Anton Bovier

Markov Processes

Lecture Notes Summer 2012, Bonn

November 24, 2012

universitätbonn iam

Contents

Chapter 1
Markov processes in discrete time

Markov processes are among the most important stochastic processes that are used to model real live phenomena that involve disorder. This is because the construction of these processes is very much adapted to our thinking about such processes. Moreover, Markov processes can be very easily implemented in numerical algorithms. This allows to numerically simulate even very complicated systems. We will always imagine a Markov process as a "particle" moving around in state space; mind, however, that these "particles" can represent all kinds of very complicated things, once we allow the state space to be sufficiently general.

Markov processes can be classified according to the properties of the nature of time and the properties of their state space. Roughly, we have the following categories:

(i) discrete time, finite state space
(ii) discrete time, countable state space
(iii) discrete time, general state space
(iv) continuous time, countable state space
(v) continuous time, general state space

The case (i) is elementary and can be studied with the help of elementary linear algebra. Case (ii) is already much more interesting, and brings new concepts such as *recurrence* and *transience*. Case (iii) is really not all that more complicated, although there are new concepts with regard to all ergodicity problems. Case (iv) is not all that different from case (ii), and the construction basically start from a discrete time Markov process where each unit of time is replaced by an exponentially distributed random time, whose parameter depends on the position in space. Fundamentally new issues here can arise if these parameters are unbounded from above or not bounded away from zero. Case (v) is really new, and poses challenging new problems that require some serious tools from functional analysis. A key new problem here is how to describe such a process in simple terms. You already know some important examples from stochastic analysis: Brownian motion, Lévy processes, and processes that are built from these: strong solutions of stochastic differential equations. However, this is not all there is, and in this lecture we will develop a more general theory of continuous time Markov processes.

As a warm-up, we recall in this first chapter the theory of Markov processes with discrete time with a slightly different twist.

1.1 Markov processes with stationary transition probabilities

In the following we denote by S the state space which we assume to be a Polish space. $\mathscr{B}$ denotes the Borel-σ-algebra on S.

The main building block for a Markov process is the so-called *transition kernel*.

Definition 1.1. A (one step) transition kernel for a discrete time Markov process with state space S is a map, $\mathscr{P} : \mathbb{N}_0 \times S \times \mathscr{B} \to [0,1]$, with the following properties:

(i) For each $t \in \mathbb{N}_0$ and $x \in S$, $\mathscr{P}_t(x,\cdot)$ is a probability measure on $(S,\mathscr{B})$.
(ii)For each $A \in \mathscr{B}$, and $t \in \mathbb{N}_0$, $\mathscr{P}_t(\cdot,A)$ is a $\mathscr{B}$-measurable function on S.

Definition 1.2. A stochastic process X with state space S and index set $\mathbb{N}_0$ is a discrete time Markov process with transition kernel $\mathscr{P}$ if, for all $A \in \mathscr{B}$, $t \in \mathbb{N}$,

$$\mathbb{P}(X_t \in A|\mathscr{F}_{t-1})(\omega) = \mathscr{P}_{t-1}(X_{t-1}(\omega),A), \mathbb{P}-\text{a.s.}. \tag{1.1.1}$$

Here $\{F_t\}_{t\in\mathbb{N}_0}$ denotes the σ-algebra generated by the random variables $X_0,\dots,X_t$.

This requirement fixes the law $\mathbb{P}$ up to one more probability measure on $(S,\mathscr{B})$, the so-called *initial distribution*, P_0.

Theorem 1.3. *Let $(S,\mathscr{B})$ be a Polish space and let $\mathscr{P}$ be a transition kernel and P_0 a probability measure on $(S,\mathscr{B})$. Then there exists a unique stochastic process satisfying (1.1.1) and $\mathbb{P}(X_0 \in A) = P_0(A)$, for all A.*

In general, we call a stochastic process whose index set supports the action of a group (or semi-group) *stationary* (with respect to the action of this (semi) group, if all finite dimensional distributions are invariant under the simultaneous shift of all time-indices. Specifically, if our index sets, I, are $\mathbb{R}_+$ or $\mathbb{Z}$, resp. $\mathbb{N}$, then a stochastic process is stationary if for all $\ell \in \mathbb{N}$, $s_1,\dots,s_\ell \in I$, all $A_1 \dots, A_\ell \in \mathscr{B}$, and all $t \in I$,

$$\mathbb{P}\left[X_{s_1} \in A_1,\dots,X_{s_\ell} \in A_\ell\right] = \mathbb{P}\left[X_{s_1+t} \in A_1,\dots,X_{s_\ell+t} \in A_\ell\right]. \tag{1.1.2}$$

We can express this also as follows: Define the shift θ, for any $t \in I$, as $(X \circ \theta_t)_s \equiv X_{t+s}$. Then X is stationary, if and only if, for all $t \in I$, the processes X and $X \circ \theta_t$ have the same finite dimensional distributions.

In the case of Markov processes, a necessary (but not sufficient) condition for stationarity is the stationarity of the transitions kernels.

Definition 1.4. A Markov process with discrete time $\mathbb{N}_0$ and state space S is said to have *stationary transition probabilities (kernels)*, if its one step transition kernel $\mathscr{P}_t$ is independent of t, i.e., if there is a probability kernel $P(x,A)$

$$\mathscr{P}_t(x,A) = P(x,A), \tag{1.1.3}$$

for all $t \in \mathbb{N}$, $x \in S$, and $A \in \mathscr{B}$.

Remark 1.5. With the notation $\mathscr{P}_{t,s}$ for the transitions kernel from time s to time t, i.e.

$$\mathbb{P}(X_t \in A|\mathscr{F}_s) = \mathscr{P}_{t,s}(A,X_s),$$

we could alternatively state that a Markov process has *stationary transition probabilities (kernels)*, if there exists a family of transition kernels $P_t(x,A)$, s.t.

$$\mathscr{P}_{s,t}(x,A) = P_{t-s}(x,A), \tag{1.1.4}$$

for all $s < t \in \mathbb{N}$, $x \in S$, and $A \in \mathscr{B}$. Note that there is a potential conflict of notation between $\mathscr{P}_t$ and P_t which should not be confused.

A key concept for Markov processes with stationary transition kernels is the notion of an *invariant* distribution.

