense in $B_0$ and, for some $\lambda > 0$,* **range**$(\lambda - G_{|D})$ *is dense in $B_0$.*

*Proof.* Follows from the preceding observations. □

The following is a very useful characterization of a core in our context.

**Lemma 3.4.16** *Let $G$ be the generator of a strongly continuous contraction semigroup, $P_t$, on $B_0$. Let $D$ be a dense subset of $\mathcal{D}(G)$. If, for all $t \geq 0$, $P_t : D \to D$, then $D$ is a core [in fact it suffices that there is a dense subset, $D_0 \subset D$, such that $P_t$ maps $D_0$ into $D$].*

*Proof.* Let $f \in D_0$ and set

$$f_n \equiv \frac{1}{n} \sum_{k=0}^{n^2} e^{-\lambda k/n} P_{k/n} f.$$

By hypothesis, $f_n \in D$. By strong continuity,

$$\begin{aligned} \lim_{n\uparrow\infty} (\lambda - G) f_n &= \lim_{n\uparrow\infty} \frac{1}{n} \sum_{k=0}^{n^2} e^{-\lambda k/n} P_{k/n} (\lambda - G) f \qquad (3.34) \\ &= \int_0^\infty e^{-\lambda t} P_t (\lambda - G) f \\ &= R_\lambda (\lambda - G) f = f \end{aligned}$$

Thus, for any $f \in D_0$, there exists a sequence of functions, $(\lambda - G) f_n \in$ **range**$(\lambda - G_D)$, that converges to $f$. Thus the closure of the range of $(\lambda - G_{|D})$ contains $D_0$. But since $D_0$ is dense in $B_0$, the assertion follows from the preceding lemma. □

Example. Let $G$ be the generator of Brownian motion. Then $C^\infty(\mathbb{R}^d)$ is a core for $G$ and $G$ is the closure of $\frac{1}{2}\Delta$ with this domain.

To show that $C^\infty$ is a core, since obviously $C^\infty$ is dense in the space of continuous functions, by the preceding lemma we need only to show that $P_t$ maps $C^\infty$ to $C^\infty$. But this is obvious from the explicit formula for the transition function of Brownian motion. Thus it remains to check that the restriction of $G$ to $C^\infty$ is $\frac{1}{2}\Delta$, which is a simple calculation (we essentially did that in [2]). Hence $G$ is the closure of $\frac{1}{2}\Delta$.

We see that these results are nice, if we know already the semigroup. In more complicated situations, we may be able to write down the action of what we want to be the generator of the Markov process we want to construct on some (small) space of function. The question when is how to know whether this specifies a (unique) strongly continuous contraction semigroup on our desired space of functions, e.g. $C_0(S)$? We may be able to show that it is dissipative, but then, is $\mathbf{range}(\lambda - G)$ dense in $C_0$?

The martingale problem formulation is a powerful tool to address such question.

We begin with a relatively simple observation.

**Lemma 3.4.17** *Let $X$ be a Feller-Dynkin process with transition function $P_t$ and generator $G$. Define, for $f, g \in B(S)$,*

$$M_t \equiv f(X_t) - \int_0^t g(X_s)ds. \tag{3.35}$$

*Then, if $f \in \mathcal{D}(G)$ and $g = Gf$, $M_t$ is a $\mathcal{F}_t$-martingale.*

*Proof.* The proof goes exactly as in the discrete time case.

$$\begin{aligned}
\mathbb{E}[M_{t+u}|\mathcal{F}_t] &= \mathbb{E}[f(X_{t+u})|\mathcal{F}_t] - \int_0^t (Gf)(X_s)ds - \int_t^{t+u} \mathbb{E}[Gf(X_s)|\mathcal{F}_t]ds \qquad (3.36)\\
&= \int P_u(X_t, dy)f(y) - \int_0^t (Gf)(X_s)ds - \int_0^u \int P_s(X_t, dy)(Gf)(y)ds\\
&= f(X_t) - \int_0^t (Gf)(X_s)ds\\
&\quad + \int P_u(X_t, dy)f(y) - f(X_t) - \int_0^u \int P_s(X_t, dy)(Gf)(y)ds\\
&= M_t + \int P_u(X_t, dy)f(y) - f(X_t) - \int_0^u (P_sGf)(X_t)ds.
\end{aligned}$$

But

$$(P_r G f)(z) = \frac{d}{dr}(P_r f)(z),$$

and so

$$\int P_u(X_t, dy) f(y) - f(X_t) - \int_0^u (P_r G f)(X_t) ds = 0,$$

from which the claim follows. □

By "the martingale problem" we will consider the inverse problem associated to this observation.

**Definition 3.4.2** Given a linear operator, $G$, with domain $\mathcal{D}(G)$ and **range**$(G) \subset C_b(S)$, a $S$-valued (càdlàg ) process defined on a filtered càdlàg space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \in \mathbb{R}_+))$, is called a solution of the martingale problem associated to the operator $G$, if for any $f \in \mathcal{D}(G)$, $M_t$ defined by (3.35) is a $\mathcal{F}_t$-martingale.

**Remark 3.4.1** One may relax the càdlàg assumptions. Ethier and Kurtz [6] work in a more general setting, which entails a number of subtleties regarding the relevant filtrations that I want to avoid.

One of the key points in the theory of martingale problems will be the fact that $G$ may not need to be the full generator (i.e. the generator with maximal domain), but just a core, i.e. an operator defined on a smaller subspace of functions. This really makes the power of this approach.

Before we continue, we need some new notion of convergence in Banach spaces.

**Definition 3.4.3** A sequence $f_n \in B(S)$ is said to converge *pointwise boundedly* to a function $f \in B(S)$, iff

(i) $\sup_n \|f_n\|_\infty < \infty$, and
(ii) for every $x \in S$, $\lim_{n\uparrow\infty} f_n(x) = f(x)$.

A set $M \in B(S)$ is called bp-closed, if for any sequence $f_n \in M$ s.t. $bp - \lim f_n = f \in B(S)$, then $f \in M$. The bp-closure of a set $D \subset B(S)$ is the smallest bp-closed set in $B(S)$ that contains $D$. A set $M$ is called bp-dense, if its closure is $B(S)$.

**Lemma 3.4.18** *Let $f_n$ be such that $bp - \lim f_n = f$ and $bp - \lim G f_n = G f$. Then, if $f_n(X_t) - \int_0^t (G f_n)(X_s)$ is a martingale for all $n$, then $f(X_t) - \int_0^t (G f)(X_s)$ is a martingale.*

*Proof.* Straightforward. □

The implication of this lemma is that to find a unique solution of the martingale problem, it suffices to know the generator on a core.

**Proposition 3.4.19** *Let $G_1$ be an operator with $\mathcal{D}(G_1)$ and* $\mathbf{range}(G_1)$*, and let $G$ be an extension of $G_1$. Assume that the bp-closures of the graphs of $G_1$ and $G$ are the same. Then a stochastic process $X$ is a solution for the martingale problem for $G$ if and only if it is a solution for the martingale problem for $G_1$.*

*Proof.* Follows from the preceding lemma. □

The strategy will be to understand when the martingale problem has a unique solution and to show that this then is a Markov process. In that sense it will be comforting to see that only dissipative operators can give rise to the solution of martingale properties.

We first prove a result that gives an equivalent characterization of the martingale problem.

**Lemma 3.4.20** *Let $\mathcal{F}_t$ be a filtration and $X$ an adapted process. Let $f, g \in B(S)$. Then, for $\lambda \in \mathbb{R}$, (3.35) is a martingale if and only if*

$$e^{-\lambda t} f(X_t) + \int_0^t e^{-\lambda s} \left(\lambda f(X_s) - g(X_s)\right) ds \tag{3.37}$$

*is a martingale.*

*Proof.* The details are left as an exercise. To see why this should be true, think of $P_t^\lambda \equiv e^{-\lambda t} P_t$ as a new semi-group. Its generator should be $(G_\lambda)$, which suggests that (3.37) should be a martingale whenever (3.35) is, and vice versa. □

**Lemma 3.4.21** *Let $G$ be a linear operator with domain and range in $B(S)$. If a solution for the martingale problem for $G$ exists for any initial condition $X_0 = x \in S$, then $G$ is dissipative.*

*Proof.* Let $f \in \mathcal{D}(G)$ and $g = Gf$. Now use that (3.37) is a martingale with $\lambda > 0$. Taking expectations and sending $t$ to infinity gives thus

$$f(X_0) = f(x) = \mathbb{E}\left[\int_0^\infty e^{-\lambda s} \left(\lambda f(X_s) - g(X_s)\right) ds\right]$$

and thus,

$$|f(x)| \leq \int_0^\infty e^{-\lambda s}\mathbb{E}|\lambda f(X_s)-g(X_s)|ds \leq \int_0^\infty e^{-\lambda s}\|\lambda f-g\| = \lambda^{-1}\|\lambda f-g\|ds,$$

which proves that $G$ is dissipative. □

Next, we know that martingales usually have a càdlàg modification. This suggests that, provided the set of functions on which we have defined our martingale problem is sufficiently rich, this property should carry over to the solution of the martingale problem as well. The following theorem shows when this holds.

**Theorem 3.4.22** *Assume that $S$ is separable, and that $\mathcal{D}(G) \subset C_b(S)$. Suppose moreover that $\mathcal{D}(G)$ is separating and contains a countable subset that separates points. If $X$ is a solution of the associated martingale problem and if for any $\varepsilon > 0$ and $T < \infty$ there exists a compact set $K_{\varepsilon,T} \subset S$, such that*

$$\mathbb{P}\left(\forall t \in [0,T] \cap \mathbb{Q} : X_t \in K_{\varepsilon,T}\right) > 1-\varepsilon, \tag{3.38}$$

*then $X$ has càdlàg modification.*

*Proof.* By assumption there exists a sequence $f_i \in \mathcal{D}(G)$ that separates points in $S$. Then

$$M_t^{(i)} \equiv f_i(X_t) - \int_0^t g_i(X_s)ds$$

with $g_i \equiv Gf_i$ are martingales and so by Doob's regularity theorem regularisable with probability one; since $\int_0^t g_i(X_s)ds$ is manifestly continuous, if follows that $f_i(X_t)$ is regularisable. In fact there exists a set of full measures such that all $f_i(X_t))$ are regularisable. Moreover, by hypothesis (3.38), the set $\{X_t(\omega), t \in [0,T]\}$ has compact closure for almost all $\omega$ for all $T$. Let $\Omega'$ denote the set of full measure where all the properties above hold. Then, for all $\omega \in \Omega'$, and all $t \geq 0$, there exists sequences $\mathbb{Q} \ni s_n \downarrow t$, such that $\lim_{s_n\downarrow t} X_{s_n}(\omega)$ exists and whence

$$f_i(\lim_{s_n\downarrow t} X_{s_n}(\omega)) = \lim_{\mathbb{Q}\ni s\downarrow t} f_i(X_s(\omega)).$$

Since the sequence $f_i$ separates points, it follows that $\lim_{\mathbb{Q}\ni s\downarrow t} X_s(\omega) \equiv Y_t(\omega)$ exists for all $t$. In fact, $X$ has a càdlàg regularization. Finally we need to show that $f_i(Y_t) = f_i(X_t)$ , a.s., in order to show that $Y$ is a modification of $X$. But this follows from the fact that the integral term in the formula for $M_t$ is continuous in $t$, and hence

$$f_i(Y_t) = \mathbb{E}f_i(Y_t)|\mathcal{F}_t]\mathbb{E}f_i(Y_t)|\mathcal{F}_t] = \lim_{s\downarrow t} \mathbb{E}(f_i(X_s)|\mathcal{F}_t) = f_i(X_t), \text{ a.s.}$$

by the fact that $M_t^{(i)}$ is a martingale. □

### 3.4.1 Uniqueness

We have seen that solutions to the martingale problem provide candidates for nice Markov processes. The main issues to understand is when a martingale problem has a *unique* solution, and whether in that case is represents a Markov process. When talking about uniqueness, we will of course always think that an initial distribution, $\mu_0$, is given. The data for the martingale problem is thus a pair $(G, \mu)$, where $G$ is a linear operator with its domain $\mathcal{D}(G)$ and $\mu$ is a probability measure on $S$.

The following first result is not terribly surprising.

**Theorem 3.4.23** *Let $S$ be separable and let $G$ be a linear dissipative operator on $B(S)$ with $\mathcal{D}(G) \subset B(S)$. Suppose there exists $G'$ with $\mathcal{D}(G') \subset \mathcal{D}(G)$ such that $G$ is an extension of $G'$. Let $\overline{\mathcal{D}(G')} = \overline{\mathbf{range}(\lambda - G')} \equiv L$, and let $L$ be separating. Let $X$ be a solution for the martingale problem for $(G, \mu)$. Then $X$ is a Markov process whose semigroup on $L$ is generated by the closure of $G'$, and the martingale problem for $(G, \mu)$ has a unique solution.*

*Proof.* Assume $G'$ closed. We know that it generates a unique strongly continuous contraction semigroup on $L$, hence a unique Markov process with generator $G'$. Thus we only have to show that the solution of the martingale problem satisfies the Markov property with respect to that semigroup.

Let $f \in \mathcal{D}(G')$ and $\lambda > 0$. Then, by Lemma 3.4.20,

$$e^{-\lambda t} f(X_t) + \int_0^t e^{-\lambda s}(\lambda f(X_s) - G' f(X_s))ds$$

is a martingale,

$$f(X_t) = \mathbb{E}\left[\int_0^\infty e^{-\lambda s}\left(\lambda f(X_{t+s} - G' f(X_{t+s}))\right) ds \middle| \mathcal{F}_t\right]. \tag{3.39}$$

To see this note that for any $T > 0$, by simple algebra,

$$\int_0^T e^{-\lambda s}\left(\lambda f(X_{t+s} - G'f(X_{t+s}))\right) ds \tag{3.40}$$
$$= e^{\lambda t}\int_0^{t+T} e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds - e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds$$
$$= e^{\lambda t}\left[\int_0^{t+T} e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds + e^{-(t+T}f(X_{t+T})\right] - e^{-T\lambda}f(X_{t+T})$$
$$-e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds$$

Hence,

$$\mathbb{E}\left[\int_0^T e^{-\lambda s}\left(\lambda f(X_{t+s}) - G'f(X_{t+s}))\right) ds\Big|\mathcal{F}_t\right] \tag{3.41}$$
$$= f(X_t) + e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds$$
$$-e^{-\lambda T}\mathbb{E}\left[f(X_{t+T})\big|\mathcal{F}_t\right] - e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G'f(X_s))\right) ds$$
$$= f(X_t) - e^{-\lambda T}\mathbb{E}\left[f(X_{t+T})\big|\mathcal{F}_t\right].$$

Letting $T$ tend to infinity, we get (3.39).

We will use the following lemma.

**Lemma 3.4.24** *Let $P_t$ be a SCCSG on $B_0$ and $G$ its generator. Then, for any $f \in B_0$,*

$$\lim_{n\uparrow\infty}(1 - n^{-1}G)^{-[nt]}f = P_t f. \tag{3.42}$$

*Proof.* Set $V(t) \equiv (1 - tG]^{-1}$. We want to show that $V(1/n)^[tn] \to P_t$. But

$$n[V(1/n)f - f] = n\left[(1 - n^{-1}G)^{-1}f - f\right] = G_n f,$$

where $G_n$ is the Hille-Yosida approximation of $G$. Hence

$$V(1/n)^{tn}f = \left[1 + n^{-1}G_n\right]^{tn}.$$

Now one can show that for any linear contraction $B$ (**Exercise!**),

$$\|B^n f - e^{n(B-1)}f\| \leq \sqrt{n}\|BF - f\||.$$

We will apply this for $B = \frac{1}{n}G_n + 1$. Thus

$$\left\|\left[1 + n^{-1}G_n\right]^{tn}f - \exp(tG_n)f\right\| \leq n^{-1/2}\|G_n f\|.$$

Since the right-hand side converges to zero for $f \in \Delta(G)$, and $\exp(tG_n)f \to P_t f$, by the Hille-Yosioda theorem, we arrive at the claim of the lemma for $f \in \Delta(G)$. But since $\Delta(G)$ is dense, the result holds for all $B_0$ by standard arguments. □

Now from (3.39)

$$\begin{aligned}(1-n^{-1}G')^{-1}f(X_t) &= n\frac{1}{n-G'}f(X_t) \\ &= \mathbb{E}\left[n\int_0^\infty e^{-ns}f(X_{t+s})ds\Big|\mathcal{F}_t\right] \\ &= \mathbb{E}\left[\int_0^\infty e^{-s}f(X_{t+n^{-1}s})ds\Big|\mathcal{F}_t\right]\end{aligned} \tag{3.43}$$

Iterating this formula and re-arranging the resulting multiple integrals, and using the formula for the area of the $k$-dimensional simplex, gives

$$\begin{aligned}&(1-n^{-1}G')^{-[nu]}f(X_t) \\ &= \mathbb{E}\left[\int_0^\infty e^{-s_1-s_2\cdots-s_{[un]}}f(X_{t+n^{-1}(s_1+\cdots+s_{[un]})})ds_1\dots ds_{[un]}\Big|\mathcal{F}_t\right] \\ &= \mathbb{E}\left[\int_0^\infty e^{-s}\frac{s^{[un]-1}}{\Gamma([un])}f(X_{t+n^{-1}(s)})ds\Big|\mathcal{F}_t\right]\end{aligned} \tag{3.44}$$

We write, for $f \in \mathcal{D}(G')$,

$$f(X_{t+n^{-1}(s)}) = f(X_{t+u} + \int_u^{s/n} G'f(X_{t+v})dv$$

and insert this into (3.44). Finally, since

$$\int_0^\infty e^{-s}\frac{s^{[un]-1}}{\Gamma([un])}ds = 1,$$

we arrive at

$$\begin{aligned}(1-n^{-1}G')^{-[nu]}f(X_t) &= \mathbb{E}\left[f(X_{t+u})\big|\mathcal{F}_t\right] \\ &+ \mathbb{E}\left[\int_0^\infty e^{-s}\frac{s^{[un]-1}}{\Gamma([un])}\int_u^{s/n} G'f(X_{t+v)})dvds\Big|\mathcal{F}_t\right]\end{aligned} \tag{3.45}$$

We are finished if the second term tends to zero. But, re-expressing the volume of the sphere through multiply integrals, we see that

$$\begin{aligned}&\left|\mathbb{E}\left[\int_0^\infty e^{-s}\frac{s^{[un]-1}}{\Gamma([un])}\int_u^{s/n} G'f(X_{t+v)})dvds\Big|\mathcal{F}_t\right]\right| \\ &\leq \|G'f\|_\infty \int_0^\infty ds_1\dots ds_{[un]}\left|n^{-1}(s_1+\cdots+s_{[un]})-u\right| e^{-s_1-\cdots-s_{[un]}}\end{aligned} \tag{3.46}$$

But the last integral is nothing but the expectation of $\left|n^{-1}\sum_{i=1}^{[un]} e_i - u\right|$ where $e_i$ are iid exponential random variable. Hence the law of large numbers implies that this converges to zero. Thus we have the desired relation

$$P_u f(X_t) = \mathbb{E}[f(X_{t+u})|\mathcal{F}_t]$$

for all $f \in \mathcal{D}(G')$. In the usual way, this relation extends to the closure of $\mathcal{D}(G')$ which by assumption if $L$. □

Finally we establish an important uniqueness criterion an the strong Markov property for solutions of uniquely posed martingale problems.