Definition 1.6. Let P be the transition kernel of a Markov process with stationary transition kernels. Then a probability measure, π, on $(S,\mathscr{B})$ is called an invariant (probability) distribution, if

$$\int \pi(dx)P(x,A) = \pi(A), \tag{1.1.5}$$

for all $A \in \mathscr{B}$. More generally, a positive, σ-finite measure, π, satisfying (1.1.5), is called an *invariant measure*.

Lemma 1.7. *A Markov process with stationary probability kernels and initial distribution $P_0 = \pi$ is a stationary stochastic process, if and only if π is an invariant probability distribution.*

Proof. Exercise. $\square$

In the case when the state space, S, is finite, we have seen that there is always at least one invariant measure, which then can be chosen to be a probability measure. In the case of general state spaces, while there still will always be an invariant measure (through a generalisation of the Perron-Frobenius theorem to the operator setting), there appears a new issue, namely whether there is an invariant measure that is finite, viz. whether there exists a invariant probability distribution.

1.2 The strong Markov property

The setting of Markov processes is very much suitable for the application of the notions of stopping times. Recall that for a closed set, D, we set

$$\tau_D \equiv \inf(t > 0 : X_t \in D). \tag{1.2.1}$$

In fact, one of the very important properties of Markov processes is the fact that we can split expectations between past and future also at stopping times.

Theorem 1.8. *Let X be a Markov process with stationary transition kernels. Let $\mathscr{F}_n$ be a filtration such that X is adapted, and let T be a stopping time. Let F and G be $\mathscr{F}$-measurable functions, and let F in addition be measurable with respect to the pre-T-σ-algebra $\mathscr{F}_T$. Then*

$$\mathbb{E}\left[\mathbb{1}_{T<\infty} F G \circ \theta_T | \mathscr{F}_0\right] = \mathbb{E}\left[\mathbb{1}_{T<\infty} F \mathbb{E}'\left[G|\mathscr{F}'_0\right](X_T)\middle|\mathscr{F}_0\right] \tag{1.2.2}$$

where $\mathbb{E}'$ and $\mathscr{F}'$ refer to an independent copy, X', of the Markov process X.

Proof. We have

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}_{T<\infty} F G \circ \theta_T | \mathscr{F}_0\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}_{T<\infty} F G \circ \theta_T | \mathscr{F}_T\right] | \mathscr{F}_0\right] \\ &= \mathbb{E}\left[\mathbb{1}_{T<\infty} F \mathbb{E}\left[G \circ \theta_T | \mathscr{F}_T\right] | \mathscr{F}_0\right]. \end{aligned} \tag{1.2.3}$$

Now $\mathbb{E}[G \circ \theta_T | \mathscr{F}_T]$ depends only on X_T (and will thus often be denoted simply as $\mathbb{E}[G \circ \theta_T | X_T]$) and by stationarity is equal to $\mathbb{E}'[G|\mathscr{F}'_0](X_T)$, which yields the claim of the theorem. □

1.3 Markov processes and martingales

We now take a different look at Markov processes that will become important and more difficult in the continuous time case. First we want to see how the transition kernels can be seen as operators acting on spaces of measures respectively spaces of function.

If μ is a σ-finite measure on S, and P is a Markov transition kernel, we define the measure μP as

$$\mu P(A) \equiv \int_S P(x,A)\mu(dx), \tag{1.3.1}$$

and similarly, for the t-step transition kernel, P_t,

$$\mu P_t(A) \equiv \int_S P_t(x,A)\mu(dx). \tag{1.3.2}$$

By the Markov property, we have

$$\mu P_t(A) = \mu P^t(A). \tag{1.3.3}$$

Note that the action of ? on measures conserves the total mass, i.e.

$$\mu P(\Sigma) = \int_S P(x,S)\mu(dx) = \mu(S). \tag{1.3.4}$$

The action on measures has of course the following natural interpretation in terms of the process: if $\mathbb{P}(X_0 \in A) = \mu(A)$, then

$$\mu(X_t \in A) = \mu P_t(A). \tag{1.3.5}$$

Alternatively, if f is a bounded, measurable function on S, we define

$$(Pf)(x) \equiv \int_S f(y)P(x,dy), \tag{1.3.6}$$

and

$$(P_t f)(x) \equiv \int_S f(y)P_t(x,dy), \tag{1.3.7}$$

where again

$$P_t f = P^t f. \tag{1.3.8}$$

Lemma 1.9. *Let* $\|f\|_\infty \equiv \sup_{x\in S} |f(x)$ *denote the supremums norm. Then for any bounded function* f*,*

$$\|Pf\|_\infty \leq \|f\|_\infty. \tag{1.3.9}$$

Proof. Simply note that

$$\|Pf\|_\infty = \left\| \int_S P(x,dy)f(y) \right\|_\infty \leq \|f\|_\infty \int_S P(x,dy) = \|f\|_\infty. \tag{1.3.10}$$

□

We say that P_t is a semi-group acting on the space of measures, respectively on the space of bounded measurable functions. The interpretation of the action on functions is given as follows.

Lemma 1.10. *Let* P_t *be a Markov semi-group acting on bounded measurable functions* f*. Then*

$$(P_t f)(x) = \mathbb{E}\left(f(X_t)|\mathscr{F}_0\right)(x) \equiv \mathbb{E}_x f(X_t). \tag{1.3.11}$$

Proof. We only need to show this for $t = 1$. Then, by definition,

$$\mathbb{E}_x f(X_1) = \int_S f(y)\mathbb{P}[X_1 \in dy|\mathscr{F}_0](x) = \int_S f(y)P(x,dy).$$

□

Notice that, by telescopic expansion, we have the elementary formula

$$P_t f - f = \sum_{s=0}^{t-1} P_s(P - \mathbb{1}) f \equiv \sum_{s=0}^{t-1} P_s L f, \tag{1.3.12}$$

where we call $L \equiv P - \mathbb{1}$ the (discrete) generator of our Markov process (this formula will have a complete analog in the continuous-time case).

An interesting consequence is the following observation:

Lemma 1.11. *[Discrete time martingale problem]. Let L be the generator of a Markov process, X_t, and let f be a bounded measurable function. Then*

$$M_t \equiv f(X_t) - f(X_0) - \sum_{s=0}^{t-1} L f(X_s) \tag{1.3.13}$$

is a martingale.