**Theorem 3.4.25** *Let $S$ be a separable space and let $G$ be a linear operator on $B(S)$. Suppose that the for any initial distribution, $\mu$, any two solutions, $X, Y$, of the martingale problem for $(G, \mu)$ have the same* one-dimensional *distributions, i.e. for any $t \geq 0$, $\mathbb{P}(X_t \in A) = \mathbb{P}(Y_t \in A)$ for any Borel set $A$. Then the following hold:*

*(i) Any solution of the martingale problem for $G$ is a Markov process and any two solutions of the martingale problem with the same initial distribution have the same finite dimensional distributions (i.e. uniqueness holds).*

*(ii) If $\mathcal{D}(G) \subset C_b(S)$ and $X$ is a solution of the martingale problem with càdlàg sample paths, then for any a.s. finite stopping time, $\tau$,*

$$\mathbb{E}[f(X_{t+\tau})|\mathcal{F}_\tau] = \mathbb{E}[f(X_{t+\tau})|X_\tau], \tag{3.47}$$

*for all $f \in B(S)$.*

*(iii) If in addition to the assumptions in (ii), there exists a càdlàg solution of the martingale problem for any initial measure of the form $\delta_x$, $x \in S$, then the strong Markov property holds, i.e.*

$$\mathbb{E}[f(X_{t+\tau})|\mathcal{F}_\tau] = P_t f(X_\tau). \tag{3.48}$$

*Proof.*

Let $X$ be the solution of the martingale problem with respect to some filtration $\mathcal{G}_t$. We want to prove that it is a Markov process. Let $F \in \mathcal{G}_r$ have positive probability. The, for any measurable set $B$ let

$$P_1(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[\mathbb{1}_B|\mathcal{G}_r]\right]}{\mathbb{P}(F)} \tag{3.49}$$

and

$$P_2(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[\mathbb{1}_B|X_r]\right]}{\mathbb{P}(F)}. \tag{3.50}$$

Let $Y_s \equiv X_{r+s}$. We see that, since $\mathbb{E}[f(X_r)|X_r] = f(X_r) = \mathbb{E}[f(X_r)|G_r]$,

$$P_1(Y_0 \in \Gamma) = P_2(Y_0 \in \Gamma) = \mathbb{P}[X_r \in \Gamma|F] \tag{3.51}$$

Now chose any $0 \leq t_1 < t_2 < \cdots < t_{n+1}$, $f \in \mathcal{D}(G)$, $g = Gf$, and $h_k \in B(S)$, $(k \in \mathbb{N}$. Define

$$\eta(Y) \equiv \left( f(Y_{t_{n+1}}) - f(Y_{t_n}) - \int_{t_n}^{t_{n+1}} g(Y_s)ds \right) \prod_{k=1}^{n} h_k(Y_{t_k}). \tag{3.52}$$

$Y$ is a solution of the martingale problem if and only if $\mathbb{E}\eta(Y) = 0$ for all possible choices of the parameters (Check this!).

Now $\mathbb{E}\left[\eta(X_{r+\cdot})|\mathcal{G}_r\right] = 0$, since $X$ is a solution of the martingale problem. A fortiori, $\mathbb{E}\left[\eta(X_{r+\cdot})|X_r\right] = 0$, and so

$$E_1[\eta(Y)] = E_2[\eta(Y)] = 0,$$

where $E_i$ denote the expectation w.r.t. the measures $P_i$. Hence, $Y$ is a solution to the martingale problem for $G$ under both $P_1$ and $P_2$, and by (3.51),

$$E_1[f(Y_t)] = E_2[f(Y_t)],$$

for any bounded measurable function. Thus, for any $F \in \mathcal{G}_r$,

$$\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[f(X_{r+s})|\mathcal{G}_r]\right] = \mathbb{E}\left[\mathbb{1}_F \mathbb{E}[f(X_{r+s})|X_r]\right],$$

and hence

$$\mathbb{E}[f(X_{r+s})|\mathcal{G}_r] = \mathbb{E}[f(X_{r+s})|X_r].$$

Thus $X$ is a Markov process.

To prove uniqueness one proceeds as follows. Let $X$ and $Y$ be two solutions of the martingale problem for $(G, \mu)$. We want to show that

$$\mathbb{E}\left[\prod_{k=1}^{n} h_k(X_{t_k})\right] = \mathbb{E}\left[\prod_{k=1}^{n} h_k(Y_{t_k})\right]. \tag{3.53}$$

By hypothesis, this holds for $n = 1$, so we will proceed by induction, assuming (3.53) for all $m \leq n$. For with we define two new measures

$$\widetilde{P}(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_B \prod_{k=1}^{n} h_k(X_{t_k})\right]}{\mathbb{E}\left[\prod_{k=1}^{n} h_k(X_{t_k})\right]}, \tag{3.54}$$

$$\widetilde{Q}(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_B \prod_{k=1}^{n} h_k(Y_{t_k})\right]}{\mathbb{E}\left[\prod_{k=1}^{n} h_k(Y_{t_k})\right]}. \tag{3.55}$$

Set $\widetilde{X}_t \equiv X_{t+t_n}$ and $\widetilde{Y}_t \equiv Y_{t+t_n}$. As in the proof of the Markov property,

$\widetilde{X}$ and $\widetilde{Y}$ are solutions of the martingale problems under $\widetilde{P}$ and $\widetilde{Q}$, respectively. Now for $t = 0$, we get from the induction hypothesis that

$$\widetilde{\mathbb{E}}^P f(\widetilde{X}_0) = \widetilde{\mathbb{E}}^Q f(\widetilde{Y}_0)$$

where the expectations are w.r.t. the measures defined above. Thus $\widetilde{X}$ and $\widetilde{Y}$ have the same initial distribution. Now we can use the fact the by hypothesis, any two solutions of our martingale problem with the same initial conditions have the same one-dimensional distributions. But this provides immediately the assertion for $m = n + 1$ and concludes the inductive step.

The proofs of the strong properties (ii) and (iii) follows from similar constructions using stopping times $\tau$ instead of $r$, and optional sampling theorem for bounded continuous functions of càdlàg martingales. E.g., to get (ii), note that

$$\mathbb{E}[\eta(X_{\tau+s})|\mathcal{G}_\tau] = 0.$$

For part (iii) we construct the measures $P_i$ replacing $r$ by $\tau$ and so get instead of the Markov property the strong Markov property. □

Note that in the above theorem, we have made no direct assumptions on the choice of $\mathcal{D}(G)$ (in particular, it need not separate point, as in the previous theorem). The assumption is implicit in the requirement that uniqueness of the one-dimensional marginals must be satisfies. This is then also the main message: a martingale problem that gets uniqueness of the one-dimensional marginals implies uniqueness of the finite dimensional marginals. This theorem is in fact the usual way to prove uniqueness of solutions of martingale problems.

**Duality.** One still needs methods to verify the hypothesis of the last theorem. A very useful one is the so-called duality method.

**Definition 3.4.4** Consider two separable metric spaces $(S, \rho)$ and $(E, r)$. Let $G_1, G_2$ be two linear operators on $B(S)$, resp. $B(E)$. Let $\mu, \nu$ be probability measures on $S$, resp. $E$, $\alpha : S \to \mathbb{R}$, $\beta : E \to \mathbb{R}$, $f : S \times E \to \mathbb{R}$, measurable functions. Then the martingale problems for $(G_1, \mu)$ and $(G_2, \nu)$ are *dual* with respect to $(f, \alpha, \beta)$, of for any solution, $X$, of the martingale problem for $(G_1, \mu)$ and any solution $Y$ of $(G_2, \nu)$, the following hold:

(i) $\int_0^t (|\alpha(X_s)| + |\beta(Y_s)|)ds < \infty$, a.s.,

(ii)

$$\int \mathbb{E}\left[\left|f(X_t, y)\exp\left(\int_0^t \alpha(X_s)ds\right)\right|\right]\nu(dy) < \infty, \tag{3.56}$$

$$\int \mathbb{E}\left[\left|f(x, Y_t)\exp\left(\int_0^t \beta(Y_s)ds\right)\right|\right]\mu(dx) < \infty, \tag{3.57}$$

(iii) and,

$$\int \mathbb{E}\left[\left|f(X_t, y)\exp\left(\int_0^t \alpha(X_s)ds\right)\right|\right]\nu(dy) \tag{3.58}$$
$$= \int \mathbb{E}\left[\left|f(x, Y_t)\exp\left(\int_0^t \beta(Y_s)ds\right)\right|\right]\mu(dx)$$

for any $t \geq 0$.

**Proposition 3.4.26** *With the notation of the definition, let $\mathcal{M} \subset \mathcal{M}_1(S)$ contain the set of all one-dimensional distributions of all solutions of the martingale problem for $G_1$ for which the distribution of $X_0$ has compact support. Assume that $(G_1, \mu)$ and $(G_2, \delta_y)$ are dual with respect to $(f, 0, \beta)$ for every $\mu$ with compact support and any $y \in E$. Assume further that the set $\{f(\cdot, y) : y \in E\}$ is separating on $\mathcal{M}$. If for every $y \in E$ there exists a solution of the martingale problem $(G_2, \delta_y)$, then uniqueness holds for each $\mu$ in the martingale problem $(G_1, \mu)$.*

*Proof.* Let $X$ and $\widetilde{X}$ be solutions for the martingale problem for $(G_1, \mu)$ where $\mu$ has compact support, and let $Y^y$ be a solution to the martingale problem $(G_2, \delta_y)$. By duality we have then that

$$\mathbb{E}[f(X_t, y)] = \int \mathbb{E}\left[f(x, Y_t^y)\exp\left(\int_0^t \beta(Y_s^y)ds\right)\right]\mu(dx) = \mathbb{E}[f(\widetilde{X}_t, y)] \tag{3.59}$$

Now we assumed that the class of functions $\{f(\cdot, y) : y \in E\}$ is separating on $\mathcal{M}$, so the one-dimensional marginals of $X$ and $\widetilde{X}$ coincide.

If $\mu$ does not have compact support, take a compact set $K$ with $\mu(K) > 0$ and consider the two solutions $X$ and $\widetilde{X}$ conditioned on $X_0 \in K$, $\widetilde{X}_0 \in K$. They are solutions of the martingale problem for the initial distribution conditioned on $K$, and hence have the same one-dimensional distributions. Thus

$$\mathbb{P}[X_t \in \Gamma | X_0 \in K] = \mathbb{P}[\widetilde{X}_t \in \Gamma | \widetilde{X}_0 \in K]$$

for any $K$, which again implies, since $\mu$ is inner regular, the equality of the one dimensional distributions and thus uniqueness by Theorem 3.4.25. □

This theorem leaves a lot to good guesswork. It is more or less an art to find dual processes and there are no clear results that indicate when and why this should be possible. Nonetheless, the method is very useful and widely applied.

Let us see how one might wish to go about finding duals. Let us assume that we have two independent processes, $X, Y$, on spaces $S_1, S_2$, and two functions $g, h \in B(S_1 \times S_2)$, such that

$$f(X_t, y) - \int_0^t g(X_s, y) ds \tag{3.60}$$

and

$$f(x, Y_t) - \int_0^t h(x, Y_s) ds \tag{3.61}$$

are martingales with respect to the natural filtrations for $X$, respectively $Y$. Then (3.58) is the integral of

$$\frac{d}{ds}\mathbb{E}\left[f(X_s, Y_{t-s}) \exp\left(\int_0^s \alpha(X_u)du + \int_0^{t-s} \beta(Y_u)du\right)\right]. \tag{3.62}$$

Computing (assuming that we can pull the derivative into the expectation) gives that (3.62) equals

$$\mathbb{E}\ \left[\Big(g(X_s, Y_{t-s}) - h(X_s, Y_{t-s}) + (\alpha(X_s) - \beta(Y_{t-s}))\, f(X_s, Y_{t-s})\Big)\right.$$
$$\left.\times \exp\left(\int_0^s \alpha(X_u)du + \int_0^{t-s} \beta(Y_u)du\right)\right]. \tag{3.63}$$

This latter quantity is equal to zero, if

$$g(x, y) + \alpha(x) f(x, y) = h(x, y) + \beta(y) f(x, y). \tag{3.64}$$

To see how this can be used, we look at the following simple example. Let $S_1 = \mathbb{R}$ and $S_2 = \mathbb{N}_0$. The process $X$ has generator $G_1$ defined on smooth functions by $G_1 = \frac{d^2}{dx^2} - x\frac{d}{dx}$ and $Y$ has generator $G_2 f(y) = y(y-1)(f(y-2) - f(y))$. Clearly the process $Y$ can be realized as a Markov jump process that jumps down by 2 and is absorbed in the states 0 and 1. The Second process is called Ornstein-Uhlenbeck process. Now choose the function $f(x, y) = x^y$. If $X$ is a solution of the martingale problem for $G_1$, we get, assuming the necessary integrability conditions, that will be satisfied if the initial distribution of $X_0$ has bounded support), that

$$X_t^y - \int_0^t \left(y(y-1)X_s^{y-2} - yX_s^y\right) ds \tag{3.65}$$

are martingales. Of course, this suggest to choose

$$g(x,y) = y(y-1)x^{y-2} - yx^y, \tag{3.66}$$

Similarly,

$$x^{Y_t} - \int_0^t Y_s(Y_s-1)\left(x^{Y_s-2} - x^{Y_s}\right) ds \tag{3.67}$$

is a martingale and hence

$$h(x,y) = y(y-1)\left(x^{y-2} - x^y\right). \tag{3.68}$$

Now we may set $\alpha = 0$ and see that we can satisfy (3.64) by putting

$$\beta(y) = y^2 - 2y. \tag{3.69}$$

Thus we get

$$\mathbb{E}\left[X_t^{Y_0}\right] = \mathbb{E}\left[X_0^{Y_t} \exp\left(\int_0^t (Y_u - 2Y_u)\, du\right)\right]. \tag{3.70}$$

This explains in a way what is happening here: the jump process $Y$ together with the initial distribution of the process $X$ determines the moments of the process $X_t$. One may check that in the present case, these are actually growing sufficiently slowly to determine the distribution of $X_t$, this in turn is, as we know, sufficient to determine the law of the process $X$.

The general structure we encounter in this example is rather typical. One will often try to go for an integer-valued dual process that determines the moments of the process of interest. Of course, success is not guaranteed.

The tricky part is to guess good functions $f$ and a good dual process $Y$. To show existence for the dual process is often not so hard. We will now turn briefly to the existence question in general.

### 3.4.2 Existence

We have seen that a uniquely solvable martingale problem provides a way to construct a Markov process. We need to have ways to produce solutions of martingale problems. The usual way to do this is through approximations and weak convergence.

**Lemma 3.4.27** *Let $G$ be a linear operator with domain and range in $C_b(S)$. Let $G_n, n \in \mathbb{N}$ be a sequence of linear operators with domain and range in $B(S)$. Assume that, for any $f \in \mathcal{D}(A)$, there exists a sequence, $f_n \in \mathcal{D}(G_n)$ , such that*

$$\lim_{n\uparrow\infty} \|f_n - f\| = 0, \text{and} \lim_{n\uparrow\infty} \|G_n f_n - Gf\| = 0. \tag{3.71}$$

*Then, if for each $n$, $X^n$ is a solution of the martingale problem for $G_n$ with càdlàg sample paths, and if $X^n$ converges to $X$ weakly, then $X$ is a càdlàg solution to the martingale problem for $G$.*

*Proof.* Let $0 \leq t_i \leq t < s$ be elements of the set $\mathcal{C}(X) \equiv \{u \in \mathbb{R}_+ : \mathbb{P}[X_u = X_{u-}] = 1\}$. Let $h_i \in C_b(S)$, $i \in \mathbb{N}$. Let $f, f_n$ be as in the hypothesis of the lemma. Then

$$\mathbb{E}\left[\left(f(X_s) - f(X_t) - \int_t^s Gf(X_u)du\right)\prod_{i=1}^k h_i(X_{t_i})\right] \tag{3.72}$$
$$= \lim_{n\uparrow\infty} \mathbb{E}\left[\left(f_n(X_s^n) - f_n(X_t^n) - \int_t^s Gf_n(X_u^n)du\right)\prod_{i=1}^k h_i(X_{t_i}^n)\right]$$
$$= 0$$

Now the complement of the set $\mathcal{C}(X)$ is at most countable, and then the relation (3.72) carries over to all points $t_i \leq t < s$. But this implies that $X$ solves the martingale problem for $G$. □

The usefulness of the result is based on the following lemma, which implies that we can use Markov jump processes as approximations.

**Lemma 3.4.28** *Let $S$ be compact and let $G$ be a dissipative operator on $C(S)$ with dense domain and $G1 = 0$. Then there exists a sequence of positive contraction operators, $T_n$, on $B(S)$ given by transition kernels, such that, for $f \in \mathcal{D}(G)$,*

$$\lim_{n\uparrow\infty} n(T_n - 1)f = Gf. \tag{3.73}$$

*Proof.* I will only roughly sketch the ideas of the proof, which is closely related to the Hille-Yosida theorem. In fact, from $G$ we construct the resolvent $(n-G)^{-1}$ on the range of $(n-G)$. Then For a dissipative $G$, the operators $n(n-G)^{-1}$ are bounded (by one) on **range**$(n-G)$. Thus, by the Hahn-Banach theorem, they can be extended to $C(S)$ as bounded operators. Using the Riesz representation theorem one can then associate to $n(n-G)^{-1}$ a probability measure, s.t.

$$n(n-G)^{-1}f(x) = \int f(y)\mu_n(x, dy),$$

and hence $n(n-G)^{-1} \equiv T_n$ defines a Markov transition kernel. Finally, ist remains to show that $n(T_n - 1)f = \frac{nG}{n-G} f = T_n G f$ converges to $Gf$, for $f \in \mathcal{D}(G)$, which is fairly straightforward. □

The point of the lemma is that it shows that the martingale problem for $G$ can be approximated by martingale problems with *bounded* generators of the form B

$$G_n F(x) = n \int (f(y) - f(x)) \mu_n(x, dy).$$

For such generators, the construction of a solution can be done explicitly in various ways, e.g. by constructing the transition function through the convergent series for $\exp(tG_n)$.

Such Markov processes are called *Markov jump processes* because they can be realized in a very simple through a time change of a discrete time Markov chain. To be a bit more general, let $G$ be a generator of the form

$$Gf(X) = \lambda \int (f(y) - f(x)) \mu(x, dy).$$

Let $Y_k, k \in \mathbb{N}$ be a Markov chain with state space $S$ and transition kernel

$$P_1(x.A) = \mu(x, A).$$

Then let $\tau_i$ be a family of iid exponential random variables with parameter one. Define

$$X_t \equiv \begin{cases} Y_0, & \text{if } 0 \leq t < \frac{\tau_0}{\lambda(Y_0)}, \\ Y_k, & \text{if } \sum_{\ell=0}^{k-1} \frac{\tau_\ell}{\lambda(Y_\ell)} \leq t < \sum_{\ell=0}^{k} \frac{\tau_\ell}{\lambda(Y_\ell)} \end{cases}. \tag{3.74}$$

Then $X_t$ is a Markov process with generator $G$. In other words, the process $X$ follows the same trajectory as the Markov chain $Y$, but waits an exponential time of mean $1/\lambda(Y_k)$ before making the next move when it reaches a state $Y_k$.

I leave it as an exercise to check this fact.

## 3.5 Convergence results

**This section is still under construction!!!**

An obvious question to be asked in the theory of Markov processes is to what extend convergence of sequences of semi-groups implies convergence of the corresponding processes. As a preparation we need to connect convergence of semi-groups and generators.

**Theorem 3.5.29** *Let $P_t^{(n)}, P_t^{n)}$ be SCCSG's on a Banach space $B_0$ with generators $G_n, G$, respectively. Let $D$ be a core for $G$. Then the following are equivalent:*

*(i) For each $f \in B_0$, $P_t^{(n)} f \to P_t f$ for all $t \in \mathbb{R}_t$, uniformly on bounded intervals.*
*(ii) For each $f \in B_0$, $P_t^{(n)} f \to P_t f$ for all $t \in \mathbb{R}_t$.*
*(iii) For each $f \in D$, there exists $f_n \in \mathcal{D}(G_n)$ for each $n$, such that $f_n \to f$ and $G_n f \to Gf$.*

*Proof.* It is clear that (i) is stronger than (ii). Next we show that (ii) implies (iii): Let $\lambda > 0$, $f \in \mathcal{D}(A)$, and $g \equiv (\lambda - G)f$, so that $f = R_\lambda g$. Set $f_n \equiv R_\lambda^{(n)} g \in \mathcal{D}(G_n)$. Since

$$R_\lambda^{(n)} g = \int_0^\infty e^{-\lambda t} P_t^{(n)} g,$$

(ii) together with Lebesgue's dominated convergence theorem implies that $f_n \to f$. But since $(\lambda - G_n) f_n = g$, it follows that $G_n f_n \to Gf$.

It remains to show that (iii) implies (i): Let $P_t^{(n),\lambda}$ be defined as in the Hille-Yosida theorem. For $f \in D$, let $f_n$ be defined as above. Then

$$\begin{aligned} P_t^{(n)} f - P_t f = & P_t^{(n)}(f - f_n) + [P_t^{(n)} f_n - P_t^{(n),\lambda} f_n] \\ & + P_t^{(n),\lambda}(f_n - f) + [P_t^{(n),\lambda} f - P_t^\lambda f] \\ & + [P_t^\lambda f - P_t f]. \end{aligned}$$

Trivially, the first and the third term tend to zero, since

$$\|P_t^{(n),\lambda}(f - f_n)\| \leq \|f - f_n)\| \downarrow 0.$$

Also, the last term can be made arbitrarily small by taking $\lambda$ to infinity (by the Hille-Yosida theorem).

To deal with the remaining two terms, we need an auxiliary results:

**Lemma 3.5.30** *Let $P_t$ be a SCCSG with generator $G$ and $P^\lambda$, $G^\lambda$ be the Hille-Yosida approximants, for any $f \in \mathcal{D}(G)$,*

$$\|P_t^\lambda f - P_t f\| \leq t \|G^\lambda f - Gf\|. \tag{3.75}$$

*Proof.* This follows immediately from the Hille-Yosida-theorem and the bound (3.18). □

Thus we see that

$$\sup_{0\leq t\leq t_0}\left\|P_t^{(n)}f_n - P_t^{(n),\lambda}f_n\right\| \leq t_0\|G_nf_n - G_n^\lambda f_n\|$$
$$\leq t_0\|G_nf_n - Gf\| + t_0\|Gf - G^\lambda f\| + t_0\|G_n^\lambda(t - f_n)\|.$$

The first term tends to zero with $n$ by assumption and so does the last term since $\|G_n^\lambda\| \leq \lambda$. The second term can be made arbitrarily small by taking $\lambda$ to infinity for $f \in \mathcal{D}(G)$. Thus we have shown that $P_t^{(n)}f \to P_tf$, uniformly in compact $t$-sets, on a dense set of functions $f$. But this implies the same convergence on the closure, by the boundedness of the semi-groups.

This concludes the proof of the theorem. □

The following theorem gives a first answer.

**Theorem 3.5.31** *Let $S$ be a locally compact and separable space. Let $P_t^{(n)}$, $n \in N$ be a sequence of Feller semi-groups on $C_0(S)$ and let $X_n$ be the corresponding Markov processes with càdlàg paths. Suppose that $P_t$ is a Feller semi-group on $C_0(S)$ such that, for all $f \in C_0(S)$ and for all $t \in \mathbb{R}_+$,*

$$\lim_{n\uparrow\infty} P_t^{(n)}f = P_tf. \tag{3.76}$$

*Then, if $\mathbb{P}(X_n(0) \in A) \to \nu(A)$, for all Borel sets $A$, then there exists a Markov process $X$ with càdlàg paths and initial distribution $\nu$ such that $X_n \stackrel{\mathcal{D}}{\to} X$.*

*Proof.* Clearly weak convergence will involve some tightness argument. We will in fact use the following lemma.

**Lemma 3.5.32** *Let $S$ be a Polish space and let $\{X_n\}$ be a family of processes with càdlàg sample paths. Suppose that for every $\eta > 0$, and $T < \infty$, there exist compact sets $K_{\eta,T} \subset S$, such that*

$$\inf_n \mathbb{P}(X_n(t) \in K_{\eta,T}, \forall 0 \leq t \leq T) < 1 - \eta. \tag{3.77}$$

*Let $H$ be a dense subset of $C_0(S,\mathbb{R})$. Then $\{X_n\}$ is conditionally compact, if and only if $\{f \circ X_n\}$ is relatively compact in $D_{\mathbb{R}}[0,\infty)$, for each $f \in H$.*

The proof of this lemma can be found in [6] (Chapter 3.8).

Let $G_n$ be the generators of the semi-groups $P_t^{(n)}$. By the preceding theorem, for any $f \in \mathcal{D}(G)$, there exist functions $f_n \in \mathcal{D}(G_n)$, such that $f_n \to f$ and $G_n f_n \to Gf$. Then we know that

$$f_n(X_n(T)) - \int_0^t G_n f_n(X_n(s)ds$$

is a martingale. One can show (see Chapter 3.9 in [6]) that this implies that $f \circ X_n$ is relatively compact, and hence by Lemma 3.5.32, $X_n$ is relatively compact.

□

# 4

# Itô calculus

In this chapter we will develop the basics of the theory of stochastic integration and, closely related, stochastic integral, resp. differential equations. We will be far from the most general setting possible, but our treatment will of course include the most important case of integration with respect to Brownian motion. Apart form our standard texts, there is an ample literature on stochastic calculus. For further reading see e.g. the texts by Karatzas and Shreve [11] and Itô and McKeane [9].

In this chapter we will always work on a filtered space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, \tau \in \mathbb{R})\}$ that satisfies the conditions of the "usual setting" of Definition 1.4.2. We will be interested to define *stochastic integrals* of the form

$$\int_0^t X dM, \tag{4.1}$$

where $M$ is a martingale and $X$ is a progressive process. In fact, the full ambition of stochastic analysis is to find the largest class of pairs of processes $M$ and $X$ for which such an integral can be reasonable defined (which will lead to the notion of *semi-martingale*, but here we will limit our ambition to the considerably simpler case when $M$ is a *continuous*, square-integrable martingale, i.e. when $M$ has continuous paths (a.s.) and $\mathbb{E}M_t^2 < \infty$ for all $t \leq \infty$. This includes the important case when $M$ is a Brownian motion. In fact, we could limit ourselves to this particular case in a first step, and you are welcome to think that $M_t = B_t$ if that helps. But doing so we would loose some structural information which would be regrettable.

## 4.1 Stochastic integrals

In [2] we have defined the discrete stochastic integral of a progressive process with respect to a (sub-) martingale, $C \bullet M$. The key property of this construction was that it preserved the martingale properties of $M$. We want to do the same for continuous martingales.

In the theory of stochastic integration it will be useful to relax the notions associated with martingale properties to *local* ones.

**Definition 4.1.1** A stochastic process $M$ is called a *local martingale* if there exists a sequence of stopping times, $\tau_n \leq \tau_{n+1}$ such that $\tau_n \uparrow \infty$, such that the processes $M^{\tau_n} \equiv M_{\cdot \wedge \tau_n}$ are martingales. The same terminology applies to sub and super-martingales, as well as to various integrability properties.

**Remark 4.1.1** In the sequel I will sometimes state results for martingales. They can all be extended to local martingales.

Let us note as a first step that the definition of the stochastic integral can be done in the standard way as Stieltjes integral in the case when the integrand has (locally) bounded variation.

**Proposition 4.1.1** *Let $M$ be a càdlàg (local) martingale, and let $V$ be a continuous, adapted process that is locally of bounded variation. Then*

$$W(t) = \int_0^t V_s dM_s = V(t)M(t) - V(0)M(0) - \int_0^t M_s dV_s \tag{4.2}$$

*is a local martingale.*

*Proof.* We can find stopping times $\gamma_n$ such that both $|M^{\gamma_n}|$ is bounded by $n$ and the total variation of $V$,

$$R_V(t) \equiv \sup_{\{u_k\}} \sum_{k=0}^{m-1} |V(u_{k+1}) - V(u_k)| \tag{4.3}$$

is smaller than $n$. We have that

$$\int_0^t V_s dM_s^{\gamma_n} \equiv \lim_{u_k^n} \sum_{k=0}^{m-1} V(u_k^n)(M^{\gamma_n}(u_{k+1}^n) - M^{\gamma_n}(u_k^n)),$$

where $\{u_k^n\}$ is any sequence of partitions of $[0,t]$ such that $\max(|u_{k+1}^n - u_k^n|) \to 0$. This limit exists since by elementary reshuffling,

$$\sum_{k=0}^{m-1} V(u_k^n)(M^{\gamma_n}(u_{k+1}^n) - nM^{\gamma_n}(u_k^n)) = V^{\gamma_n}(t)M^{\gamma_n}(t) - V(0)M^{\gamma_n}(0)$$
$$- \sum_{k=0}^{m-1} M^{\gamma_n}(u_{k+1}^n)(V^{\gamma_n}(u_{k+1}^n) - V^{\gamma_n}(u_k^n)). \tag{4.4}$$

Since $V$ is of bounded variation, and $M^{\gamma_n}$ is bounded, the latter sum converges to $\int_0^t M_s dV_s$, both almost surely and in $\mathcal{L}^1$, as $n \uparrow \infty$. As a consequence the same holds true for the left-hand side, and, since for any finite $n$, the left hand side is a martingale, this property remains true in the limit $\uparrow \infty$. Finally, we pass to the limit $n \uparrow \infty$, which exists since $\gamma_n \uparrow \infty$ (and thus eventually will be larger than $t$, a.s.. □

We see that the challenge will be to define stochastic integrals when also the integrand is not of bounded variation. Before doing so we need to return briefly to the theory of martingales.

### 4.1.1 Square integrable continuous (local) martingales

Let $M$ be a càdlàg martingale. We want to define its quadratic variation process $[M]$ in analogy to the discrete time case. This will be contained in the following very fundamental proposition.

**Proposition 4.1.2** *Let $M$ be a continuous square integrable martingale. Then there exists a unique increasing process, $[M]$, such that the process $M^2 - [M]$ is a uniformly integrable continuous martingale.*

*Proof.* We will only consider the case when $M$ is continuous. We can also assume that $M$ is bounded; otherwise we consider the martingale stopped on exceeding a value $N$. Now define stopping times

$$T_0^n = 0, \quad T_{k+1}^n = \inf\{t > T_k^n : |M(t) - M(T_k^n) \geq 2^{-n}\}.$$

Set $t_k^n \equiv t \wedge T_k^n$. Then we can write (by telescopic expansions)

$$M_t^2 = 2\sum_{k\geq 1} M(t_{k-1}^n)(M(t_k^n) - M(t_{k-1}^n)) + \sum_{k\geq 1} \left(M(t_k^n) - M(t_{k-1}^n)\right)^2. \tag{4.5}$$

Let

$$H_t^n \equiv \sum_{k\geq 1} M(T_{k-1}^n) 1\!\!1_{T_{k-1}^n < t \leq T_k^n}.$$

Note that the process $H^n$ is *left-continuous*, which makes it previsible. This is of course the natural choice from the point of view that we want to define stochastic integrals that are martingales. Then the first term on the right of (4.5) is $H^n \bullet M$ (see [2], Chapter 4), and we know that this is a $L^2$-bounded martingale. We define

$$A_t^n \equiv \sum_{k\geq 1} \left(M(t_k^n) - M(t_{k-1}^n)\right)^2 . \tag{4.6}$$

Then

$$M_t^2 = 2(H^n \bullet M)_t + A_t^n.$$

By construction $H^n$ approximates $M$ very well:

$$\sup_t |H_t^n - H_t^{n+1}| \leq 2^{-n-1} \tag{4.7}$$

$$\sup_t |H_t^n - M_t| \leq 2^{-n} \tag{4.8}$$

The sets $J_n(\omega) \equiv \{T_k^n(\omega); k \in \mathbb{N}\}$ refine each other, i.e. $J_n(\omega) \subset J_{n+1}(\omega)$, and

$$A^n(T_k^n) \leq A^n(T_{k+1}^n). \tag{4.9}$$

Now it is elementary to see that

$$\begin{aligned}
\mathbb{E}[(((H^n - H^{n+1}) \bullet M)_\infty)^2] &= \mathbb{E}\sum_{k\geq 1}(H_{k-1}^n - H_{k-1}^{n+1})^2(M_k - M_{k-1})^2 \\
&\leq 2^{-2n-2}\mathbb{E}\sum_{k\geq 1}(M_k - M_{k-1})^2 \\
&= \mathbb{E}M_\infty^2. 
\end{aligned} \tag{4.10}$$

Thus the continuous martingales $(H^n \bullet M)$ converge, as $n \uparrow \infty$, uniformly to a continuous martingale, $N$. This implies that the processes $A^n$ converge to some continuous process $A$, and

$$M_t^2 = 2N_t + A_t.$$

Due to the fact that the sets $J_n$ form refinements and that $A_n$ increases on the stopping times $T_k^n$, it follows that

$$A(T_k^n) \leq A(T_{k+1}^n),$$

for all $k, n$. So $A$ is increasing on the closure of $J(\omega) \equiv \cup_n J_n(\omega)$. Thus if $J(\omega)$ is dense, $A$ is increasing. The remaining option is that the complement of $J(\omega)$ contains some open interval $I$. But in that case, since no $T_k^n$ in in $I$, $M$ must be constant on $I$, and so is then $A$. Thus $A$ is a continuous increasing process such that

$$M_t^2 - A_t$$

is a continuous martingale; hence $A = [M]$.

It remains to show the uniqueness of this process. For this we use the following (maybe surprising) lemma.

**Lemma 4.1.3** *If $M$ is a continuous (local) martingale that has paths of finite variation, then, if $M_0 = 0$, then $M_t = 0$ for all $t$.*

*Proof.* Again by stopping $M$ when at $\tau_n \equiv \inf\{t : V_M(t) > n\}$, where

$$V_M(t) = \lim_n \sum_k |M(u_k^\ell) - M(u_{k-1}^n)|$$

is the total variation process, we may assume that $M$ has bounded total variation. Then, obviously,

$$\begin{aligned} A_t^n &= \sum_k \left(M(t_k^n) - M(t_{k-1}^n)\right)^2 \\ &\leq 2^{-n} \sum_k \left|M(t_k^n) - M(t_{k-1}^n)\right| \leq 2^{-n} V_M(t) \end{aligned} \tag{4.11}$$

which tends to zero as $n \uparrow \infty$. Thus $M^2$ is a martingale. So $\mathbb{E}M_t^2 = 0$, for all $t$, and a positive random variable of zero mean is zero a.s. □

Now we derive uniqueness from this: Assume that there are two processes $A, A'$ with the desired properties. Then $A - A'$ is the difference of two uniformly integrable martingales, hence itself a uniformly integrable martingale. On the other hand, as $A$ and $A'$ are increasing and hence of finite variation, their difference is of finite variation, and thus identically equal to zero by the preceding lemma. □

**Remark 4.1.2** The condition that $M$ is square integrable is not necessary. One can extend the construction by considering the stopped martingales $M^{\tau_n}$ where $\tau_n = \inf\{t : |M_t \geq t\}$. $M^{\tau_n}$ is square integrable, and so $[M^{\tau_n}]$ exists. Moreover, we can set $[M]_t = [M^{\tau_n}]_t$ for $t \leq \tau_n$. Since $[M^{\tau_{n+1}}]_t = [M^{\tau_n}]_t$ for $t \leq \tau_n$, this construction can be extended consistently to all $t$ since $\tau_n \uparrow \infty$.

It will be convenient to know the following fact:

**Proposition 4.1.4** *Let $M$ be a càdlàg martingale. Then , for each*

*$t \geq 0$ and any sequence of partitions $\{u^n_k\}$, of the interval $[0,t]$ such that $\lim_{n\uparrow\infty} \max_k |u^n_k - u^n_{k-1}| = 0$,*

$$\sum_k \left(M(u^n_{k+1}) - M(u^n_k)\right)^2 \xrightarrow{\mathcal{D}} [M]_t. \tag{4.12}$$

*Moreover, if $M$ is square integrable, then the convergence also holds in $\mathcal{L}^1$.*

The proof of this proposition is somewhat technical and will not be given, but see e.g. [6].

Let us note that in the case when $M$ is Brownian motion, we have already seen in [2] that

**Lemma 4.1.5** *If $B_t$ is standard Brownian motion, then $[B]_t = t$.*

Let us recall from the discrete time theory that there were two brackets, $< M >$ and $[M]$ associated to a martingale: the first corresponds to the process given by Proposition 4.5, and the second is the quadratic variation process. In the case of continuous martingales, both are the same.

### 4.1.2 Stochastic integrals for simple functions

We have already seen that the stochastic integral can be defined as a Stieltjes integral for integrators of bounded variation. We will now show the crucial connection between the quadratic variation process of the stochastic integral and the process $[M]$ first in the case when the integrand, $X$ is a step function.

Let $\mathcal{E}_b$ be the space of all bounded step functions, i.e. functions $X$ of the form

$$X_t = \sum_{i\geq 1} X(i) 1\!\!1_{t_{i-1}<t\leq t_i},$$

for some sequence $0 = t_0 < t_1 < \cdots < t_n < \ldots$ and values $X_i \in \mathbb{R}$. Note that $X(t_i) = X_i$. Clearly then, our stochastic integral for such function is defined and equals

$$\int_0^t X dM = \sum_{i\geq 1; t_i \leq t} X_i(M(t_i) - M(t_{i-1})) + X_{m(t)}(M(t) - M(t_{m(t)}),$$

where $m(t) = \max(m : t_m \leq t)$.

The following lemma states the crucial properties of stochastic integrals.