Proof. Let $t, r \geq 0$. Then

$$\begin{aligned} \mathbb{E}(M_{t+r}|\mathscr{F}_t) &= \mathbb{E}(f(X_{t+r})|\mathscr{F}_t) - \mathbb{E}(f(X_0)|\mathscr{F}_t) - \sum_{s=0}^{t+r-1} \mathbb{E}(L f(X_s)|\mathscr{F}_t) \\ &= P^r f(X_t) - f(X_t) + f(X_t) - f(X_0) \\ &\quad - \sum_{s=t}^{t+r-1} \mathbb{E}(L f(X_s)|\mathscr{F}_t) - \sum_{s=0}^{t-1} \mathbb{E}(L f(X_s)|\mathscr{F}_t) \\ &= f(X_t) - f(X_0) - \sum_{s=0}^{t-1} (L f(X_s) \\ &\quad + P^r f(X_t) - f(X_t) - \sum_{s=0}^{r-1} P^s(L f(X_t)) \\ &= M_t + 0. \end{aligned} \tag{1.3.14}$$

This proves the lemma. □

Remark 1.12. (1.3.13) is of course the Doob decomposition of the process $f(X_t)$, since $\sum_{s=0}^{t-1} L f(X_s)$ is a previsible process.

What is important about this observation is that it gives rise to a characterisation of the generator that will be extremely useful in the general continuous time setting.

Namely, one can ask whether the requirement that M_t be a martingale given a family of pairs (f, Lf) characterises fully a Markov process.

Theorem 1.13. *Let X be a discrete time stochastic process on a filtered space such that X is adapted. Then X is a Markov process with transition kernel $P \equiv \mathbb{1} + L$, if and only if, for all bounded measurable functions, f, the expression on the right-hand side of (1.3.13) is a martingale.*

Proof. Lemma 1.11 already provides the "only if" part, so it remains to show the "if" part.

First, if we assume that X is a Markov process, setting $r = 1$ in (1.3.13) and taking conditional expectations given $\mathscr{F}_0$, we see that $\mathbb{E}f(X_1) - f(X_0) = (Lf)(X_0)$, implying that the transition kernel must be $\mathbb{1} + L$.

It remains to show that X is indeed a Markov process. To see this, we just use the above calculation, which gives

$$\begin{aligned}
\mathbb{E}(f(X_{t+r})|\mathscr{F}_t) &= \mathbb{E}(M_{t+r}|\mathscr{F}_t) + f(X_0) \\
&\quad + \sum_{s=0}^{t-1}(Lf)(X_s) + \sum_{s=t}^{t+r-1}\mathbb{E}((Lf)(X_s)|\mathscr{F}_t) \\
&= M_t + f(X_0) + \sum_{s=0}^{t-1}(Lf)(X_s) + \sum_{s=t}^{t+r-1}\mathbb{E}((Lf)(X_s)|\mathscr{F}_t) \\
&= f(X_t) + \sum_{s=0}^{r-1}\mathbb{E}((Lf)(X_{t+s})|\mathscr{F}_t) \qquad (1.3.15)
\end{aligned}$$

Now let again $r = 1$. Then

$$\mathbb{E}(f(X_{t+1})|\mathscr{F}_t) = f(X_t) + (Lf)(X_t) = ((\mathbb{1} + L)f)(X_t) \equiv Pf(X_t), \qquad (1.3.16)$$

In view of the definition of discrete time Markov processes, choosing $f = \mathbb{1}_A$, for $A \in \mathscr{B}(S)$, this gives (1.1.1), and hence X is a Markov process. Thus the theorem is proven. □

In view of continuous time Markov processes it is, however, instructive to see that we can also derive easily the more general fomula

$$\mathbb{E}(f(X_{t+s})|\mathscr{F}_t) = (\mathbb{1} + L)^s f(X_t) \equiv P^s f(X_t), \qquad (1.3.17)$$

from the martingale problem. We have seen that it holds for $s = 1$; Now proceed by induction: assume that it holds for all bounded measurable functions for $s \le r - 1$. We must show that it then also holds for $s = r$. To do this, we use (1.3.15) and use the induction hypothesis for the terms in the sum (where $s \le r - 1$) with f replaced by Lf. This gives

$$\begin{aligned}
\mathbb{E}(f(X_{t+r})|\mathscr{F}_t) &= f(X_t) + \sum_{s=0}^{r-1}((\mathbb{1} + L)^s Lf)(X_t) \qquad (1.3.18) \\
&= f(X_t) + \sum_{s=0}^{r-1}(((\mathbb{1} + L)^s(L + \mathbb{1})f)(X_t) - ((\mathbb{1} + L)^s f)(X_t)) \\
&= ((\mathbb{1} + L)^r f)(X_t),
\end{aligned}$$

as claimed. Hence (1.3.17) holds for all s, by induction.

Remark 1.14. The analog of this theorem in the continuous time case will bring out the full strength of this approach. A crucial point is that it will not be necessary to consider all bounded functions, but just sufficiently rich classes. This allows to formulate martingale problems even then one cannot write down the generator in a explicit form. The idea of characterising Markov processes by the associated martingale problem goes back to Stroock and Varadhan, see [16].

1.4 Harmonic functions and martingales

We have seen that measures that satisfy $\mu L = 0$ are of special importance in the theory of Markov processes. Also of central importance are functions that satisfy $Lf = 0$. In this section we will assume that the transition kernels of our Markov processes have bounded support, so that for some $K < \infty$, $|X_{t+1} - X_t| \leq K < \infty$ for all t.

Definition 1.15. Let L be the generator of a Markov process. A measurable function that satisfies

$$Lf(x) = 0, \forall x \in S, \tag{1.4.1}$$

is called a *harmonic function*. A function is called *subharmonic* (resp. *superharmonic*, if $Lf \geq 0$, resp. $Lf \leq 0$.

Theorem 1.16. *Let X_t be a Markov process with generator L. Then, a non-negative function f is*

(i) harmonic, if and only if $f(X_t)$ is a martingale;
(ii)subharmonic, if and only if $f(X_t)$ is a submartingale;
(iii)super-harmonic, if and only if $f(X_t)$ is a supermartingale;

Proof. Simply use Lemma 1.11. □

Remark 1.17. Theorem 1.16 establishes a profound relationship between potential theory and martingales. It also explains, the strange choice of super and sub in martingale theory.