**Lemma 4.1.6** *Let $M$ be a continuous square integrable martingale and $X \in \mathcal{E}$. Then $\int_0^t XdM$ as defined above is a continuous square integrable martingale and*

$$\left[\int_0^\cdot XdM\right]_t = \int_0^t X^2 d[M]. \tag{4.13}$$

*Proof.* We have already seen that $\int XdM$ is a martingale. To see that it is square integrable, note that

$$\mathbb{E}\left(\int_0^t XdM\right)^2 = \sum_{i\geq 0} \mathbb{E}\left[X_i^2(M(t_i) - M(t_{i-1}))^2\right] \leq C\mathbb{E}M_t^2 < \infty,$$

by assumption. To show (4.13), we have to show that

$$\left(\int_0^t XdM\right)^2 - \int_0^t X^2 d[M]$$

is a martingale. To prove this, we need to compute

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_t^{s+t} XdM\right)^2 - \int_2^t X^2 d[M]\Big|\mathcal{F}_t\right] \qquad (4.14)\\
&= \mathbb{E}\Bigg[\sum_{i,j} X_i X_j (M(t_{i+1}) - M_{t_i})(M(t_{j+1}) - M(t_j))\\
&\qquad - \sum_i X(t_i)^2([M]_{t_{i+1}} - [M]_{t_i})\Big|\mathcal{F}_t\Bigg]\\
&= \sum_i \mathbb{E}\Bigg[X_i^2 \mathbb{E}\left[(M(t_{i+1}) - M_{t_i})^2\big|\mathcal{F}_{t_i}\right]\\
&\qquad - X(t_i)^2 \mathbb{E}\left[([M]_{t_{i+1}} - [M]_{t_i})\big|\mathcal{F}_{t_i}\right]\Bigg]\Big|\mathcal{F}_t\Bigg]\\
&= 0,
\end{aligned}$$

since of course

$$\mathbb{E}\left[(M(t_{i+1}) - M_{t_i})^2\big|\mathcal{F}_{t_i}\right] = \mathbb{E}\left[([M]_{t_{i+1}} - [M]_{t_i})\big|\mathcal{F}_{t_i}\right].$$

This proves the lemma. □

The lemma states the key properties that we want the general stochastic integral to share. Naturally, our ambition will be to extend the integral to integrands $X$ for which the objects characterizing it make sense.

Note that, in particular, it follows from (4.13) that

$$\mathbb{E}\left(\int_0^t XdM\right)^2 = \mathbb{E}\int_0^t X^2 d[M]. \tag{4.15}$$

This means that the map $X \to \int_0^t XdM$, from the space of left-continuous step-functions equipped with the norm

$$\|X\|_{2,d[M]} \equiv \left(\mathbb{E}\int_0^t X^2 d[M]\right)^{1/2}$$

to the space of local, square integrable martingales with the norm $\mathcal{L}^2(\mathbb{P})$ is an *isometry*, called the *Itô isometry*. We will extend this isometry to to all of $\mathcal{L}^2(d[M])$ to define the *Itô integral* .

To do so we need an approximation result.

**Lemma 4.1.7** *If $M$ is a square integrable martingale and $X$ is in $\mathcal{L}^2([M])$. Then there exists a sequence of bounded, left-continuous step functions, $X_n$, such that*

$$\lim_{n\uparrow\infty} \mathbb{E}\int_0^t (X - X_n)^2 d[M] = 0. \tag{4.16}$$

*Proof.* We go in several steps. First, we can approximate any $X$ by bounded functions $X_n$: set $X_n(t) = X(t)\mathbb{1}_{|X_t|\leq n}$. Then $(X_n - X)^2 \downarrow 0$, so that the convergence in (4.1.1) follows by monotone convergence. Thus we may from now on assume that $X$ is bounded. For bounded $X$ we then construct the approximants (assume $n$ is so large that $N^{-1} < t$):

$$X_n(t) \equiv \frac{1}{[M]_t - [M]_{t-1/n} + n^{-1}} \int_{t-1/n}^t X_u(d[M]_u + du). \tag{4.17}$$

One verifies that in fact $\lim_{n\uparrow\infty} X_n(t) = X(t)$, while $X_n$ is continuous. Since $X$ is bounded, convergence as in (4.16) follows by dominated convergence. Thus we can assume that $X$ is continuous. In that case, we approximate

$$X_n(t) = X\left(\frac{[nt]}{n}\right), \tag{4.18}$$

which is a left-continuous step function if $[x] \equiv \min\{n \in \mathbb{N} : n \geq x\}$. Then again convergence as in (4.16) follows by dominated convergence. $\square$

We can now extend the definition of the stochastic integral.

**Theorem 4.1.8** *Let $M$ be a continuous, square integrable local martingale, and let $X \in \mathcal{L}^2(d[M])$. Then there exists a unique continuous square integrable local martingale, $\int_0^\cdot X dM$, such that, whenever a sequence of left-continuous step-functions, $X_n$, satisfies*

$$\sum_{n\in\mathbb{N}} \left\{ \mathbb{E}\left[\int_0^n (X_n - X)^2 d[M]\right]\right\}^{1/2} < \infty, \tag{4.19}$$

*then*

$$\lim_{n\uparrow\infty} \sup_{0\le t\le T} \left| \int_0^t (X_n - X) dM_s \right| = 0, \tag{4.20}$$

*almost surely and in $\mathcal{L}^2$. Moreover,*

$$\left[\int_0^\cdot X dM\right]_t = \int_0^t X^2 d[M]. \tag{4.21}$$

*Proof.* Note first that, by taking subsequences, Lemma 4.1.7 implies that we can always find sequences of step functions that satisfy (4.19). Hence

$$\begin{aligned} &\mathbb{E}\sum_n \sup_{0\le t\le T} \left| \int_0^t X_{n+1} dM - \int_0^t X_n dM \right| \qquad (4.22)\\ &\le \sum_n \left( \mathbb{E}\left[ \sup_{0\le t\le T} \left| \int_0^t (X_{n+1} - X_n) dM \right|^2 \right] \right)^{1/2} \\ &\le \sum_n \left( 4\mathbb{E}\left[ \left| \int_0^T (X_{n+1} - X_n) dM \right|^2 \right] \right)^{1/2} \\ &\le 2\sum_n \left( \mathbb{E}\left[ \int_0^T (X_{n+1} - X_n)^2 d[M] \right] \right)^{1/2} < \infty \end{aligned}$$

Here we used the maximum inequality and the finiteness of the last expression follows from the assumption (4.19).

It follows from the Borel-Cantelli lemma that there exists a set, $A \in \mathcal{F}$, of measure zero, such that

$$\mathbb{1}_{A^c} \int_0^\cdot X_n dM$$

converges uniformly on bounded time intervals. The limiting process is continuous, square integrable and adapted (since we assumed completeness of $\mathcal{F}_0$). Thus we have (4.20) almost surely. To prove uniform convergence in $\mathcal{L}^2$, note that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t (X_n - X)dM\right|^2\right] \qquad (4.23)\\
&= \mathbb{E}\left[\liminf_{m\uparrow\infty}\sup_{0\le t\le T}\left|\int_0^t (X_n - X_m)dM\right|^2\right]\\
&\le \liminf_{m\uparrow\infty}\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t (X_n - X_m)dM\right|^2\right]\\
&\le \liminf_{m\uparrow\infty} 4\mathbb{E}\left[\left|\int_0^T (X_n - X_m)dM\right|^2\right]\\
&= \liminf_{m\uparrow\infty} 4\mathbb{E}\left[\int_0^T (X_n - X_m)^2 d[M]\right]\\
&= 4\mathbb{E}\left[\int_0^T (X_n - X)^2 d[M]\right]
\end{aligned}
$$

which converges to zero as $n \uparrow \infty$. Thus convergence in $\mathcal{L}^2$ holds for (4.20).

Finally, that fact that $\int_0^\cdot XdM$ is a local martingale follows form the fact that this holds for the approximants and the uniform convergence we have just established. Similarly, the formula for the bracket follows.

□

**Remark 4.1.3** Theorem 4.19 extends the isometry $X \to \int X_d M$ from the dense set of left-continuous bounded step functions to the full space $\mathcal{L}^2(d[M])$.

**Remark 4.1.4** Theorem 4.1.8 is not the end of the possible extension of the definition of stochastic integrals. Using localization arguments as indicated in the definition of the bracket $[M]$, one can extend the space of integrators to continuous local martingales without the assumption of square integrability.

## 4.2 Itô's formula

We now come to the most useful formula involving the notion of stochastic integrals, the celebrated *Itô formula* . It is the analog of the fundamental theorem of calculus for functions of stochastic processes with unbounded variation.

We consider a stochastic process $X$ of the form

$$X_t = X_0 + V_t + M_t, \tag{4.24}$$

where $V_t$ is a continuous, adapted process of bounded variation, $M_t$ is a local martingale (you may assume $\mathcal{L}^2$, but see the remark above), and $V_0 = M_0 = 0$. Let $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be continuously differentiable in the first and twice continuously differentiable in the second argument.

**Theorem 4.2.9** *With the assumptions above, the following holds:*

$$\begin{aligned} f(t, X_t) - f(0, X_0) \qquad\qquad & \qquad\qquad\qquad\qquad (4.25) \\ = \int_0^t \frac{\partial}{\partial s} f(s, X_s) ds + \int_0^t \frac{\partial}{\partial x} f(s, X_s) dV_s & \\ + \int_0^t \frac{\partial}{\partial x} f(s, X_s) dM_s & \\ + \frac{1}{2} \int_0^t \frac{\partial^2}{\partial x^2} f(s, X_s) d[M]_s. & \end{aligned}$$

**Remark 4.2.1** The Itô formula can be stated more conveniently in differential form as

$$\begin{aligned} df(t, X_t) = \frac{\partial}{\partial t} f(t, X_t) dt + \frac{\partial}{\partial x} f(t, X_t) dX_t \\ + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, X_t) d[X]_t, \end{aligned} \tag{4.26}$$

with the understanding that $d[X] = d[M]$, since the quadratic variation of the finite variation process $V$ is zero.

*Proof.* As usual, we first localize. Let

$$\tau_n \equiv \inf \{t \geq 0 : (|X_0| + |M_t| + R_V(t)) \geq n\}.$$

Then $\tau_n$ are stopping times tending to infinity, and we can prove first (4.25) with $t$ replaced by $t \wedge \tau_n$, and then let $n$ tend to infinity to extend the result to all $t$. Thus in the sequel we can assume $M$ bounded and $V$ of bounded variation. Let $\{t_k\}$ be a partition of $[0, t]$, and set $\Delta_k X \equiv X_{t_k} - X_{t_{k-1}}$, etc.. Then

$$
\begin{aligned}
f(t, X_t) - f(0, X_0) & \qquad\qquad (4.27)\\
&= \sum_{k=0}^{m-1} \left( f(t_{k+1}, X_{t_{k+1}}) - f(t_k, X_{t_k}) - f(t_k, X_{t_{k+1}}) + f(t_k, X_{t_{k+1}}) \right) \\
&= \sum_{k=0}^{m-1} \left( \int_{t_k}^{t_{k+1}} \frac{\partial}{\partial t} f(u, X_{k+1}) du \right. \\
&+ \left. \frac{\partial}{\partial x} f(t_k, X_k) \Delta_k X + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t_k, \xi_k)(\Delta_k X)^2 \right),
\end{aligned}
$$

for some $\xi_k$ with $|X_{t_k} - \xi_k| \leq |\Delta_k X|$, by Taylor's theorem. Clearly, as we refine the partition $\{t_k\}$, the first two terms tend to the integrals, resp. stochastic integrals appearing in the Itô formula. It is not very difficult to see that the last term produces the integral of $\frac{\partial^2}{\partial x^2} f(t, x)$ with respect to the bracket of $M$. To see this, note first that

$$
\begin{aligned}
\sum_k \Delta X_k^2 = \sum_k (V_{t_{k+1}} - V_{t_k})^2 + 2 \sum_k (V_{t_{k+1}} - V_{t_k})(M_{t_{k+1}} - M_{t_k}) \\
+ \sum_k (M_{t_{k+1}} - M_{t_k})^2 \qquad (4.28)
\end{aligned}
$$

If we take a sequence of partitions such that $\max_k |t_{k+1} - t_k| \downarrow 0$, then the first two terms clearly tend to zero (since $V$ has bounded variation and $M$ is continuous, so $|M_{t_{k+1}} - M_{t_k}|$ tends to zero. Also, since $f$ is $C^2$ and $X$ is continuous, it follows that

$$
\max_k \left| \frac{\partial^2}{\partial x^2} f(t_k, \xi_k) - \frac{\partial^2}{\partial x^2} f(t_k, X_{t_k}) \right| \downarrow 0. \qquad (4.29)
$$

Thus we are left to show that

$$
\sum_k \frac{\partial^2}{\partial x^2} f(t_k, X_{t_k}) \left( M_{t_{k+1}} - M_{t_k} \right)^2 \to \int \frac{\partial^2}{\partial x^2} f(s, X_s) d[M]_s. \qquad (4.30)
$$

But this is relatively straightforward. □

To see how useful the Itô formula can be, we will use it to prove Lévy's famous theorem that says that Brownian motion can be characterized as the unique local martingale whose bracket is equal to $t$.

**Theorem 4.2.10** *Let $X$ be a continuous local martingale such that $[X]_t = t$. Then $X$ is Brownian motion.*

*Proof.* Let $f(t,x) \equiv \exp\left(i\theta x + \frac{1}{2}\theta^2 t\right)$. Clearly $f$ (reps. the real and imaginary parts of $f$) satisfies the hypothesis of Theorem 4.2.9. Since $X$ is a local martingale, and

$$\frac{\partial}{\partial t} f(t,x) = \frac{1}{2}\theta^2 f(t,x) = -\frac{1}{2}\frac{\partial^2}{\partial x^2} f(t,x),$$

and since $d[x]_t = dt$, by hypothesis, Itô's formula implies that $f(t, X_t)$ is a local martingale, i.e. one checks that

$$\mathbb{E}\left[f(t+s, X_{t+s})|\mathcal{F}_t\right] = f(t, X_t) \tag{4.31}$$

Writing this out implies that

$$\mathbb{E}\left[e^{i\theta(X_{t+s}-X_t)}|\mathcal{F}_t\right] = e^{-\frac{s}{2}\theta^2},$$

for all $\theta, s, t$, so the increments of $X$ are independent and Gaussian, implying the $X$ is Brownian motion. □

## 4.3 Black-Sholes formula and option pricing

In this section we give a derivation of the Black-Sholes formula as a simple application of Itô's formula.

The basic idea of option pricing can be expressed in a rather fundamental intrinsically mathematical way. A stochastic integral can be interpreted as the evolution of the wealth of an investor who invests according to a previsible strategy $C$ in a stock whose price evolves as a continuous martingale, $M$ (we disregard here interest rates or inflation). An (European) *option* is a function, $F : S \to \mathbb{R}$, that corresponds to a payoff of an amount of money, $F(M_T)$, at fixed time $T$. If a bank engages in such a contract, it must ensure that it will charge a price for this option that allows it, by following a previsible investment strategy, to procure the payoff $F(M_T)$ at the end of the period from the proceeds of the received option price. Thus the issue is whether we can represent the payoff as a sum of initial price, $\Pi$, plus a consecutive wealth process:

$$F(M_T) = \int_0^T C dM + \Pi_0, \text{ a.s.} \tag{4.32}$$

where of course $\Pi_0$ should be minimal. In purely mathematical terms, this corresponds to asking for a representation formula of the random variable $F(M_T)$.

Let us now see how the Itô formula relates to the issue of option pricing. As we said, on option is a contract that guarantees the pay-out of an amount $F(M_T)$ at time $T$. Thus, the value, $V(T, M_T)$, at time

$T$ is precisely $F(M_T)$. We are interested to know that the value of the option is at earlier times $t < T$, given the stock price $M_t$. To this end we consider the value function $V(t, M)$ as a function of two variables. Then Itô's formula tells us that

$$dV(t, M_t) = \frac{\partial}{\partial t} V(t, M_t) dt \\ + \frac{\partial}{\partial M} V(t, M_t) dM_t \\ + \frac{1}{2} \frac{\partial^2}{\partial M^2} V(t, M_t) d[M]_t \tag{4.33}$$

An investment strategy can replicate the martingale part, $\frac{\partial}{\partial M} V(t, M_t) dM_t$, whereas the other terms should cancel. I.e. we will demand that

$$0 = \frac{\partial}{\partial t} V(t, M) dt \\ + \frac{1}{2} \frac{\partial^2}{\partial M^2} V(t, M) d[M]_t \tag{4.34}$$

Note that if $M_t$ is exponential Brownian motion, then

$$dM_t = \sigma(t) M_t dB_t$$

and

$$d[M]_t = \sigma^2(t) M_t^2 dt$$

so that the differential equation becomes

$$0 = \frac{\partial}{\partial t} V(t, M) + \frac{1}{2} \frac{\partial^2}{\partial M^2} V(t, M) \sigma^2(t) M^2 \tag{4.35}$$

This equation can now be considered as a partial equation for the function $V$ with the final conditions

$$V(T, M) = F(M).$$

Let us now see that this then gives the desired reproduction strategy. Take an initial amount of capital $X_0 = V(0, M_0)$. Invest according to the strategy $C(t, M_t) = \frac{\partial}{\partial M} V(t, M_t)$ in $M$. As a result, at time $T$, you will have accumulated the wealth

$$X(T) = \int_0^T \frac{\partial}{\partial M} V(t, M_t) dM_t + X_0. \tag{4.36}$$

But according to (4.33), since (4.34) holds,

$$F(M_T) = V(T, M_T) = V(0, M_0) + \int_0^T \frac{\partial}{\partial M} V(t, M_t) dM_t = X(T)! \tag{4.37}$$

Thus our investment strategy has produced exactly the desired amount of money needed to cover the pay-out of the option. $V(0, M_0)$ is then the reasonable price for the option.

## 4.4 Girsanov's theorem

Girsanov's theorem is a particularly useful result to study properties of processes that can be seen as modifications of Brownian motions. For simplicity I will consider only the one-dimensional situation, but the obvious extensions to multi-dimensional settings hold true as well.

Suppose we are living on a filtered space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$ that satisfies the usual assumptions and we are given a Brownian motion $B$ and an adapted process $X$ that is square integrable with respect to $dt$, i.e. an integrand for Brownian motion.

Suppose we want to study the process

$$W_t \equiv B_t - \int_0^t X_s ds. \tag{4.38}$$

For example, we could think of the case $X_s = b(s, B_s)$, for some bounded measurable function (as in the last section). The simplest case of course would be $b(s, X_s) = b$, so

$$W_t = B_t - bt,$$

which is Brownian motion with a constant drift $b$.

How can we compute properties of $W$? In particular, can we find a new probability measure, $\tilde{P}$, such that under this new measure, $W$ becomes simple? Girsanov's theorem is a striking affirmative answer to this question.

**Theorem 4.4.11** *Let $B, X, W$ be as above and define*

$$Z_t(X) \equiv \exp\left(\int_0^t X dB - \frac{1}{2}\int_0^t X_s^2 ds\right) \tag{4.39}$$

*and let $\widetilde{\mathbb{P}}$ be defined by*

$$\widetilde{\mathbb{P}}_T(A) \equiv \mathbb{E}\left[Z_T(X) \mathbb{1}_A\right]. \tag{4.40}$$

*Then, if $Z$ is a martingale, the process $W_t$, $t \leq T$ is a Brownian motion under $\widetilde{\mathbb{P}}_T$.*

**Remark 4.4.1** One can check using Itô's formula that $Z_t$ solves

$$dZ_t = Z_t X_t dB_t$$

and hence is always a positive local martingale, and so, by Fatou's lemma, a super-martingale. It is a martingale whenever $\mathbb{E}Z_t = 1$.

*Proof.* Let us first show a more abstract looking result. To formulate this, and to proof Girsanov's theorem, we need to introduce the notion of a bracket between two martingales:

$$[M, N]_t \equiv [M + N]_t - \frac{1}{2}\left([N]_t + [M]_t\right) \tag{4.41}$$

One may verify that with this notion, one has the following generalization of Itô's formula to the case of functions of several variables:

$$\begin{aligned} f(t, X_t) &- f(0, X_0) \qquad\qquad (4.42)\\ &= \int_0^t \frac{\partial}{\partial s} f(s, X_s) ds + \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x_i} f(s, X_s) dV_i(s) \\ &+ \sum_{i=1}^d \int_0^t \frac{\partial}{\partial x_i} f(s, X_s) dM_i(s) \\ &+ \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} f(s, X_s) d[M_i, M_j]_s. \end{aligned}$$

**Lemma 4.4.12** *Let $M$ be a continuous local martingale and let $Z_t \equiv \exp\left(M_t - \frac{1}{2}[M]_t\right)$. Assume that $Z$ is uniformly integrable. Let $\mathbb{Q}$ be the measure that is absolutely continuous with respect to $\mathbb{P}$ such that the Radon-Nikodým derivative $\frac{d\mathbb{Q}}{dP} = Z_\infty$. Then, if $X$ is a continuous local martingale under $\mathbb{P}$; then $X - [X, M]$ is a local martingale under $\mathbb{Q}$.*

*Proof.* As usual we stop at a time $\tau_n \equiv \inf\{t \geq 0 : |X_t| + [M, N]_t \geq n\}$. By assumption, $Z_t$ is a uniformly integrable martingale and $Z_t = \mathbb{E}[Z_\infty | \mathcal{F}_t]$, a.s.. Let

$$Y \equiv X^{\tau_n} - [X^{\tau_n}, M].$$

Note that $Z$ is the solution of the stochastic differential equation

$$dZ_t = Z_t dM_t,$$

which can be verified using Itô's formula. Next we use Itô's formula (4.42) to compute

$$
\begin{aligned}
d(Z_tY_t) &= Z_t dY_t + Y_t dZ_t + d[Z,Y]_t \qquad (4.43)\\
&= Z_t(dX_t - d[X,M]_t) + (X_t - [X,M]_t)Z_t dM_t + d[Z,Y]\\
&= Z_t(dX_t - d[X,M]_t) + (X_t - [X,M]_t)Z_t dM_t + Z_t d[M,X]\\
&= Z_t dX_t + (X_t - [X,M]_t)Z_t dM_t.
\end{aligned}
$$

Here we used that first $[Z,Y] = [Z,X]$, since $Y - X$ is of bounded variation, and then the fact that $Z = \int Z dM$, and finally the theorem of Kunita-Watanabe that states that

$$\left[\int Z dM, X\right] = \left[\int Z dM, \int dX\right] = \int Z d[M,X],$$

which extends the formula for the bracket of a stochastic integral to that of the co-bracket of two such integrals in a natural way. Thus $ZY$ is a stochastic integral and hence a martingale under $\mathbb{P}$. Therefore, for $A \in \mathcal{F}_s$,

$$\mathbb{E}_{\mathbb{Q}}\left[(Y_t - Y_s)\mathbb{1}_A\right] = \mathbb{E}\left[(Z_\infty Y_t - Z_\infty Y_s)\mathbb{1}_A\right] = \mathbb{E}\left[(Z_tY_t - Z_sY_s)\mathbb{1}_A\right] = 0, \qquad (4.44)$$

and so $Y_t$ is a martingale under $\mathbb{Q}$. Thus the un-stopped $X - [X,M]$ is a local martingale. □

We can now conclude the proof of Theorem 4.4.11 rather easily. We see that we are in the setting of Lemma 4.4.12 with $X_t = B_t$, $M_t = \int_0^t X dB$ and $Y_t = W_t = B_t - \int_0^t X d[B] = B_t - \int_0^t X ds$. Thus we know that $W_t$ is a local martingale. To show that it is Brownian motion, it suffices to compute its bracket. But since $\int_0^t X_s ds$ is an ordinary integral, it is of bounded variation and hence $[W]_t = [B]_t = t$, so $W$ as a continuous local martingale with bracket $t$ is Brownian motion by Lévy's theorem. □

In the special case when $W_t = B_t - bt$, $Z_t = \exp\left(bB_t - \frac{1}{2}b^2 t\right)$.

Let us now consider a Brownian motion $B_t$ in $\mathbb{R}$ and let for $b \in \mathbb{R}$ $\tau_b$ be the first hitting time of $b$, $T_b \equiv \inf\{t > 0 : B_t \geq b\}$. Using a simple symmetry argument and the strong Markov property, one can show that

$$\mathbb{P}_0[T_b < t] = 2\mathbb{P}_0[B_t \geq b] = \sqrt{\frac{2}{\pi}} \int_{b/\sqrt{t}}^{\infty} e^{-x^2/2} dx, \qquad (4.45)$$

and hence the probability density of this variable is

$$\mathbb{P}_0[T_b \in dt] = \frac{|b|}{\sqrt{2\pi t^2}} e^{-b^2/2t},$$

and, for $a \geq 0$,

$$\mathbb{E}e^{-aT_b} = e^{-|b|\sqrt{2a}}.$$

Now consider $W_t = B_t - \mu t$. Let $Z_t \equiv e^{\mu B_t - \mu t^2/2}$. The we have that under the measure $\mathbb{P}^\mu$, defined by $P^\mu(A) = \mathbb{E}Z_t \mathbb{1}_A$, the process $W_t$ is a Brownian motion, or $B_t$ under $\mathbb{P}^\mu$ is a Brownian motion with drift $\mu$. Since on the set $\{T_b \leq t\}$, $T_{t\vee T_b} = Z_{T_b}$, the optional sampling theorem implies

$$\begin{aligned} \mathbb{P}^\mu[T_b \leq t] &= \mathbb{E}1_{\{T_b \leq t\}} Z_t \\ &= \mathbb{E}[1_{\{T_b \leq t\}} \mathbb{E}[Z_t | \mathcal{F}_{T_b \vee t}]] \\ &= \mathbb{E}[1_{\{T_b \leq t\}} Z_{T_b \vee t}] \\ &= \mathbb{E}[1_{\{T_b \leq t\}} Z_{T_b}] \\ &= \mathbb{E}\left[1_{\{T_b \leq t\}} e^{\mu b - \frac{1}{2}\mu^2 T_b}\right] \\ &= \int_0^t e^{\mu b - \frac{1}{2}\mu^2 s} \mathbb{P}_0[T_b \in ds] \end{aligned} \tag{4.46}$$

Differentiating, we get that

$$\mathbb{P}_0^\mu[T_b \in dt] = \frac{|b|}{\sqrt{2\pi t^2}} e^{-(b-\mu t)^2/2t}. \tag{4.47}$$

One can also conclude that

$$\mathbb{P}_0[T_b < \infty] = e^{\mu b - |\mu b|},$$

so that if $b$ and $\mu$ have the same sign, the drifted Brownian motion reaches the level $b$ with probability one, whereas in the opposite case the level $b$ is hit with probability strictly smaller than one.

**Novikov's condition** As we have noted, the process $Z$ in Girsanov's construction is a martingale if and only if $\mathbb{E}Z_t = 1$, for all $t \in \mathbb{R}_+$. We need verifiable criteria for this to hold. The following proposition that we take from [12], gives such a criterion, also known as Novikov's condition.

**Proposition 4.4.13** *Let $M$ be a continuous local martingale starting in zero. If*

$$\mathbb{E}\exp\left(\frac{1}{2}[M]_\infty\right) < \infty, \tag{4.48}$$

*for all $t \in \mathbb{R}_+$, then $Z_t$ is a unifomrly integrable martingale.*

*Proof.* We show first that $M$ is a uniformly integrable martingale and that $\mathbb{E}\exp\left(\frac{1}{2}M_\infty\right) < \infty$. In fact, (4.48) implies that $M$ is a martingale bounded in $\mathcal{L}^2$, and hence uniformly integrable. Next,

$$\exp\left(\tfrac{1}{2}M_\infty\right) = Z_\infty^{1/2}\exp\left(\tfrac{1}{4}[M]_\infty\right)$$

so that by the Cauchy-Schwarz inequality,

$$\mathbb{E}\exp\left(\tfrac{1}{2}M_\infty\right) \leq [\mathbb{Z}_\infty]^{1/2}\left[\mathbb{E}\exp\left(\tfrac{1}{2}[M]_\infty\right)\right]^{1/2} \leq \left[\mathbb{E}\exp\left(\tfrac{1}{2}[M]_\infty\right)\right]^{1/2} < \infty.$$

Now since $M_t$ is a uniformly integrable martingale, $M_t = \mathbb{E}\left[L_\infty|\mathcal{F}_t\right]$, and

$$\exp\left(\tfrac{1}{2}M_t\right) \leq \mathbb{E}\left[\exp\left(\tfrac{1}{2}M_\infty\right) \mid |\mathcal{F}_t\right].$$

Therefore, $\exp\left(\frac{1}{2}M_t\right)$ is in $\mathcal{L}^1$ and a submartingale. Then for any stopping time, $T$,

$$\exp\left(\tfrac{1}{2}M_T\right) \leq \mathbb{E}\left[\exp\left(\tfrac{1}{2}M_\infty\right) \mid |\mathcal{F}_T\right],$$

which shows that the family $\{\exp\left(\frac{1}{2}L_T\right), T \text{ stopping time}\}$ is uniformly integrable. Now set, for $0 < a < 1$, $Y_t^{(a)} \equiv \exp\left(\frac{aM_t}{1+a}\right)$ and $Z_t^{(a)} \equiv \exp\left(aM_t - \frac{a^2}{2}[M_t]\right)$. Then

$$Z_t^{(a)} = Z_t^{a^2}\left(Y_t^{(a)}\right)^{1-a^2}.$$

Then for $A \in \mathcal{F}_\infty$ and $T$ a stopping time, by Hölder's inequality

$$\mathbb{E}\left[\mathbb{1}_A Z_T^{(a)}\right] \leq \mathbb{E}\left[Z_T\right]^{a^2}\mathbb{E}\left[1_A Y_T^{(a)}\right]^{1-a^2} \leq \mathbb{E}\left[1_A Y_T^{(a)}\right]^{1-a^2} \leq \mathbb{E}\left[\mathbb{1}_A \exp\left(\tfrac{1}{2}M_T\right)\right]^{2a(1-a)}$$

where the second inequality used that $Z$ is a submartingale and the last is Jensen's inequality. This implies that the family $\{Z_T^{(a)}, T \text{ stopping time}\}$ is uniformly integrable. Hence $Z^{(a)}x$ is a uniformly integrable martingale. It follows that

$$1 = \mathbb{E}\left[Z_\infty^{(a)}\right] \leq \mathbb{E}\left[Z_\infty\right]^{a^2}\mathbb{E}\left[\exp\left(\tfrac{1}{2}M_\infty\right)\right]^{2a(1-a)}.$$

Turning this arround and letting $a \uparrow 1$, we get $\mathbb{E}\left[Z_\infty\right] \geq 1$, hence $\mathbb{E}\left[Z_\infty\right] = 1$. □

# 5

# Stochastic differential equations

## 5.1 Stochastic integral equations

We will define the notion of stochastic differential equations first.

We want to construct stochastic processes where the velocities are given as functions of time and position, and that have in addition a stochastic component. We will consider the case where the stochastic component comes from a Brownian motion, $B_t$. Such an equation should look like

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{5.1}$$

with prescribed initial conditions $X_0 = x_0$. The interpretation of such an equation is not totally straightforward, due to the term $\sigma(t, X_t)dB_t$. We will interpret such an equation as the integral equation

$$X_t = x_0 + \int_0^t b(s, X(s))ds + \int_0^t \sigma(s, X(s))dB_s, \tag{5.2}$$

where the integral with respect to $B$ is understood as the Itô stochastic integral defined in the last chapter. The functions $b, \sigma$ are in the most general setting assumed to be locally bounded and measurable.

The questions one is of course interested are those of existence and uniqueness of solutions to such equations, as well as that of properties of solutions. We begin by discussing the notions of strong and weak solutions.

## 5.2 Strong and weak solutions

We will denote by $W$ the Polish space $C(\mathbb{R}_+, \mathbb{R}^n)$ of continuous paths and we denote by $\mathcal{H}$ the corresponding Borel-$\sigma$-algebra, and by $\mathcal{H}_t \equiv \sigma\{x_s, s \leq t\}$ the filtration generated by the paths up to time $t$.

The formal set-up for a stochastic differential equation involves an initial conditions and a Brownian motion, all of which require a probability space. We will denote this by

$$(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \xi, B), \tag{5.3}$$

where

(i) $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ is a filtered space satisfying the usual conditions;
(ii) $B$ is a Brownian motion (on $\mathbb{R}^d$), adapted to $\mathcal{F}_t$,
(iii) $\xi$ is a $\mathcal{F}_0$-measurable random variable.

The minimal or *canonical* set-up has $\Omega = \mathbb{R}^n \times W$, $\mathbb{P} = \mu \times \mathbb{Q}$, where $\mu$ is the law of $\xi$ and $\mathbb{Q}$ is Wiener measure and $\mathcal{F}_t$ the usual augmentation of $\mathcal{F}_t^0 \equiv \sigma\{\xi, B_s, s \leq t\}$.

The precise definition of *path-wise uniqueness* of a SDE is as follows:

**Definition 5.2.1** For a SDE, path-wise uniqueness holds, if the following holds: For any set-up $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \xi, B)$, and any two continuous semi-martingales $X$ and $X'$, such that

$$\int_0^t (|b(s, X_s)| + |\sigma(s, X_s)|^2) ds < \infty, \tag{5.4}$$

and the same condition for $X'$ hold and both processes solve the SDE with this initial condition $\xi$ and this Brownian motion $B$,

$$\mathbb{P}[X_t = X'_t, \quad \forall t] = 1. \tag{5.5}$$

If a SDE admits for any setup $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \xi, B)$ exactly one continuous semi-martingale as solution, we say that the SDE is *exact.*

The notion of *strong solutions* is naturally associated with the setting of exact SDE's.

**Definition 5.2.2** A strong solution of a SDE is a function,

$$F : \mathbb{R}^n \times W \to W, \tag{5.6}$$

such that

$$F^{-1}(\mathcal{H}_t) \subset \mathcal{B}(\mathbb{R}^n) \times \bar{\mathcal{H}}_t, \forall t \geq 0, \tag{5.7}$$

and on any set-up $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \xi, B)$, the process

$$X = F(\xi, B)$$

solves the SDE. $\bar{\mathcal{H}}_t$ is the augmentation of $\mathcal{H}_t$ with respect to the Wiener measure.

Existence and uniqueness results in the strong sense can be proven in a very similar way as in the case of ordinary differential equations, using Gronwall's inequality and the Picard iteration scheme.

The general approach is to assume local Lipshitz conditions, to prove existence of solutions for finite times, and then glue solutions together until a possible explosion.

Let us give the basic uniqueness and existence results, essentially due to Itô.

**Theorem 5.2.1** *Assume that $\sigma$ and $b$ are bounded measurable, and that in addition there exists an open set $U \subset \mathbb{R}$, and $T > 0$, such that there exists $K < \infty$, s.t.*

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y|, \tag{5.8}$$

*for all $x, y \in U, t < T$. Let $X, Y$ be two solutions of (5.2) (with the same Brownian motion $B$), and set*

$$\tau \equiv \inf\{t \geq 0 : X_t \notin U or Y_t \notin U\}. \tag{5.9}$$

*Then, if $\mathbb{E}[X_0 - Y_0]^2 = 0$, it follows that*

$$\mathbb{P}\left[X(t \wedge \tau) = Y(t \wedge \tau), \forall 0 \leq t \leq T\right] = 1. \tag{5.10}$$

*Proof.* The proof is based on Gronwall's lemma and very much like the deterministic analog. We compute

$$\begin{aligned}
&\mathbb{E}\left[\max_{0\leq s\leq t} \left(X(s\wedge\tau) - Y(s\wedge\tau)\right)^2\right] \qquad (5.11)\\
&\leq 2\mathbb{E}\left[\max_{0\leq s\leq t}\left(\int_0^{s\wedge\tau} (\sigma(u,X(u)) - \sigma(u,Y(u)))dB_u\right)^2\right]\\
&+2\mathbb{E}\left[\max_{0\leq s\leq t}\left(\int_0^{s\wedge\tau} (b(u,X(u)) - b(u,Y(u)))du\right)^2\right]\\
&\leq 8\mathbb{E}\left[\int_0^{t\wedge\tau} (\sigma(u,X(u)) - \sigma(u,Y(u)))^2 du\right]\\
&+2t\mathbb{E}\left[\int_0^{t\wedge\tau} (b(u,X(u)) - b(u,Y(u)))^2 du\right]\\
&\leq 2K^2(t+4)\mathbb{E}\left[\int_0^{t\wedge\tau} \left(X(u) - Y(u)\right)^2 du\right]\\
&\leq 2K^2(4+t)\int_0^t \mathbb{E}\left[\max_{0\leq u\leq s} \left(X(u\wedge\tau) - Y(u\wedge\tau)\right)^2 ds\right].
\end{aligned}$$

Note that in the first inequality we used that $(a+b)^2 \leq 2a^2 + 2b^2$, in the second we used the Schwartz inequality for the drift term and Doob's $L^2$-maximum inequality for the diffusion term; the next inequality uses the Lipshitz condition and in the last we used Fubini's theorem.

Gronwall's inequality then implies that

$$\mathbb{E}\left[\max_{0\leq t\leq T}\left(X(t\wedge\tau)-Y(t\wedge\tau)\right)^2\right]=0.$$

This is most easily proven as follows: Let $f$ be a non-negative function that satisfies the integral equation $f(t)\leq K\int_0^t f(s)ds$. Set $F(t)=\int_0^t f(s)ds$. Then

$$0\leq\frac{d}{dx}\left(e^{-tK}F(t)\right)\leq e^{-Kt}\left(-KF(t)+f(t)\right)\leq 0,$$

and hence $e^{-tK}F(t)\leq 0$, meaning that $F(t)\leq 0$. But since $F$ is the integral of the non-negative function $f$, this means that $f(t)=0$.

Thus we have in particular that $\mathbb{P}[\max_{0\leq t\leq T}|X_t-Y_t|=0]=1$ as claimed. □

Finally, existence of solutions (for finite times) can be proven by the usual Picard iteration scheme under Lipschitz and growth conditions.

**Theorem 5.2.2** *Let $b,\sigma$ satisfy the Lipshitz conditions (5.8) and assume that*

$$|b(t,x)|^2+|\sigma(t,x)|^2\leq K^2(1+|x|^2). \tag{5.12}$$

*Let $\xi$ be a random vector with finite second moment, independent of $B_t$, and let $\mathcal{F}_t$ be the usual augmentation, $\mathcal{F}_t$, of the filtration associated with $B$ and $\xi$. Then there exists a continuous, $\mathcal{F}_t$-adapted process $X$ which is a* strong solution *of the SDE with initial condition $\xi$. Moreover, $X$ is square integrable, i.e. for any $T>0$, there exists $C(T,K)$, such that, for all $t\leq T$,*

$$\mathbb{E}|X_t|^2\leq C(K,T)(1+\mathbb{E}|\xi|^2)e^{C(K,T)t}. \tag{5.13}$$

*Proof.* We define a map, $F$, from the space of continuous adapted processed $X$, uniformly square integrable on $[0,T]$, to itself, via

$$F(X)_t\equiv\xi+\int_0^t b(s,X_s)ds+\int_0^t\sigma(s,X_s)dB_s. \tag{5.14}$$

Note that the square integrability of $F(X)$ needs the growth conditions (5.12)

Exercise: Prove this!