A nice application of the preceding result is the maximum principle.

Theorem 1.18. *Let X be a Markov process and let D be a bounded open domain such that $\mathbb{E}\tau_{D^c} < \infty$. Assume that f is a non-negative subharmonic function on D. Then*

$$\sup_{x \in D} f(x) \leq \sup_{x \in D^c} f(x). \tag{1.4.2}$$

Proof. Let us define $T \equiv \tau_{D^c}$. Then, $f(X_T)$ is a submartingale, and thus

$$\mathbb{E}\left(f(X_T)|\mathscr{F}_0\right)(x) \geq f(x). \tag{1.4.3}$$

Since $X_T \in D^c$, it must be true that

$$\sup_{y \in D_c} f(y) \geq \mathbb{E}\left(f(X_T)|\mathscr{F}_0\right)(x) \geq f(x), \tag{1.4.4}$$

for all $x \in D$, hence the claim of the theorem. Of course we used again the Doob's optional stopping theorem. □

The theorem says that (sub) harmonic functions take on their maximum on the boundary, since of course the set D^c in (1.4.2) can be replaced by a subset, $\partial D \subset D^c$ such that $\mathbb{P}_x(X_T \in \partial D) = 1$. The above proof is an example of how intrinsically analytic results can be proven with probabilistic means. The next section will further develop this theme.

1.5 Dirichlet problems

Let us now consider a connected bounded open subset of S. We define the stopping time $T \equiv \tau_{D^c}$.

If g is a measurable function on D, we consider the Dirichlet problem associated to a generator, L, of a Markov process, X:

$$\begin{aligned} -(Lf)(x) &= g(x), \quad x \in D, \\ f(x) &= 0, \quad x \in D^c. \end{aligned} \tag{1.5.1}$$

Theorem 1.19. *Assume that $\mathbb{E}T < \infty$. Then (1.5.1) has a unique solution given by*

$$f(x) = \mathbb{E}\left(\sum_{t=0}^{T-1} g(X_t)\Big|\mathscr{F}_0\right)(x) \tag{1.5.2}$$

Proof. Consider the martingale M_t from Lemma 1.11. We know from Doob's optional stopping theorem (see e.g. [15]) that M_T is also a martingale. Moreover,

$$M_T = f(X_T) - f(X_0) - \sum_{t=0}^{T-1}(Lf)(X_t) = 0 - f(X_0) - \sum_{t=0}^{T-1}(Lf)(X_t). \tag{1.5.3}$$

But we want f such that $-Lf = g$ on D. Thus, (1.5.3) seen as a problem for f, reads

$$M_T = -f(X_0) + \sum_{t=0}^{T-1} g(X_t). \tag{1.5.4}$$

Taking expectations conditioned on $\mathscr{F}_0$, yields

$$0 = -f(X_0) + \mathbb{E}\left(\sum_{t=0}^{T-1} g(X_t)\Big|\mathscr{F}_0\right), \tag{1.5.5}$$

or

$$f(x) = \mathbb{E}_x \left(\sum_{t=0}^{T-1} g(X_t) \right) \tag{1.5.6}$$

Here we relied of course on Doob's optimal stopping theorem for $\mathbb{E}M_T = 0$.

Thus any solution of the Dirichlet problem is given by (1.5.6). To verify existence, we just need to check that (1.5.6) solves $-Lf = g$ on D. To do this we use the Markov property "backwards", to see that

$$\begin{aligned} Pf(x) &= P\mathbb{E}_x \left(\sum_{t=0}^{T-1} g(X_t) \right) = \mathbb{E}_x \left[\sum_{t=1}^{T-1} g(X_t) \right] \\ &= \mathbb{E}_x \left[\sum_{t=0}^{T-1} g(X_t) \right] - g(x) = f(x) - g(x). \end{aligned} \tag{1.5.7}$$

□

We see that the Markov process produces a solution of the Dirichlet problem. We can express the solution in terms of an integral kernel, called the Green's kernel, $G_D(x,dy)$, as

$$f(x) = \int G_D(x,dy) g(y) \equiv \mathbb{E}_x \left(\sum_{t=0}^{T-1} g(X_t) \right), \tag{1.5.8}$$

or, in more explicit terms,

$$G_D(x,dy) = \sum_{t=0}^{\infty} P_D^t(x,dy), \tag{1.5.9}$$

where

$$P_D^t(x,dy) = \int_D P(x,dz_1) \int_D P(z_1,dz_2) \int_D \dots \int_D P(z_{t-1},dy). \tag{1.5.10}$$

Note that $\int_D P(x,dz) < 1$.

The preceding theorem has an obvious extension to more complicated boundary value problems.

Let $D \subset S$ be as above and specify functions $g\colon D \to \mathbb{R}$, $u\colon D^c \to \mathbb{R}$ and $k\colon D \to [-\bar{k},\infty)$ with $\bar{k} < 1$. Consider the following set of equations for an unknown function f:

$$\begin{aligned} (-Lf)(x) + k(x) f(x) &= g(x), \quad \forall x \in D, \\ f(x) &= u(x), \quad \forall x \in D^c. \end{aligned} \tag{1.5.11}$$

The following theorem provides a stochastic representation of the solution of such Dirichlet problems.

Theorem 1.20. *Let X be a discrete-time Markov process with generator L. Assume that D is such that*

$$\mathbb{E}_x \left[\tau_{D^c} (1 - \bar{k})^{\tau_{D^c}} \right] < \infty. \tag{1.5.12}$$

Then the Dirichlet problem (1.5.11) *has a unique solution given by*

$$f(x) = \mathbb{E}_x \left[\sum_{s=0}^{\tau_{D^c}-1} \prod_{u=0}^{s} \frac{1}{1+k(X_u)} g(X_s) + \prod_{u=0}^{\tau_{D^c}-1} \frac{1}{1+k(X_u)} u(X_{\tau_{D^c}}) \right].$$