As in (5.11)

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{0\le t\le T}(F(X)_t - F(Y)_t)\right)^2 \qquad (5.15)\\
&\le 2E\left(\sup_{0\le t\le T}\left(\int_0^t (\sigma(X_s)-\sigma(Y_s))dB_s\right)^2\right)\\
&+2E\left(\sup_{0\le t\le T}\left(\int_0^t (b(X_s)-b(Y_s))d_s\right)^2\right)\\
&\le 2K^2(1+T)\int_0^T \mathbb{E}\sup_{0\le s\le t}(X_s-Y_s)^2\,dt
\end{aligned}
$$

and hence

$$
\mathbb{E}\left(\sup_{0\le t\le T}(F^k(X)_t - F^k(Y)_t)\right)^2 \le \frac{C^kT^{2k}}{k!}\mathbb{E}\left(\sup_{0\le t\le T}(X_t-Y_t)\right)^2. \qquad (5.16)
$$

Thus, for $n$ sufficiently large, $F^n$ is a contraction, and hence has a unique fixed point which solves the SDE. □

**Remark 5.2.1** The conditions for existence above are not necessary. In particular, growth conditions are important only when the solutions can actually reach the regions there the coefficients become too big. Formulations of weaker hypothesis for existence and uniqueness can be found for instance in [10], Chapter 14. Their verification in concrete cases can of course be rather tricky.

We will now consider a weaker form of solutions, in which the solution is not constructed from the BM, but the BM comes from the solution. This is like in the martingale problem formulation, and we will soon see the equivalence of the two concepts.

**Definition 5.2.3** A stochastic integral equation

$$
X_t = X_0 + \int_0^t \sigma(s,X_s)dB_s + \int_0^t b(s,X_s)ds \qquad (5.17)
$$

has a *weak solution* with initial distribution $\mu$, if there exists a filtered space $(\Omega,\mathcal{F},\mathbb{P},\{\mathcal{F}_t\})$, satisfying the usual conditions, and continuous martingales $X$ and $B$, such that

(i) $B$ is an $\mathcal{F}_t$-Brownian motion;
(ii) $X_0$ has law $\mu$;
(iii) $\int_0^t(|\sigma(s,X_s)|^2+|b(s,X_s)|)ds<\infty$, a.s., for all $t$;
(iv) (5.17) holds.

**Definition 5.2.4** A solution of (5.17) is unique in law (or *weakly unique*), if whenever $X_t$ and $X'_t$ are two solutions such that the laws of $X_0$ and $X'_0$ are the same, then the laws of $X$ and $X'$ coincide.

**Example.** The following simple example illustrates the difference between strong and weak solutions. Consider the equation

$$X_t = X_0 + \int_0^t \mathrm{sign}(X_s) dB_s. \tag{5.18}$$

Here we define $\mathrm{sign}(x) = -1$, if $x \leq 0$, and $\mathrm{sign}(x) = +1$, if $x > 0$. Obviously, $[X]_t = \int_0^t dt = t$,, so for any solution, $X_t$, that is a continuous local martingale, Lévy's theorem implies that $X_t$ is a Brownian motion, if it exists. In particular, we have weak uniqueness of the solution. Moreover, we can easily construct a solution: Let $X_t$ be a Brownian motion and set

$$B_t \equiv \int_0^t \mathrm{sign}(X_s) dX_s. \tag{5.19}$$

Then $dB_s = \mathrm{sign}(X_s) dX_s$, and hence

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

so the pair $(X, B)$ yields a weak solution! Note that the Brownian motion is constructed from $X$, not the other way around! On the other hand, there is no path-wise uniqueness: Let, say, $X_0 = 0$. Then, if $X_t$ is a solution, so is $-X_t$. Of course being Brownian motions, they have the same law. Note that the corresponding $B_t$ in the construction above would be the same. Moreover, the Brownian motion of (5.19) is measurable with respect to the filtration generated by $|X_t|$ which is smaller than that of $X_t$; thus, $X_t$ is not adapted to the filtration generated by the Brownian motion. Hence we see that there is indeed not necessarily a solution of this SDE for any $B$, and so this SDE does not have a strong solution.

**Remark 5.2.2** The example (and in particular the last remark) is hiding an interesting fact and concept, that of *local time*. This is the content of the following theorem due to Tanaka:

**Theorem 5.2.3** *Let $X$ be a continuous semi-martingale. Then there exists a continuous increasing adapted process, $\{\ell_t, t \geq 0\}$, called* the local time of $X$ at 0*, such that*

$$|X_t| - |X_0| = \int_0^t \mathrm{sign}(X_s) dX_s + \ell_t. \tag{5.20}$$

*$\ell_t$ grows only when $X$ is zero, i.e.*

$$\int_0^t 1_{X_s \neq 0} d\ell_s = 0. \tag{5.21}$$

*Proof.* The proof uses Itô's formula and an applroximation of the absolute value by $C^\infty$ functions. Chose some non-decreasing smooth funtion $\phi$ that is equal to $-1$ for $x \leq 0$ and equal to $+1$ for $x \geq 1$. Then take $f_n(x)$ suc that $f_n'(x) = \phi(nx)$ with $f_n(0) = 0$. Then Itô's formula gives

$$f_n(X_t) - f_n(X_0) = \int_0^t f_n'(X_s) dX_s + \frac{1}{2} \int_0^t f_n''(X_s) d[X]_s. \tag{5.22}$$

We denote the last term by $C_t^n$. Clearly $C_t^n$ is non-decreasing, and since $f''$ vanishes outside the interval $[0, 1/n]$, we have that

$$\int_0^t \mathbb{1}_{X_s \notin [0,1/n]} dC_s^n = 0. \tag{5.23}$$

It is also important to note that $f_n(x)$ converges to $|x|$ uniformly, and $f_n$ converges to the sign from below.

To prove the convergence of $C_t^n$, we just have to proce the convergence of the stochastic integrals.

Now consider the canonical decomposition of the semi-martingale $X_t = X_0 + M_t + A_t$, where $A_t$ can be assumed of finite variation and $M_t$ bounded; otherwise use localisation. We bound the stochastic integrals with respect to $M_t$ and $A_t$ seperately. The first is controlled be the bound

$$\left\| \int_0^\infty (\text{sign}(X_s) - f_n'(X_s))\, dM_s \right\|_2^2 \leq \mathbb{E} \int_0^\infty (\text{sign}(X_s) - f_n'(X_s))^2\, d[M]_s. \tag{5.24}$$

By the uniform convergence of the integrand to zero, it follows that the right-hand side tends to zero. Then Doob's maximum inequatlity implies that

$$\mathbb{P}\left( \sup_{t \leq \infty} \left| \int_0^\infty (\text{sign}(X_s) - f_n'(X_s))\, dM_s \right| > \varepsilon \right] \leq \varepsilon^{-2} \mathbb{E} \int_0^\infty (\text{sign}(X_s) - f_n'(X_s))^2\, d[M]_s, \tag{5.25}$$

which tends to zero with $n$. Taking possibly subsequences, we get almost sure convergence of the supremum, possibly by choosing subsequences.

The control of the integral with respect to $A_t$ is similar and simpler. Note that the convergence of $f_n'$ is monotone. From here the claimed result follows easily. □

Note that this theorem implies that in the example above, $B_t = |X_t| - \ell_t$, and since $\ell_t$ depends only on $|X|$, the measurability properties claimed above hold.

The connection between weak and strong solutions is clarified in the following theorem due to Yamada and Watanabe. It essentially says that weak existence and path-wise uniqueness imply the existence of a strong solution, and in turn weak uniqueness.

**Theorem 5.2.4** *An SDE is exact if and only if*

*(i) there exists a weak solution, and*
*(ii) solutions are path-wise unique.*

*Then uniqueness in law also holds.*

The proof of this theorem may be found in [14]

## 5.3 Weak solutions and the martingale problem

We will now show a deep and important connection between weak solutions of SDEs and the martingale problem.

The remarkable thing is that these issues can be cooked down again to the study of martingale problems. We do the computations for the one-dimensional case, but clearly everything goes through in the $d$-dimensional case exactly in the same way.

Let us first observe that, using Itô's formula, given that the equation (5.2) has a solution, then it is a solution of a martingale problem.

**Lemma 5.3.5** *Assume that $X$ solves (5.2). Define the family of operator $G_t$ on the space of $C^\infty$-functions $f : \mathbb{R} \to \mathbb{R}$, as*

$$G_t \equiv \frac{1}{2}\sigma^2(t,x)\frac{d^2}{dx^2} + b(t,x)\frac{d}{dx}. \tag{5.26}$$

*Then $X$ is a solution of the martingale problem for $G_t$.*

**Remark 5.3.1** We need here in fact a slight generalisation of the notion of martingale problems in order to include time-inhomogeneous processes. For a family of operators $G_t$ with common domain $\mathcal{D}$, we say that a process $X_t$ is a solution of the martingale problem, if for all $f : S \to \mathbb{R}$ in $\mathcal{D}$,

$$f(X_t) - \int_0^t (G_s f)(X_s) ds \tag{5.27}$$

is a margingale. A simple way of relating this to the usual martingale problem is to consider an process $(t, X_t)$ on the space $\mathbb{R}_+ \times S$. Then the operator $\widetilde{G} = (\partial_t + G_t)$ can be seen as on ordinary generator with domain a subset of $B(\mathbb{R}_+ \times S)$. If $f$ is in this domain, the martingale should be

$$M_t \equiv f(t, X_t) - f(0, X_0) - \int_0^t (\partial_s f(s, X_s) + (G_s f)(s, X_s))ds. \tag{5.28}$$

Restricting the domain of $\widetilde{G}$ to functions of the form $f(t, x) = \gamma(t)g(x)$, this reduces to

$$M_t \equiv g(X_t)\gamma(t) - g(X_0)g(0) - \int_0^t (\partial_s \gamma(s) g(X_s) + (G_s g)(X_s, s)\gamma(s))\, ds. \tag{5.29}$$

We see immediately, by setting $\gamma(t) \equiv 1$, that is $(t, X_t)$ makes (5.29) a martingale, then $X_t$ solves the time dependent martingale problem (5.27). On the other hand it is also easy to see that if $X_t$ makes (5.27) a martingale then $(t, X_t)$ makes (5.29) a martingale. Note that we have seen this already in the special case $\gamma(t) = \exp(\lambda t)$.

*Proof.* For later use we will derive a more general result. Let $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$. We use Itô's formula to express

$$f(t, X_t) - f(0, X_0) = \int_0^t \partial_s f(s, X_s) ds + \int_0^t \partial_x f(s, X_s) dX_s + \frac{1}{2}\int_0^t \partial_x^2 f(s, X_s) d[X]_s. \tag{5.30}$$

Now

$$dX_s = b(s, X_s)ds + \sigma(s, X_s)dB_s.$$

We set

$$M_t \equiv X_t - \int_0^t b(s, X_s)ds$$

and note that this is by (4.25) equal to $\int_0^t \sigma(s, X_s)dB_s$, and hence a martingale. Moreover,

$$[M]_t = \int_0^t \sigma(s, X_s)^2 d[B]_s = \int_0^t \sigma(s, X_s)^2 ds.$$

Hence

$$
\begin{aligned}
f(t, X_t) - f(0, X_0) = & \int_0^t \partial_x f(s, X_s) b(s, X_s) ds \\
& + \int_0^t \partial_s f(s, X_s) ds + \frac{1}{2} \int_0^t \sigma(s, X_s) \partial_x^2 f(s, X_s) ds \\
& + \int_0^t \partial_x f(s, X_s) dM_s,
\end{aligned}
$$

or

$$
f(t, X_t) - f(0, X_0) - \int_0^t [\partial_s f(s, X_s + (Gf)(s, X_s)] ds = - \int_0^t \partial_x f(s, X_s) dM_s, \tag{5.31}
$$

where the right-hand side is a martingale, which means that $X$ solves the martingale problem, as desired. □

This observation becomes really useful through the converse result.

**Theorem 5.3.6** *Assume that $b$ and $\sigma$ are locally bounded as above and assume that in addition $\sigma^{-1}$ is locally bounded. Let $G_t$ be given by (5.26). Assume that $X$ is a continuous solution to the martingale problem for $(G, \delta_{x_0})$, then there exists a Brownian motion, $B$, such that $(X, B)$ is a solution to the stochastic integral equation (5.2).*

*Proof.* We know that for every $f \in C^\infty(\mathbb{R})$,

$$
f(X_t) - f(X_0) - \int_0^t (G_s f)(s, X_s) ds \tag{5.32}
$$

is a continuous martingale. Choosing $f(x) = x$, it follows that

$$
X_t - X_0 - \int_0^t b(s, X_s) ds \equiv M_t \tag{5.33}
$$

is a continuous martingale. Essentially we want to show that this martingale is precisely the stochastic integral term in (5.2). To do this, we need to compute the bracket of $M$. For this we consider naturally (5.32) with $f(x) = x^2$. To simplify the notation, let us assume without loss of generality that $X_0 = 0$. This gives

$$
X_t^2 - 2 \int_0^t X_s b(s, X_s) ds - \int_0^t \sigma^2(s, X_s) ds = \widehat{M}_t, \tag{5.34}
$$

where $\widehat{M}$ is a martingale. Thus

$$M_t^2 = X_t^2 - 2X_t \int_0^t b(s, X_s)ds + \left(\int_0^t b(s, X_s)ds\right)^2 \tag{5.35}$$
$$= 2\int_0^t X_s b(s, X_s)ds + \int_0^t \sigma^2(s, X_s)ds + \widehat{M}_t$$
$$- 2X_t \int_0^t b(s, X_s)ds + \left(\int_0^t b(s, X_s)ds\right)^2 .$$

I claim that

$$2\int_0^t X_s b(s, X_s)ds - 2X_t \int_0^t b(s, X_s)ds + \left(\int_0^t b(s, X_s)ds\right)^2 \tag{5.36}$$

is also a martingale. By partial integration,

$$\int_0^t X_s b(s, X_s)ds = X_t \int_0^t b(s, X_s)ds - \int_0^t \int_0^s b(u, X_u)du dX_s.$$

Thus (5.36) equals

$$-2\int_0^t \int_0^s b(u, X_u)du dX_s + \left(\int_0^t b(s, X_s)ds\right)^2$$
$$= -2\int_0^t \int_0^s b(u, X_u)du dM_s,$$

which is a martingale. Hence

$$M_t^2 - \int_0^t \sigma^2(s, X_s)ds \tag{5.37}$$

is a martingale, so that by definition of the quadratic variation process,

$$\int_0^t \sigma^2(s, X_s)ds = [M]_t.$$

Now set

$$B(t) \equiv \int_0^t \frac{1}{\sigma(s, X_s)} dM_s.$$

Then

$$[B]_t = \int_0^t \frac{1}{\sigma(s, X_s)^2} d[M]_s = t,$$

so by Lévy's theorem, $B(t)$ is Brownian motion, and it follows that $X$ solves (5.2) with this particular realization of Brownian motion. □

We can summarize these findings in the following theorem.

**Theorem 5.3.7** *Let $\mathbb{P}^y$ be a solution of the martingale problem associated to the operator $G$ defined in (5.26) starting in $y$. Then there exists a weak solution of the SDE (5.2) with law $\mathbb{P}^y$. Conversely, if there is a weak solution of (5.2), then there exists a solution of the martingale problem for (5.26). Uniqueness in law holds if and only if the associated martingale problem has a unique solution.*

In other words, solutions of our stochastic integral equation are Markov processes with generator given by the closure of the second order (elliptic) differential operator $G$ given by (5.26). To study their existence and uniqueness, we can use the tools we developed in the theory of Markov processes. Note that we state the theorem without the boundedness assumption on $\sigma^{-1}$ from Theorem 5.3.6, which in fact can be avoided with some extra work.

As a consequence, we sketch two existence and uniqueness results for weak solutions.

**Theorem 5.3.8** *Consider the SDE with time-independent coefficients,*

$$dX_t = b(X_t) + \sigma(X_t)dB_t, \tag{5.38}$$

*in $\mathbb{R}^d$ where the coefficients $b_i$ and $\sigma_{ij}$ are bounded and continuous. Then for any measure $\mu$ such that*

$$\int \|x\|^{2m}\mu(dx) < \infty, \tag{5.39}$$

*for some $m > 0$, there exists a weak solution to (5.38) with initial measure $\mu$.*

*Proof.* We only have to prove that the martingale problem with generator

$$Gf(y) = \sum_i b_i(y)\partial_i f(y) + \frac{1}{2}\sum_{i,j,k}\sigma_{ik}(y)\sigma_{kj}(y)\partial_i\partial_j f(y),$$

for $f \in C_0^2(\mathbb{R}^d)$ has a solution. To do this, we construct an explicit solution for a sequence of operators $G^{(n)}$ that converge to $G$ and deduce from this the existence of the solution of the martingale problem for $G$.

To do this, let $t_j^{(n)} = j2^{-n}$ and set $\phi_n(t) = t_j^{(n)}\mathbb{1}_{t\in[t_j^{(n)},t_{j+1}^{(n)})}$. Then set

$$b^{(n)}(t,y) \equiv b(y(\phi_n(t))), \quad \sigma^{(n)}(t,y) \equiv \sigma(y(\phi_n(t))).$$

Then define the processes $X_t^{(n)}$ by

$$X_0^{(n)} = \xi \tag{5.40}$$
$$X_t^{(n)} = X_{t_j^{(n)}}^{(n)} + b(X_{t_j^{(n)}}^{(n)})(t - t_j^{(n)}) + \sigma(X_{t_j^{(n)}}^{(n)})(B_t - B_{t_j^{(n)}}), t \in (t_j^{(n)}, t_{j+1}^{(n)}].$$

We will denote the laws of the processes $X^{(n)}$ by $P^{(n)}$. One easily verifies that the processes $X^{(n)}$ solves the integral equation

$$X_t^{(n)} = \xi + \int_0^t b^{(n)}(s, X^{(n)})ds + \int_0^t \sigma^{(n)}(s, X^{(n)})dB_s. \tag{5.41}$$

But then $X^{(n)}$ solves the martingale problem for the (time dependent) operator

$$(G_t^{(n)} f)(y) \equiv \sum_i b_i^{(n)}(t,y)\partial_i f(y(t)) + \frac{1}{2}\sum_{i,j,k} \sigma_{ik}^{(n)}(t,y)\sigma_{kj}^{(n)}(t,y)\partial_i\partial_j f(y(t)). \tag{5.42}$$

The first thing to show is that the laws of this family of processes are tight. For this one uses the criterion given by Proposition **??**. The basic ingredient is the following:

$$\mathbb{E}\left\|X_t^{(n)} - X_s^{(n)}\right\|^{2m} \leq C_m(t-s)^m \tag{5.43}$$

for $0 \leq t, s \leq T$, where $C_m$ is uniform in $n$ and depends only on the bound on the coefficients of the sde. Moreover,

$$\mathbb{E}\|X_0^{(n)}\|^{2m} \leq C'_m < \infty \tag{5.44}$$

by assumption. To prove (5.43), we write

$$\mathbb{E}\left\|X_t^{(n)} - X_s^{(n)}\right\|^{2m} \leq \mathbb{E}\left\|\int_s^t b_n(u, X_u^{(n)})du\right\|^{2m} \tag{5.45}$$
$$+\mathbb{E}\left\|\int_s^t \sigma_n(u, X_u^{(n)})dB_u\right\|^{2m} \tag{5.46}$$
$$\leq (t-s)^{2m}\mathbb{E}\sup_{u\in[s,t]}\left\|b_n(u, X_u^{(n)})\right\|^{2m} \tag{5.47}$$
$$+K_m\mathbb{E}\left(\int_s^t \left\|\sigma_n(u, X_u^{(n)})\right\|^2 du\right)^m \tag{5.48}$$
$$\leq C(m)(t-s)^m \tag{5.49}$$

Here we used the inequality (valid for local martingales

$$\mathbb{E}|M_t|^{2m} \leq K_m\mathbb{E}[M]_t^m, \tag{5.50}$$

for the martingale $\int_s^t \sigma(u, X^{(n)})dB_u$. This inequality is a special case of the so-called Burkholder-Davis-Gundy inequality, which we will state and proof below.

Then Prohorov's theorem implies that the sequence is conditionally compact, so that we can at least extract a convergent subsequence. Hence we may assume that $P^{(n)}$ converges weakly to some probability measure $P^*$. We want to show that the process whose law is $P^*$ solves the martingale problem for the operator $G$.

For $f \in C_0^2(\mathbb{R}^d)$, one checks that $G^{(n)} f(y) \to Gf(y)$. Then Lemma (3.4.27) implies that $P^*$ is a solution of the martingale problem and hence a weak solution of the sde exists. □

**Remark 5.3.2** Note that we cheat a little here. Namely, the operators $G^n$ and the form of the approximating integral equations are more general than what we have previously assumed in that the coefficients $b^{(n)}(t, y)$ and $\sigma^{(n)}(t, y)$ depend on the past of the function $y$ and not only on the value of $y$ at time $t$. There is, however, no serious difficulty in generalising the entire theory to that case. The only crucial property that needs to be maintained is that the coefficients remain progressive processes with respect to the filtration $\mathcal{F}_t$.

**Remark 5.3.3** The preceeding theorem can be extended rather easily to the case when $b$ and $\sigma$ are time-dependent, and even to the case when they are bounded, continuous progressive functionals.

**Remark 5.3.4** The boundedness conditions on the coefficients can be replaced by the condition

$$\|b(y)\|^2 + \|\sigma(y)\|^2 \leq K\left(1 + \|y\|^2\right), \tag{5.51}$$

if the bound for the initial condition holds for some $m > 1$. The proof is simiar to the one given above, but requires to bound a moment of the maximum of $X_t^n$ via a Gronwall argument together with the BDG inequalities. I leave this as an exercise.

We now state the Burkholder-Davis-Gundy inequality.

**Lemma 5.3.9** *Let $M$ be a continuous local martingale. Then, for every $m > 0$, there exist universal constants $k_m, K_m$ depending only on $m$, such that, for any stopping time $T$,*

$$k_m \mathbb{E}[M]_T^m \leq \mathbb{E}\left(\sup_{0 \leq s \leq T} |M_s|\right)^{2m} \leq K_m \mathbb{E}[M]_T^m. \tag{5.52}$$

*Proof.* The proof (which is taken from [14]) is based on the following simple lemma, called the "good $\lambda$ inequality".

**Lemma 5.3.10** *Let $X, Y$ be non-negative random variables. Assume that there exists $\beta > 1$, such that for all $\lambda >, \delta > 0$,*

$$\mathbb{P}\left(X > \beta\lambda, Y \leq \delta\lambda\right) \leq \psi(\delta)\mathbb{P}\left(X > \lambda\right), \tag{5.53}$$

*where $\psi(\delta) \downarrow 0$, as $\delta \downarrow 0$. Then for any function positive, increlasing function $F$, such that $\sup_{x>0} \frac{F(\alpha x)}{F(x)} < \infty$, there exists a constant $C$ such that*

$$\mathbb{E}F(X) \leq C\mathbb{E}F(Y). \tag{5.54}$$

*Proof.* The statement is non-trivial only if $\mathbb{E}F(Y) < \infty$. We may also assume that $\mathbb{E}F(X) < \infty$. Now choose $\gamma$ such that for all $x$, $F(x/\beta) \geq \gamma F(x)$. Such a number must exist be hypothesis on $F$. We integrate both sides of (5.53) w.r.t. $F(d\lambda)$ and get, using partial integration,

$$\psi(\delta)\mathbb{E}F(X) \geq \int_0^\infty Fd\lambda)\mathbb{E}\mathbb{1}_{Y/\delta\leq\lambda<X/\beta} \tag{5.55}$$

$$= \mathbb{E}\left(\int_0^{X/\beta} F(d\lambda) - \int_0^{Y/\delta} F(d\lambda)\right)_+$$

$$\geq \mathbb{E}F(X/\beta) - \mathbb{E}F(Y/\delta) \geq \gamma\mathbb{E}F(X) - \mathbb{E}F(Y/\delta).$$

Now we solve this for $\mathbb{E}F(X)$ to get

$$\mathbb{E}F(X) \leq \frac{\mathbb{E}F(Y/\delta)}{\gamma - \psi(\delta)} \tag{5.56}$$

We can choose $\delta$ so small that $\psi(\delta) \leq \gamma/$. Then there exists $\mu$ such that $F(x/\delta) \leq \mu F(x)$, for all $x > 0$. This proves the inequality with $C = 2\mu/\gamma$. □

We have to establish the inquality (5.53) for $X = M_T^* \equiv \sup_{t\leq T} M_t$ and $Y = [M]_T^{1/2}$. Recall that for any continuous martingale $N_t$ starting in zero, for $\tau_x \equiv \inf(t : N_t = x)$, and $a < 0 < b$,

$$\mathbb{P}\left(\tau_a < \tau_b\right) \leq -a/(b-a). \tag{5.57}$$

Now fix $\beta > 1, \lambda > 0$, and $0 < \delta < (\beta - 1)$. Set $\tau \equiv \inf(t : |M_t| > \lambda)$. Define

$$N_t \equiv (M_{t+\tau} - M_\tau)^2 - ([M]_{t+\tau} - [M]_t). \tag{5.58}$$

One easily checks that $N_t$ is a continuous local martingale. Now condsider the event $\{M_T^* \geq \beta\lambda, [M]_T^{1/2} \leq \delta\lambda\}$. Now on this event, we have that

$$\sup_{t\leq T} N_t \geq (\beta-1)^2\lambda^2 - \delta^2\lambda^2, \tag{5.59}$$

and

$$\inf_{t\leq T} N_t \geq -\delta^2\lambda^2. \tag{5.60}$$

This implies that on this event, $N_t$ hits $(\beta-1)^2\lambda^2 - \delta^2\lambda^2$ before $-\delta^2\lambda^2$, and so by (5.57),

$$\mathbb{P}\left(M_T^* \geq \beta\lambda, [M]_T^{1/2} \leq \delta\lambda | \mathcal{F}_\tau\right) \leq \delta^2/(\beta-1)^2. \tag{5.61}$$

From this it follows that

$$\mathbb{P}\left(M_T^* \geq \beta\lambda, [M]_T^{1/2} \leq \delta\lambda\right) \leq \delta^2/(\beta-1)^2 \mathbb{P}\left(\tau < T\right) = d^2/(\beta-1)^2 \mathbb{P}\left(|M_T^*| > 0\lambda\right). \tag{5.62}$$

This proves (5.53) and hence

$$\mathbb{E}F(M_T^*) \leq C\mathbb{E}F([M]_T^{1/2}). \tag{5.63}$$

The converse inequality is obtained by the same procedure but chosing of $Y = M_T^*$ and $X = [M]_T^{1/2}$. □

A uniqueness result is interestingly tied to a Cauchy problem.

**Lemma 5.3.11** *If for every $f \in \mathbb{C}_0^\infty(\mathbb{R}^d)$ the Cauchy problem*

$$\frac{\partial u(t,x)}{\partial t} = (Gu)(t,x), \quad (t,x) \in (0,\infty)\times\mathbb{R}^d \tag{5.64}$$
$$u(0,x) = f(x), \quad x \in \mathbb{R}^d$$

*has a solution in $C([0,\infty)\times\mathbb{R}^d)\cap C^{(1,2)}((0,\infty)\times\mathbb{R}^d)$ that is bounded in any strip $[0,T]\times\mathbb{R}^d$, then any two solutions of the martingale problem for $G$ with the same initial distribution have the same finite dimensional distributions.*

*Proof.* Given the solution $u$ let $g(t,x) \equiv u(T-t,x)$. Then $g$ solves, for $0 \leq t \leq T$,

$$\frac{\partial g(t,x)}{\partial t} + (G_s g)(t,x) = 0, \quad (t,x) \in (0,\infty)\times\mathbb{R}^d \tag{5.65}$$
$$g(T,x) = f(x), \quad x \in \mathbb{R}^d$$

Then it follows from (5.31) that $g(t, X_t)$ is a local martingale for any solution of the martingale problem. Hence

$$\mathbb{E}_x f(X_T) = \mathbb{E}_x g(T, X_T) = \mathbb{E}_x g(0, X_0) = g(0,x), \tag{5.66}$$

is the same for any solution. This implies uniqueness of the one-dimensional distributions. □

Now Theorem 3.4.25 implies immediately the following corollary:

**Corollary 5.3.12** *Under the assumptions of the preceeding lemma, weak uniqueness holds for the SDE corresponding to the generator* $G$.

## 5.4 Weak solutions from Girsanov's theorem

Girsanov's theorem 4.4.11 provides a very efficient and explicite way of constructing weak solutions of certain SDE's.

**Theorem 5.4.13** *Consider the stochastic differential equation*

$$dX_t = b(t, X_t) + dB_t, \quad 0 \leq t \leq T, \tag{5.67}$$

*for fixed* $T$. *Assume that* $b : [0, T] \times \mathbb{R}^d$ *is measurable and satisfies, for some* $K < \infty$,

$$\|b(t, x)\| \leq K(1 + \|x\|). \tag{5.68}$$

*Then for any probability measure* $\mu$ *on* $\mathbb{R}^d$ *there exists a weak solution of (5.67) with initial law* $\mu$.

*Proof.* Let $X$ be a family of Brownian motions starting in $x \in \mathbb{R}$ under laws $P_x$. Then

$$Z_t \equiv \exp\left(\int_0^t b(s, X_s) \cdot dX_s - \frac{1}{2}\int_0^t \|b(s, X_s)\|^2 ds\right) \tag{5.69}$$

is a martingale under $P_x$. Thus Girsanov's theorem says that under the measure $\mathbb{Q}_x$ such that $\frac{d\mathbb{Q}_x}{d\mathbb{P}_x} = Z_T$, the process

$$W_t \equiv X_t - X_0 - \int_0^t b(s, X_s) ds \tag{5.70}$$

for $0 \leq t \leq T$ is a Brownian motion starting in 0. Thus we have a pair $(X_t, W_t)$ such that

$$X_t = X_0 + \int_0^t b(s, X_s) ds + W_t, \tag{5.71}$$

holds for $0 \leq t \leq T$, and $W_t$ is a Brownian motion, under $\mathbb{Q}_x$. This shows that we have a weak solution of (5.67). □

A complementary result also provided criteria for uniqueness in law.

**Theorem 5.4.14** *Assume that we have weak solutions* $(X^{(i)}, W^{(i)})$, $i = 1, 2$, *on filtered spaces* $(\Omega^{(i)}, \mathcal{F}^{(i)}, \mathbb{P}^{(i)}, \mathcal{F}_t^{(i)})$, *of the SDE (5.4.13) with the same initial distribution. If*

$$\mathbb{P}^{(i)}\left[\int_0^T \|b(t, X_t^{(i)}\|^2 dt < \infty\right] = 1, \tag{5.72}$$

*for $i = 1, 2$, then $(X^{(1)}, W^{(1)})$ and $(X^{(2)}, W^{(2)})$ have the same distribution under their respective probability measures $\mathbb{P}^{(i)}$.*

*Proof.* Define stopping times

$$\tau_k^{(i)} \equiv T \wedge \inf\left\{0 \leq t \leq T : \int_0^t \|b(t, X_t^{(i)}\|^2 dt = k\right\}. \tag{5.73}$$

We define the martingales

$$\xi_t^{(k)}(X^{(i)}) \equiv \exp\left(-\int_0^{t\wedge\tau_k^{(i)}} b(s, X_s^{(i)}) dW_s^{(i)} - \frac{1}{2}\int_0^{t\wedge\tau_k^{(i)}} \|b(s, X_s^{(i)})\|^2 ds\right), \tag{5.74}$$

and the corresponding transformed measures $\widetilde{\mathbb{P}}_k^{(i)}$. Then by Girsanov's theorem,

$$X_{t\wedge\tau_k^{(i)}}^{(i)} \equiv X_0^{(i)} + \int_0^{\tau_k^{(i)}} b(s, X_s^{(i)}) ds + W_{t\wedge\tau_k^{(i)}}^{(i)} \tag{5.75}$$

is a Brownian motion with unital distribution $\mu$, stopped at $\tau_k^{(i)}$. In particular, these processes have the same law for $i = 1, 2$. Now the $W^{(i)}$ and the stopping times $\tau_k^{(i)}$ can be expressed in terms of these processes, and probabilities of events of the form

$$\{((X_{t_1}^{(i)}, W_{t_1}^{(i)}), \dots (X_{t_n}^{(i)}, W_{t_n}^{(i)})) \in \Gamma, \tau_k^{(i)} = t_n\},$$

for any collections $t_1 < t_2 < \cdots < t_n$ thus have the same probabilities. Passing to the limit $k \uparrow \infty$ using that due to our assumption, $\mathbb{P}^{(i)}[\tau_k^{(i)} = T] \to 1$ we get uniqueness in law for the entire time interval $[0, T]$. □

## 5.5 Large deviations

In this section we will give a short glimpse in what is know as the *theory of large deviations large deviations* in the context of simple diffusions. I will emphasize the use of Girsanov's theorem and skid over numerous other interesting issues. There are many nice books on large deviation theory, in particular [**?**, **?**, **?**].

We begin with a discussion of *Schilder's theorem* for Brownian motion.

A we know very well, a Brownian motion $B_t$ starting at the origin will, at time $t$, typically be found at a distance not greater than $\sqrt{t}$ from the origin, in particular, $B_t/t$ converges to zero a.s. We will be interested in computing the probabilities that the BM follows an exceptional path that lives on the sale $t$. To formalize this idea, we fix a time scale $T$

(which we might also call $1/\varepsilon$), and a smooth path $\gamma : [0,1] \to \mathbb{R}^d$. We want to estimate

$$\mathbb{P}\left[\sup_{0\leq s\leq 1} \|T^{-1}B_{sT} - \gamma(s)\| \leq \varepsilon\right]. \tag{5.76}$$

It will be convenient to adopt the notation $\|f\|_\infty \equiv \sup_{0\leq s\leq 1} \|f(s)\|$. We will first prove a lower bound on the probabilities of the form (5.76).

**Lemma 5.5.15** *Let $B$ be Brownian motion, set $B^T_s \equiv T^{-1}B_{Ts}$, and let $\gamma$ be a smooth path in $\mathbb{R}^d$ starting in the origin. Then*

$$\lim_{\varepsilon\downarrow 0}\lim_{T\uparrow\infty} T^{-1}\ln \mathbb{P}\left[\|B^T - \gamma\|_\infty \leq \varepsilon\right] \geq -I(\gamma) \equiv -\frac{1}{2}\int_0^1 \|\dot\gamma(s)\|^2 ds. \tag{5.77}$$

*Proof.* For notational simplicity we consider the case $d = 1$ only. Note that $B^T_s = T^{-1}B_{sT}$ has the same distribution as $T^{-1/2}B_s$. Thus we must estimate the probabilities

$$\mathbb{P}\left[\sup_{t\leq 1} \|B_t - \sqrt{T}\gamma(t)\| \leq \sqrt{T}\varepsilon\right]. \tag{5.78}$$

To do this, we observe that by Girsanov's theorem, the process

$$\widehat{B}_t \equiv B_t - \sqrt{T}\gamma(t) \tag{5.79}$$

is a Brownian motion under the measure $\mathbb{Q}$ defined through

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\sqrt{T}\int_0^t \dot\gamma(s)dB_s - \frac{T}{2}\int_0^t \|\dot\gamma(s)\|^2 ds\right). \tag{5.80}$$

Hence

$$\begin{aligned}
&\mathbb{P}\left[\|B - \sqrt{T}\gamma\|_\infty \leq \sqrt{T}\varepsilon\right] \qquad (5.81)\\
&= \mathbb{P}\left[\|\widetilde{B}\|_\infty \leq \sqrt{T}\varepsilon\right]\\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)dB_s + \frac{T}{2}\int_0^1 \|\dot\gamma(s)\|^2 ds}\mathbb{1}_{\|\widetilde{B}\|_\infty \leq \sqrt{T}\varepsilon}\right]\\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)d\widetilde{B}_s - \frac{T}{2}\int_0^1 \|\dot\gamma(s)\|^2 ds}\mathbb{1}_{\|\widetilde{B}\|_\infty \leq \sqrt{T}\varepsilon}\right]\\
&= e^{-\frac{T}{2}\int_0^1 \|\dot\gamma\|^2(s)ds}\mathbb{Q}\left[\|\widetilde{B}\|_\infty \leq \sqrt{T}\varepsilon\right]\mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)d\widetilde{B}_s}\middle|\|\widetilde{B}\|_\infty \leq \sqrt{T}\varepsilon\right]\\
&= e^{-\frac{T}{2}\int_0^1 \|\dot\gamma(s)\|^2 ds}\mathbb{P}\left[\|B\|_\infty \leq \sqrt{T}\varepsilon\right]\mathbb{E}_{\mathbb{P}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)dB_s}\middle|\|B\|_\infty \leq \sqrt{T}\varepsilon\right].
\end{aligned}$$

Now we may use Jensen's inequality to get that

$$\mathbb{E}_{\mathbb{P}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)dB_s}\Big|\|B\|_\infty \leq \sqrt{T}\varepsilon\right] \tag{5.82}$$
$$\geq \exp\left(-\sqrt{T}\mathbb{E}_{\mathbb{P}}\left[\int_0^1 \dot\gamma(s)dB_s\Big|\|B\|_\infty \leq \sqrt{T}\varepsilon\right]\right) = 1.$$

On the other hand, it is easy to see, using e.g. the maximum inequality, that, for any $\varepsilon > 0$,

$$\lim_{T\uparrow\infty} \mathbb{P}\left[\|B\|_\infty \leq \sqrt{T}\varepsilon\right] = 1. \tag{5.83}$$

Hence,

$$\liminf_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[\|B-\sqrt{T}\gamma\|_\infty \leq \sqrt{T}\varepsilon\right] \geq -\frac{1}{2}\int_0^t \|\dot\gamma(s)\|^2 ds, \tag{5.84}$$

which is the desired lower bound. □

To prove a corresponding upper bound, we proceed as follows. Fix $n\in\mathbb{N}$ and set $t_k = k/n$, $k = 0,\dots n$. Set $\alpha \equiv T/n$. Let $L$ be the linear interpolation of $B^T_s$ such that for all $t_k$, $B^T_{t_k} = L_{t_k}$. Then

$$\begin{aligned}\mathbb{P}\left[\|B^T - L\|_\infty > \delta\right] &\leq \sum_{k=1}^n \mathbb{P}\left[\max_{t_{k-1}\leq t\leq t_k}\|B^T_t - L_t\| > \delta\right]\\ &\leq n\mathbb{P}\left[\max_{0\leq t\leq\alpha}\|B^T_t - L_t\| > \delta\right]\\ &= n\mathbb{P}\left[\max_{0\leq t\leq\alpha}\|B_t - \frac{t}{\alpha}B_\alpha\| > \delta\sqrt{T}\right]\\ &\leq n\mathbb{P}\left[\max_{0\leq t\leq\alpha}\|B_t - \frac{t}{\alpha}B_\alpha\| > \delta\sqrt{T}\right]\\ &\leq n\mathbb{P}\left[\max_{0\leq t\leq\alpha}\|B_t\| > \delta\sqrt{T}/2\right],\end{aligned}$$

where we used that $\max_{0\leq t\leq\alpha}\|B_t - \frac{t}{\alpha}B_\alpha\| > x$ implies that $\max_{0\leq t\leq\alpha}\|B_t\| > x/2$. The last probability can be estimated using the following exponential inequality (for one-dimensional Brownian motion)

$$\mathbb{P}[\sup_{0\leq s\leq t}|B_s| > xt] \leq 2\exp\left(-\frac{x^2 t}{2}\right) \tag{5.85}$$

which is obtained easily using that $Z_t \equiv \exp\left(\alpha B_t - \frac{1}{2}\alpha^2 t\right)$ is a martingale and applying Doob's submartingale inequality (see the proof of the Law of the iterated logarithm in [**?**]).