Proof. The most convenient way to prove Theorem 1.20 is again via the martingale problem characterisation of Markov processes. Indeed, we check that, for any bounded function f,

$$\begin{aligned} M_t \equiv & \prod_{s=0}^{t-1} \frac{1}{1+k(X_s)} f(X_t) - f(X_0) \\ & + \sum_{s=0}^{t-1} \prod_{u=0}^{s} \frac{1}{1+k(X_u)} \left[k(X_s) f(X_s) - (Lf)(X_s)\right] \end{aligned} \tag{1.5.13}$$

is a martingale. Moreover, Doob's optional stopping theorem applies for $M_{\tau_{D^c}}$ under condition (1.5.12). Thus as before, if f solves the Dirichlet problem (1.5.11), it must hold that

$$\begin{aligned} 0 = \mathbb{E}_x M_{\tau_D^c} = \mathbb{E}_x \Bigg(& \prod_{s=0}^{\tau_{D^c}-1} \frac{1}{1+k(X_s)} u(X_{\tau_{D^c}}) - f(x) \\ & + \sum_{s=0}^{\tau_{D^c}-1} \prod_{u=0}^{s} \frac{1}{1+k(X_u)} g(X_s) \Bigg), \end{aligned} \tag{1.5.14}$$

which implies that (1.5.13) must hold. Finally one shows that this solves the Equation (1.5.11) as in the proof of Theorem 1.19. □

Note that the solution to the Dirichlet problem is unique, unless the homogeneous problem

$$\begin{aligned} (-Lf)(x) + k(x) f(x) &= 0, \quad \forall x \in D, \\ f(x) &= 0, \quad \forall x \in D^c, \end{aligned} \tag{1.5.15}$$

admits a non-zero solution. The most interesting case for us is when $k \equiv \lambda$ is constant. In that case, if (1.5.15) admits a non-zero solution, then λ is called an *eigenvalue* and the corresponding solution an *eigenfunction* of the Dirichlet problem.

Theorem 1.20 is a two way game: it allows to produce solutions of analytic problems in terms of stochastic processes, and it allows to compute interesting probabilistic problems analytically. As an example, assume that $D^c = A \cup B$ with $A \cap B = \emptyset$. Set $h = \mathbb{1}_A$. Then, clearly, for $x \in D$,

$$\mathbb{E}_x h(X_T) = \mathbb{P}_x(X_T \in A) \equiv \mathbb{P}_x(\tau_A < \tau_B), \tag{1.5.16}$$

and so $\mathbb{P}_x(X_T \in A)$ can be represented as the solution of the boundary value problem

$$\begin{aligned}(Lf)(x) &= 0, \quad x \in D, \\ f(x) &= 1, \quad x \in A, \\ f(x) &= 0, \quad x \in B.\end{aligned} \tag{1.5.17}$$

This is a generalisation of the *ruin* problem for the random walk.

Exercise. Derive the formula for $\mathbb{P}_x(\tau_A < \tau_B)$ directly from the Markov property without using Lemma 1.11.

1.6 Feynman-Kac formulas

The formalism expained in the preceding section has a useful extension to the solution of time-dependent problems of the form

$$\partial_t f(x,t) - Lf(x,t) + k(x)f(x,t) = g(x), \quad x \in S, t \in [0,T], \tag{1.6.1}$$

$$f(x,T) = \psi(x), \quad x \in S, \tag{1.6.2}$$

where $\partial_t f(x,t) \equiv f(x,t) - f(x,t-1)$ denotes the discrete deriative with respect to time. k, g, ψ are given functions, and T is a fixed time.

To obtain a stochstic representation of the solution of such equations, we proceed by extending the telescopic expansions that yield martingales to functions f that depend on both X_t and t. This alows to show that

$$\begin{aligned}M_t \equiv & \prod_{s=0}^{t-1} \frac{1}{1+k(X_s)} f(X_t,t) - f(X_0,0) \\ & + \sum_{s=0}^{t-1} \prod_{u=0}^{s} \frac{1}{1+k(X_u)} \left[k(X_s) f(X_s,s) - (Lf)(X_s,s) + \partial_s f(X_s,s)\right],\end{aligned} \tag{1.6.3}$$

where $\partial_s f(X,s) \equiv f(X,s) - f(X,s-1)$, is a martingale. Therefore, for $t < T$,

$$\mathbb{E}(M_T | \mathscr{F}_t) = M_t. \tag{1.6.4}$$

Now the left-hand side of (1.6.4) is equal to

$$
\begin{aligned}
&\prod_{s=0}^{t-1}\frac{1}{1+k(X_s)}\,\mathbb{E}\left(\prod_{s=t}^{T-1}\frac{1}{1+k(X_s)}\,\psi(X_T)|\mathscr{F}_t\right)-f(X_0,0) \qquad (1.6.5)\\
&+\sum_{s=0}^{t-1}\prod_{u=0}^{s}\frac{1}{1+k(X_u)}\,g(X_s)\\
&+\sum_{s=t}^{T-1}\prod_{u=0}^{t-1}\frac{1}{1+k(X_u)}\,\mathbb{E}\left(\prod_{u=t}^{s}\frac{1}{1+k(X_u)}\,g(X_s)\Big|\mathscr{F}_t\right)\\
&=M_t-\prod_{s=0}^{t-1}\frac{1}{1+k(X_s)}\,f(X_t,t)\\
&+\prod_{s=0}^{t-1}\frac{1}{1+k(X_s)}\,\mathbb{E}\left(\prod_{s=t}^{T-1}\frac{1}{1+k(X_s)}\,\psi(X_T)+\sum_{s=t}^{T-1}\prod_{u=t}^{s}\frac{1}{1+k(X_u)}\,g(X_s)\Big|\mathscr{F}_t\right).
\end{aligned}
$$

Thus we arrive at the representation of the solution of (1.6.1)

$$
f(x,t)=\mathbb{E}\left(\prod_{s=t}^{T-1}\frac{1}{1+k(X_s)}\,\psi(X_T)+\sum_{s=t}^{T-1}\prod_{u=t}^{s}\frac{1}{1+k(X_u)}\,g(X_s)\Big|\mathscr{F}_t\right)(x). \qquad (1.6.6)
$$

This representation is called a *Feynman-Kac formula*. In the case when $g\equiv 0$ and $k\equiv 0$, it simplifies to

$$
f(x,t)=\mathbb{E}\left(\psi(X_T)|\mathscr{F}_t\right)(x)=\mathbb{E}_x\psi(X_{T-t}). \qquad (1.6.7)
$$

So far we have considered the case without boundary conditions. From the derivation above, it is, however, also easy to see how to deal with problems of the form

$$
\begin{aligned}
\partial_t f(x,t)-Lf(x,t)+k(x)f(x,t) &= g(x), \quad x\in D, t\in[0,T], \qquad (1.6.8)\\
f(x,t) &= \psi(x), \quad x\in D^c, t\in[0,T], \qquad (1.6.9)\\
f(x,T) &= \psi(x), \quad x\in S.
\end{aligned}
$$

Namely, defining the optional time $\hat{T}\equiv T\wedge\tau_{D^c}$, and noticing that $M_{\hat{T}}$ is always a martingale, we obtain Thus we arrive at the representation of the solution of (1.6.1)

$$
f(x,t)=\mathbb{E}\left(\prod_{s=t}^{\hat{T}-1}\frac{1}{1+k(X_s)}\,\psi(X_{\hat{T}})+\sum_{s=t}^{\hat{T}-1}\prod_{u=t}^{s}\frac{1}{1+k(X_u)}\,g(X_s)\Big|\mathscr{F}_t\right)(x). \qquad (1.6.10)
$$

1.7 Doob's *h*-transform

Let us consider a Markov process, X, with generator $P-1$. We may want to consider modifications of the process. One important type modification is to condition it to reach some set in particular places (e.g. consider a random walk in a finite interval; we may be interested to consider this walk conditioned on the fact that it exits on a specific side of the interval; this may correspond to consider a sequence of games conditioned on the player to win).

How and when can we do this, and what is the nature of the resulting process? In particular, is the resulting process again a Markov process, and if so, what is its generator?

As an example, let us try to condition a Markov process to hit a domain B for the first time in a subset $A \subset B$. We may assume that $\mathbb{E}\tau_B < \infty$. Define $h(x) \equiv \mathbb{P}_x[\tau_A = \tau_B]$, if $x \notin B$. Let $\mathbb{P}$ be the law of X. Let us define a new measure, $\mathbb{P}^h$, on the space of paths as follows: If Y is a $\mathscr{F}_t$-measurable random variable, then

$$\mathbb{E}^h[Y|\mathscr{F}_0] = \frac{1}{h(X_0)}\mathbb{E}[h(X_t)Y|\mathscr{F}_0]. \tag{1.7.1}$$

Lemma 1.21. *With the notation above, if Y is a $\mathscr{F}_{\tau_B-1}$-measurable function,*

$$\mathbb{E}^h_x[Y] = \mathbb{E}_x[Y|\tau_A = \tau_B]. \tag{1.7.2}$$

Proof. This is an application of the strong Markov property. By definition,

$$\begin{aligned}\mathbb{E}^h_x[Y] &= \frac{1}{h(x)}\mathbb{E}_x[Yh(X_{\tau_B-1})] \qquad (1.7.3)\\ &= \frac{1}{h(x)}\mathbb{E}_x\left[Y\mathbb{E}'[\mathbb{1}_{\tau_A=\tau_B}|\mathscr{F}'_0](X_{\tau_B-1})\right]\\ &= \frac{1}{h(x)}\mathbb{E}_x\left[Y\mathbb{E}\left[\mathbb{1}_{\tau_A=\tau_B}|\mathscr{F}_{\tau_B-1}\right]\right]\\ &= \frac{1}{\mathbb{P}_x[\tau_A=\tau_B]}\mathbb{E}_x\left[Y\mathbb{1}_{\tau_A=\tau_B}\right]\\ &= \mathbb{E}_x[Y|\tau_A=\tau_B].\end{aligned}$$

Here the first equality is just the definition of h and reproduces the form of the right-hand side of the strong Markov property; the second equality is the strong Markov property; the last equality uses that fact that the event $\{\tau_A = \tau_B\}$ depends only on what happens after $\tau_B - 1$, and so $\mathbb{1}_{\tau_A=\tau_B}\theta_{\tau_B-1} = \mathbb{1}_{\tau_A=\tau_B}$. □

Let us now look at the transformed measure $\mathbb{P}^h$ in the general case. The first thing to check is of course whether this defines in a consistent way a probability measure. Some thought shows that all that we need is the following lemma.

Lemma 1.22. *Let Y be $\mathscr{F}_s$-measurable. Then, for any $t \geq s$,*

$$\mathbb{E}^h[Y|\mathscr{F}_0] \equiv \frac{1}{h(X_0)}\mathbb{E}[h(X_s)Y|\mathscr{F}_0] = \frac{1}{h(X_0)}\mathbb{E}[h(X_t)Y|\mathscr{F}_0]. \tag{1.7.4}$$

In particular, $\mathbb{P}^h[\Omega|\mathscr{F}_0] = 1$.

Proof. Just introduce a conditional expectation:

$$\mathbb{E}[h(X_t)Y|\mathscr{F}_0] = \mathbb{E}[\mathbb{E}[h(X_t)Y|\mathscr{F}_s]|\mathscr{F}_0] = \mathbb{E}[Y\mathbb{E}[h(X_t)|\mathscr{F}_s]|\mathscr{F}_0], \tag{1.7.5}$$

and use that $h(X_t)$ is a martingale

$$= \mathbb{E}[Yh(X_s)|\mathscr{F}_0],$$

from which the result follows. □

This lemma shows in particular, why it is important that h be a harmonic function.

Now we turn to the question of whether the law $\mathbb{P}^h$ is a Markov process. To this end we turn to the martingale problem. We will show that there exists a generator, L^h, such that

$$M_t^h \equiv f(X_t) - f(X_0) - \sum_{s=0}^{t-1}(L^h f)(X_s) \tag{1.7.6}$$

is a martingale under the law $\mathbb{E}^h$, i.e. that, for $t > t'$,

$$\mathbb{E}^h[M_t^h|\mathscr{F}_{t'}] = M_{t'}^h. \tag{1.7.7}$$

Note first that, by definition

$$\begin{aligned}\mathbb{E}^h[M_t^h|\mathscr{F}_{t'}] = &\frac{1}{h(X_{t'})}\mathbb{E}[h(X_t)f(X_t)|\mathscr{F}_{t'}] - f(X_0) - \sum_{s=0}^{t'-1}(L^h f)(X_s)\\ &- \sum_{s=t'}^{t-1}\frac{1}{h(X_{t'})}\mathbb{E}[h(X_s)L^h f(X_s)|\mathscr{F}_{t'}].\end{aligned} \tag{1.7.8}$$