This gives us

$$\mathbb{P}\left[\max_{0\leq t\leq \alpha}\|B_t\| > \delta\sqrt{T}/2\right] \leq d\mathbb{P}\left[\max_{0\leq t\leq \alpha}|B_t| > \delta\sqrt{T}/2\sqrt{d}\right] \quad (5.86)$$
$$\leq = 2de^{-\frac{\delta^2 n^2 T}{8d}}$$

and so

$$\mathbb{P}\left[\|B^T - L\|_\infty > \delta\right] \leq n2e^{-\frac{\delta^2 n^2 T}{8d}} \quad (5.87)$$

which can be made as small as desired by choosing $n$ large enough.

The simplest way to proceed now is to estimate the probability that the value of the *action functional*, $I$, on $L$, has an exponential tail with rate $T$, i.e. that, for $n$ large enough,

$$\limsup_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[I(L) \geq \lambda\right] \leq \lambda. \quad (5.88)$$

This is proven easily using the exponential Chebyshev inequality, since

$$I(L) = \frac{n}{2}\sum_{k=1}^{n}\left\|B^T_{t_{k+1}} - B^T_{t_k}\right\|^2 = \frac{1}{2T}\sum_{i=1}^{dn}\eta_i^2$$

where $\eta_i$ are iid standard normal random variables. But

$$\mathbb{E}e^{\rho\eta_i^2} \leq C_\lambda \leq \infty,$$

for all $\rho < 1$, and so

$$\mathbb{P}\left[\frac{1}{2T}\sum_{i=1}^{dn}\eta_i^2 > \lambda\right] \leq e^{-\rho\lambda T}\mathbb{E}e^{\rho\sum_{i=1}^{nd}\eta_i^2/2} \quad (5.89)$$
$$\leq e^{-\rho\lambda T}C_\rho^{nd}$$

for all $\rho < 1$, and so (5.88) follows, for any $n$.

We can deduce from the two estimates the following version of the upper bound:

**Proposition 5.5.16** *Let $K_\lambda \equiv \{\phi : I(\phi) \leq \lambda\}$. Then*

$$\limsup_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[\mathrm{dist}(\mathrm{B^T}, \mathrm{K}_\lambda) \geq \delta\right] \leq -\lambda. \quad (5.90)$$

Clearly the meaning of this proposition is that the probability to find a Brownian that is not near a path whose action is less than $\lambda$ has probability less than $\exp(-\lambda T)$.

The two bounds, together with the fact that the levels sets $K_\lambda$ (of $I$ are compact (a fact we will not prove), imply the usual formulation of a *large deviation principle*:

**Theorem 5.5.17** *For any Borel set $A \subset W$,*

$$-\inf_{\gamma\in \operatorname{int} A} I(\phi) \leq \liminf_{T\uparrow\infty} T^{-1} \ln \mathbb{P}\left[B^T \in A\right] \\ \leq \limsup_{T\uparrow\infty} T^{-1} \ln \mathbb{P}\left[B^T \in A\right] \leq -\inf_{\gamma\in \bar{A}} I(\phi), \tag{5.91}$$

*where* $\operatorname{int} A$ *and* $\bar{A}$ *denote the interior respectively closure of* $A$*.*

The next step will be to pass to an analogous result for the solution of the SDE (5.67) with a scaled down Brownian term. i.e. we want to consider the equation

$$X_t = T^{-1/2} B_t + \int_0^t b(X_s) ds. \tag{5.92}$$

(for notational simplicity we take zero initial conditions). The easiest (although somewhat particular) way to do this is to construct the map $F : W \to W$, as

$$F(\gamma) = f, \tag{5.93}$$

where $f$ is the solution of the integral equation

$$f(t) = \int_0^t b(f(s))ds + \gamma(t). \tag{5.94}$$

We may use Gronwall's lemma to show that this mapping is continuous. Then $X = F(B^T)$, and

$$\mathbb{P}[X \in A] = \mathbb{P}[B^T \in F^{-1}(A)]. \tag{5.95}$$

Hence, since the continuous map maps open/reps. closed sets in open/resp. closed sets, we can use LDP for Brownian motion to see that

$$\mathbb{P}[X \in A] \leq \sup_{\gamma\in F^{-1}(\bar{A})} I(\gamma) = \sup_{F(\gamma)\in \bar{A}} I(\gamma) = \sup_{\gamma\in\bar{A}} I(F^{-1}(\gamma)), \tag{5.96}$$

and similarly for the lower bound. Hence the process $X^T$ satisfies a large deviation principle with rate function $\widetilde{I}(\gamma) = I(F^{-1}(\gamma))$, and since

$$F^{-1}(\gamma)(t) = \gamma(t) - \int_0^t b(\gamma_s) ds,$$

$$\widetilde{I}(\gamma) = \frac{1}{2}\int_0^1 \|\dot{\gamma}_s - b(\gamma_s)\|^2 \, ds \tag{5.97}$$

This transportation of a rate function from one family of processes to their image is called sometimes a *contraction principle*.

**Properties of action functionals** . The rate function $I(\gamma)$ has the form of a classical action functional in Newtonian mechanics, i.e. it is of the form

$$I(\gamma) = \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s), s) ds, \tag{5.98}$$

where the Lagrangian, $\mathcal{L}$, takes the special form

$$\mathcal{L}(\gamma(s), \dot{\gamma}(s), s) = \|\dot{\gamma}(s) - b(\gamma(s), s)\|_2^2. \tag{5.99}$$

The principle of least action in classical mechanics then states that the systems follows a the trajectory of minimal action subject to boundary conditions. This leads to the Euler-Lagrange equations,

$$\frac{d}{dt}\frac{\partial}{\partial \dot{\gamma}}\mathcal{L}(\gamma, \dot{\gamma}, s) = \frac{\partial}{\partial \gamma}\mathcal{L}(\gamma, \dot{\gamma}, s). \tag{5.100}$$

In our case, these take the form

$$\frac{d^2}{dt^2}\gamma(t) = \frac{\partial}{\partial t} b(\gamma(t), t) + b(\gamma(t), t)\frac{\partial}{\partial \gamma(t)} b(\gamma(t), t). \tag{5.101}$$

One can readily identify a special class of solution of this second order equation, namely solutions of the first order equations

$$\dot{\gamma}(t) = b(\gamma(t), t), \tag{5.102}$$

which have the property that they yield absolute minima of the action, $I(\gamma) = 0$. Of course, being first order equations, they admit only one boundary or initial condition.

Typical questions one will ask in the probabilistic context are: what is the probability of a solution connecting $a$ and $b$ in time $t$. The large deviation principle yields the ansert

$$\mathbb{P}\left[|X_0 - a| \leq d, |X_t - b| \leq \delta\right] \sim \exp\left(-\varepsilon^{-1} \inf_{\gamma:\gamma(0)=a,\gamma(t)=b} I(\gamma)\right), \tag{5.103}$$

which leads us to solve (5.101) subject to boundary conditions $\gamma(0) = a, \gamma(t) = b$. In general this will not solve (5.102), and thus the optimal solution will have positive action, and the event under consideration will have an exponentially small probability. On the other hand, under certain conditions one may find a zero-action solution if one does not fix the time of arrival at the endpoint:

$$\begin{aligned} &\mathbb{P}\left[|X_0 - a| \leq d, |X_t - b| \leq \delta, \text{for some}\, t < \infty\right] \\ &\sim \exp\left(-\varepsilon^{-1} \inf_{\gamma:\gamma(0)=a,\gamma(t)=b,\text{for some}\, t\infty} I(\gamma)\right). \end{aligned} \tag{5.104}$$

Clearly the infimum will be zero, if the solution of the initial value problem (5.102) with $\gamma(0) = a$ has the property that for some $t < \infty$, $\gamma(t) = b$, or if $\gamma(t) \to b$, as $t \uparrow \infty$.

**Exercise.** Consider the case of one dimension with $b(x) = -x$. Compute the minimal action for the problem (5.103) and characterize the situations for which a minimal action solution exists.

A particularly interesting question is related to the so called *exit problem*. Assume that we we consider an event as in (5.104) that admits an zero-action path $\gamma$, such that $\gamma(0) = a, \gamma(T) = (b)$. Define the time reversed path $\hat\gamma(t) \equiv \gamma(T-t)$. Clearly $\frac{d}{dt}\hat\gamma(t) = -\dot\gamma(T-t)$. Hence a simple calculation shows that

$$I(\hat\gamma) - I(\gamma) = 2\int_0^T b(\gamma(s)) \cdot \dot\gamma(s) ds = \int_\gamma b(\gamma) d\gamma. \tag{5.105}$$

Let us now specialize to the case when the vector field $b$ is the gradient of a potentia, $b(x) = \nabla F(x)$. Then

$$\int_\gamma b(\gamma) d\gamma = F(\gamma(T)) - F(\gamma(0)) = F(b) - F(a). \tag{5.106}$$

Hence

$$I(\hat\gamma) = I(\gamma) + F(b) - F(a), \tag{5.107}$$

If $I(\gamma) = 0$, then $I(\hat\gamma) = F(b) - F(a)$, and this is the miminal possible value for any curve going from $b$ to $a$. This shows the remarkable fact that the most likely path going uphill against a potential is the time-reversal of the solution of the gradient flow. Estimates of this type are the basis of the so-called Wentzell-Freidlin theory [**?**].

## 5.6 SDE's from conditioning: Doob's $h$-transform

With Girsanov's theorem we have seen that drift can be produced through a change of measure. Another important way in which drift can arise is conditioning. We have seen this already in the case of discrete time Markov chains. Again we will see that the martingale formulation plays a useful rôle.

As in the discrete case, the key result is the following.

**Theorem 5.6.18** *Let $X$ be a Markov process, i.e. a solution of the martingale problem for an operator $G$ and let $h$ be a strictly positive harmonic function. Define the measure $\mathbb{P}^h$ s.t. for any $\mathcal{F}_t$ measurable random variable,*

$$\mathbb{E}_x^h[Y] = \frac{1}{h(x)}\mathbb{E}_x[h(X_t)Y]. \tag{5.108}$$

*Then $\mathbb{P}^h$ is the law of a solution of the martingale problem for the operator $G^h$ defined by*

$$(G^h f)(x) \equiv \frac{1}{h(x)}(Lhf)(x). \tag{5.109}$$

As an important example, let us consider the case of Brownian motion in a domain $D \subset \mathbb{R}^d$, killed in the boundary of $D$. We will assume that $D$ is a harmonic function in $D$ and let $\tau_D$ the first exit time of $D$. Then

$$G^h = \frac{1}{2}\Delta + \frac{\nabla h}{h} \cdot \nabla,$$

and hence under the law $\mathbb{P}^h$, the Brownian motion becomes the solution of the SDE

$$dX_t = \frac{\nabla h(X_t)}{h(X_t)}dt + dB_t. \tag{5.110}$$

On the other hand, we have seen that, if $h$ is the probability of some event, e.g.

$$H(x) = \mathbb{P}_x[X_{\tau_D} \in A],$$

for some $A \in \partial D$, then

$$\mathbb{P}^h[\cdot] = \mathbb{P}[\cdot | X_{\tau_D} \in A] \tag{5.111}$$

This means that the Brownian motion conditioned to exit $D$ in a given place can be represented as a solution of an SDE with a particular drift. For instance, let $d = 1$, and let $D = (0, R)$. Consider the Brownian motion conditioned to leave $D$ at $R$. It is elementary to see that

$$\mathbb{P}_x[X_{\tau_D} = R] = x/R.$$

Thus the conditioned Brownian motion solves

$$dX_t = \frac{1}{X_t}dt + dB_t. \tag{5.112}$$

Note that we can take $R \uparrow \infty$ without changing the SDE. Thus, the solution of (5.112) is Brownian motion conditioned to never return to the origin. This is understandable, as the strength of the drift away form zero goes to infinity (quickly) near 0. Still, it is quite a remarkable fact that conditioning can be exactly reproduced by the application of the right drift.

Note that the process defined by (5.112) has also another interpretation. Let $W = (W_1, \ldots, W_d)$ be $d$-dimensional Brownian motion. Set $R_t = \|W(t)\|_2$. Then $R_t$ is called the *Bessel process* with dimension $d$. It turns out that this process is also the (weak) solution of a stochastic differential equation, namely:

**Proposition 5.6.19** *The Bessel process in dimension $d$ is a weak solution of*

$$dR_t = \frac{d-1}{2R_t} + dB_t. \tag{5.113}$$

*Proof.* Let us first construct the Brownian motion $B_t$ from the $d$-dimensional Brownian motions $W$ as follows. Set

$$B_t^{(i)} \equiv \int_0^t \frac{W_i(s)}{R_s} dW_i(s)$$

and

$$B_t \equiv \sum_{i=1}^d B_t^{(i)}.$$

The processes in $B_t^{(i)}$ are continuous square integrable martingales since

$$\mathbb{E}\left(\int_0^t \frac{W_i(s)}{R_s} dW_i(s)\right) = \mathbb{E}\int_0^t \left(\frac{W_i(s)}{R_s}\right)^2 ds \leq t;$$

Moreover the

$$[B]_t = \sum_{i,j} [B^{(i)}, B^{(j)}]_t = \sum_i \int_0^t \left(\frac{W_i(s)}{R_s}\right)^2 ds = t,$$

so by Lévy's theorem, $B$ is Brownian motion. Thus we can write (5.113) as

$$dR_t = \sum_i \frac{1}{R_t} dW_i(t) + \frac{1}{2}\frac{d-1}{R_t} dt.$$

But this is precisely the result of applying Itô's formula to the function $f(W) = \|W\|_2$. Note that this derivation is slightly sloppy, since the function $f$ is not differentiable at zero, but the result is correct anyway (for a fully rigorous proof see e.g. [11], Chapter 3.3). □

In particular, we see that the one-dimensional Brownian motion conditioned to stay strictly positive for all positive times is the 3-dimensional Bessel process. This shows in particular that in dimension 3 (and trivially higher), Brownian motion never returns to the origin. Looking at the SDE describing the Bessel process, one might guess that the value of $d$, as soon as $d > 1$, should not be so important for this property, since there is always a divergent drift away from 0. We will now show that this is indeed the case.

**Proposition 5.6.20** *Let $R_t$ be the solution of the SDE (5.113) with $d \geq 2$ and initial condition $R_0 = r \geq 0$. Then*

$$\mathbb{P}\left[\forall t > 0 : R_t > 0\right] = 1. \tag{5.114}$$

*Proof.* Let first $r > 0$. Let

$$\tau_k \equiv \inf\left\{t \geq 0 : R_t = k^{-k}\right\},$$

$$\sigma_k \equiv \inf\left\{t \geq 0 : R_t = k\right\}$$

and $T_k \equiv \tau_k \wedge \sigma_k \wedge n$. Now use Itô's formula for the function $h(R_{T_k})$, where $h(x) = \frac{1}{1-\alpha} x^{-\alpha+1}$, if $(d-1)/2 = \alpha \neq 1$, and $h(x) = \ln x$, if $d = 2$. The point is that $h$ is a harmonic function w.r.t. the operator $G = \frac{d^2}{dx^2} + \alpha \frac{1}{x} \frac{d}{dx}$, and hence $h(R_t)$ is a martingale. Moreover, since $T_k$ is a bounded stopping time, it follows that

$$\mathbb{E}_r\left[h\left(R_{T_k}\right)\right] = h(r). \tag{5.115}$$

Finally,

$$\mathbb{E}_r\left[h\left(R_{T_k}\right)\right] = h(k)\mathbb{P}_r[T_k = \sigma_k] + h(k^{-k})\mathbb{P}_r[T_k = \tau_k] + h(B_n)\mathbb{P}_r[T_k = n]. \tag{5.116}$$

Hence

$$\mathbb{P}_r[T_k = \tau_k] \leq \frac{h(r)}{h(k^{-k})} \leq \begin{cases} k^{-(\alpha-1)k} r^{-\alpha+1}, & \text{if } d \neq 2, \\ \frac{\ln r}{k \ln k}, & \text{if } d = 2. \end{cases} \tag{5.117}$$

Now all what is left to show is that $\mathbb{P}[n < \tau_k \wedge \sigma_k] \downarrow 0$, as $n \uparrow \infty$. But this is obvious from the fact that $R_t \geq r + B_t$, and $\mathbb{P}_0[B_t \leq n]$ tends to zero as $n \uparrow \infty$. Hence,

$$\lim_{n \uparrow \infty} \mathbb{P}_r[T_k = \tau_k] = \mathbb{P}_r[\tau_k < \sigma_k]$$

which in turn tends to zero with $k$. Now set $\tau \equiv \inf\{t > 0 : B_t = 0\}$. For every $k$, $\tau_k < \tau$, so that, again since $\sigma_k \uparrow \infty$, a.s.,

$$\mathbb{P}\left[\tau < \infty\right] \leq \lim_{k \uparrow \infty} \mathbb{P}_r\left[\tau < \sigma_k\right] \leq \lim_{k \uparrow \infty} \mathbb{P}_r\left[\tau_k < \sigma_k\right] = 0. \tag{5.118}$$

This proves the case $r > 0$. For $r = 0$, just use that, by the strong Markov property, for any $\varepsilon > 0$,

$$\mathbb{P}_0\left[R_t > 0, \forall \varepsilon < t < \infty\right] = \mathbb{E}_0 \mathbb{P}_{B_\varepsilon}[[R_t > 0, \forall 0 < t < \infty] = 1, \qquad (5.119)$$

since $\mathbb{P}_0[R_\varepsilon > 0] = 1$. Finally let $\varepsilon \downarrow 0$ to complete the proof. □

**Remark 5.6.1** The method used above is important beyond this example. It has a useful generalization in that one need not chose for $h$ a harmonic function. In fact all goes through if $h$ is chosen to be super-harmonics. In many situations it may be difficult to find a harmonic function, whereas one may well be able to to find a useful super-harmonic function.

# Bibliography

[1] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[2] A. Bovier. Stochastic processes 1. Discrete time. Lecture notes, Berlin Mathematical School, 2006.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[4] Frank den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[5] Nelson Dunford and Jacob T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[6] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[7] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.

[8] K. Ito. *Stochastic processes*. Lecture Notes Series, No. 16. Matematisk Institut, Aarhus Universitet, Aarhus, 1969.

[9] Kiyosi Itô and Henry P. McKeane. *Diffusion processes and their sample paths*. Springer, New York, 1965.

[10] J. Jacod. *Calcul stochastique et problèmes de martingales*. Springer, New York, 1979.

[11] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Graduate Texts in Mathematics. Springer, New York, 1988.

[12] J.F. LeGall. Mouvement brownien et calcul stochastique. Lecture notes, Univeristé Paris Sud, 2008.

[13] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.

[14] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[15] W. Whitt. *Stochastic-process limits.* Springer-Verlag, New York, 2002.

# Index