The middle terms are part of $M_{t'}^h$ and we must consider $\mathbb{E}[f(X_t)h(X_t)|\mathscr{F}_{t'}]$. This is done by applying the martingale problem for $\mathbb{P}$ and the function fh. This yields

$$\mathbb{E}[f(X_t)h(X_t)|\mathscr{F}_{t'}] = f(X_{t'})h(X_{t'}) + \sum_{s=t'}^{t-1}\mathbb{E}[(L(fh))(X_s)|\mathscr{F}_{t'}]$$

Inserting this in (1.7.8) gives

$$\begin{aligned}\mathbb{E}^h[M_t^h|\mathscr{F}_{t'}] = &f(X_{t'}) - f(X_0) - \sum_{s=0}^{t'-1}(L^h f)(X_s)\\ &+\frac{1}{h(X_{t'})}\sum_{s=t'}^{t-1}\Big[\mathbb{E}[(L(fh))(X_s)|\mathscr{F}_{t'}] - \mathbb{E}[h(X_s)L^h f(X_s)|\mathscr{F}_{t'}]\Big]\\ = &M_{t'}^h\\ &+\frac{1}{h(X_{t'})}\sum_{s=t'}^{t-1}\Big[\mathbb{E}[(L(fh))(X_s)|\mathscr{F}_{t'}] - \mathbb{E}[h(X_s)L^h f(X_s)|\mathscr{F}_{t'}]\Big].\end{aligned}$$

The second term will vanish if we choose L^h defined through $Lf(x) = h(x)^{-1}(L(hf))(x)$, i.e.

$$L^h f(x) \equiv \frac{1}{h(x)} \int P(x,dy)h(y)f(y) - f(x). \tag{1.7.9}$$

Hence we see that under $\mathbb{P}^h$, X solves the martingale problem corresponding to the generator L^h, and so is a Markov process with transition kernel $P^h = L^h + 1$. The process X under $\mathbb{P}^h$ is called the (Doob) h-transform of the original Markov process.

Exercise. As a simple example, consider a simple random walk on $\{-N, -N+1, \dots, N\}$. Assume we want to condition this process on hitting $+N$ before $-N$. Then let

$$h(x) = \mathbb{P}_x[\tau_N = \tau_{\{N\}\cup\{-N\}}] = \mathbb{P}_x[\tau_N < \tau_{-N}].$$

Compute $h(x)$ and use this to compute the transition rates of the h-transformed walk? Plot the probabilities to jump down in the new process!

Chapter 2
Continuous time martingales

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 [15].

2.1 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 2.1. 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.

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

Definition 2.2. A function $y : \mathbb{Q}_+ \to \mathbb{R}$ is called *regularisable*, iff

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

(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 2.3. 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 2.4. *Let $y : \mathbb{Q}_+ \to \mathbb{R}$. Then y is regularisable if and only if, for all $N \in \mathbb{N}$ and $a < b \in \mathbb{R}$,*

$$\sup\{|y(q)| : q \in \mathbb{Q} \cap [0,N]\} < \infty, \tag{2.1.1}$$

and

$$U_N(y,[a,b]) < \infty. \tag{2.1.2}$$

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 (2.1.2). Thus the limit $\lim_{q \downarrow t} y(q)$ exists, and by (2.1.1) 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 2.4 is that it can be used to show that the property to be regularisable is measurable.

Corollary 2.5. *Let $\{Y_q, q \in \mathbb{Q}_+\}$ be a stochastic process defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and let*

$$G \equiv \{\omega \in \Omega : q \to Y_q(\omega) \text{ is regularisable}\} \tag{2.1.3}$$

Then $G \in \mathscr{F}$.

Proof. By Theorem 2.4, 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 2.6. *Let $y : \mathbb{Q}_+ \to \mathbb{R}$ be a regularisable function. Define, for any $t \in \mathbb{R}_+$,*

$$f(t) \equiv \lim_{q \downarrow t} y(q). \tag{2.1.4}$$

Then f is càdlàg .

The proof is obvious and left to the reader.

2.2 Filtrations, supermartingales, and càdlàg processes

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

Definition 2.7. A filtration $(\mathscr{G}_t, t \in \mathbb{R}_+)$ of $(\Omega, \mathscr{G}, \mathbb{P})$ is an increasing family of sub-σ-algebras $\mathscr{G}_t$, such that, for $0 \leq s < t$,

$$\mathscr{G}_s \subset \mathscr{G}_t \subset \mathscr{G}_\infty \equiv \sigma\left(\bigcup_{r \in \mathbb{R}_+} \mathscr{G}_r\right) \subset \mathscr{G}. \tag{2.2.1}$$

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

Definition 2.8. A stochastic process, $\{X_t, t \in \mathbb{R}_+\}$, is called *adapted* to the filtration $\{\mathscr{G}_t, t \in \mathbb{R}_+\}$, if, for every t, X_t is $\mathscr{G}_t$-measurable.

Definition 2.9. 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 $\{\mathscr{G}_t, t \in \mathbb{R}_+\}$;
(ii)For all $t \in \mathbb{R}_+$, $\mathbb{E}|X_t| < \infty$;
(iii)For all $s \leq t \in \mathbb{R}_+$,

$$\mathbb{E}(X_t | \mathscr{G}_s) = X_s, \text{ a.s.}. \tag{2.2.2}$$

Sub- and super-martingales are define in the same way, with "$=$" in (2.2.2) replaced by "$\geq$" resp. "$\leq$".

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 2.10. *Let Y be a supermartingale on a filtered space $(\Omega, \mathscr{G}, \mathbb{P}; (\mathscr{G}_t, t \in \mathbb{R}_+))$. 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 $\mathscr{L}^1$.

Proof. This is an application of the Lévy-Doob downward theorem (see [1], Thm. 4.2.9).

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

Definition 2.11. 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 2.12. Note that we do not just ask that almost all sample paths are càdlàg .

2.3 Doob's regularity theorem

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

Theorem 2.13. *Let $(Y_t, t \in \mathbb{R}_+)$ be a supermartingale defined on a filtered space $(\Omega, \mathscr{G}, \mathbb{P}, (\mathscr{G}_t, t \in \mathbb{R}_+))$. Define the set*

$$G \equiv \{\omega \in \Omega : \text{the map}\, \mathbb{Q}_+ \ni q \to Y_q(\omega) \in \mathbb{R} \text{ is regularisable}\}. \tag{2.3.1}$$

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

$$X_t(\omega) \equiv \begin{cases} \lim_{q\downarrow t} Y_q(\omega), & \text{if } \omega \in G, \\ 0, & \text{else} \end{cases} \tag{2.3.2}$$

is a càdlàg process.

Proof. The proof makes use of our observations in Theorem 2.4. There are only countably many triples (N, a, b) with $N \in \mathbb{N}$, $a < b \in \mathbb{Q}$. Thus in view of Theorem 2.4, we must show that with probability one,

$$\sup_{q\in\mathbb{Q}\cap[0,N]} |Y_q| < \infty, \tag{2.3.3}$$

and

$$U_N([a,b]; Y|_{\mathbb{Q}}) < \infty, \tag{2.3.4}$$

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

$$\mathbb{P}\left[\sup_{q \in \mathbb{Q} \cap [0,N]} |Y_q| > 3c\right] = \lim_{m \uparrow \infty} \mathbb{P}\left[\sup_{q \in D(m)} |Y_q| > 3c\right] \le c^{-1}\left(4\mathbb{E}|Y_0| + 3\mathbb{E}|Y_N|\right), \tag{2.3.5}$$

by Lemma 4.4.15 in [1]. Taking $c \uparrow \infty$ (2.3.3) 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 [1], we get that

$$\mathbb{E}\left[U_N([a,b];Y|_{\mathbb{Q}}\right] = \lim_{m \uparrow \infty} \mathbb{E}\left[U_N([a,b];Y|_{D(m)})\right] < \infty \le \frac{\mathbb{E}|Y_N| + |a|}{b-a}, \tag{2.3.6}$$

uniformly in m, and so (2.3.4) also follows.

Now Theorem 2.4 implies the asserted result.

We may think that Theorem 2.13 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

$$Y_t(\omega) = \begin{cases} 0, & \text{if } t \le 1, \\ q(\omega), & \text{if } t > 1, \end{cases} \tag{2.3.7}$$

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

$$X_t(\omega) = \begin{cases} 0, & \text{if } t < 1, \\ q(\omega), & \text{if } t \ge 1, \end{cases} \tag{2.3.8}$$

Now first, X_t is not adapted to the filtration $\mathscr{G}_t$, since X_1 is not measurable with respect to $\mathscr{G}_1$. 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 $\mathscr{G}_t$, since

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

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

$$\mathscr{G}_{t+} \equiv \bigcap_{s>t} \mathscr{G}_s = \bigcap_{\mathbb{Q} \ni q > t} \mathscr{G}_q \tag{2.3.9}$$

and let

$$\mathscr{N}(\mathscr{G}_\infty) \equiv \{G \in \mathscr{G}_\infty : \mathbb{P}[G] \in \{0,1\}\}. \tag{2.3.10}$$

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

$$\mathscr{H}_t \equiv \sigma(\mathscr{G}_{t+}, \mathscr{N}(\mathscr{G}_\infty)). \tag{2.3.11}$$

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

Lemma 2.15. *If Y_t is a supermartingale with respect to the filtration $\mathscr{G}_t$, and X_t is its càdlàg version defined in Theorem 2.13, then X_t is adapted to the partially augmented filtration $\mathscr{H}_t$.*

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 2.13

Theorem 2.16. *With the assumptions and notations of Lemma 2.15, the process X_t is a supermartingale with respect to the filtrations $\mathscr{H}_t$. Moreover, X is a modification of Y if and only if Y is right-continuous in the sense that, for every $t \in \mathbb{R}_+$,*

$$\lim_{s \downarrow t} \mathbb{E}|Y_t - Y_s| = 0. \tag{2.3.12}$$

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

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

By the Lévy-Doob downward theorem (Theorem 4.2.9 in [1]),

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

Thus

$$\mathbb{E}[Y_s|\mathscr{H}_t] \leq X_t.$$

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

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

On the other hand, Lemma 2.10 and Theorem 2.13, $Y_{q(n)} \to X_u$ in $\mathscr{L}^1$, so

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

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

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 [15]) is as as follows:

Definition 2.17. A filtered càdlàg space is a quadruple $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_t,t\in\mathbb{R}))$, where $(\Omega,\mathscr{F},\mathbb{P})$ is a probability space and $\mathscr{F}_t$ is a filtration of $\mathscr{F}$ that satisfies the following properties:

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

If $(\Omega,\mathscr{G},\mathbb{P},(\mathscr{G}_t,t\in\mathbb{R}_+))$ is a filtered space, then the the minimal enlargement of this space, $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_t,t\in\mathbb{R}_+))$ 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 2.18. *If $(\Omega,\mathscr{G},\mathbb{P},(\mathscr{G}_t,t\in\mathbb{R}_+))$ is filtered space, and $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_t,t\in\mathbb{R}_+))$ its right-continuous regularization, then*

(i) $\mathscr{F}$ is the $\mathbb{P}$-completion of $\mathscr{G}$ (i.e. the smallest σ-algebra containing $\mathscr{G}$ and all sets of $\mathbb{P}$-outer measure zero;
(ii) If $\mathscr{N}$ denotes the set of all $\mathbb{P}$-null sets in $\mathscr{F}$, then

$$\mathscr{F}_t \equiv \bigcap_{u>t}\sigma(\mathscr{G}_u,\mathscr{N}) = \sigma(\mathscr{G}_{t+},\mathscr{N}); \tag{2.3.13}$$

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

$$F\Delta G \in \mathscr{N}, \tag{2.3.14}$$

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

Proof. Exercise.

Proposition 2.19. *The process X constructed in Theorem 2.13 is a supermartingale with respect to the filtration $\mathscr{F}_t$.*

Proof. Since by (2.3.14) $\mathscr{F}_t$ and $\mathscr{H}_t$ differ only by sets of measure zero, $\mathbb{E}(X_{t+s}|\mathscr{F}_t)$ and $\mathbb{E}(X_{t+s}|\mathscr{H}_t)$ 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 2.20. *Let $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t, t \in \mathbb{R}_+))$ 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 $u \geq t$, $\mathbb{E}(Y_u)|\mathscr{F}_t) \leq Y_t$, a.s.. Construct the process X as in Theorem 2.13 Then

$$\mathbb{E}(X_t|\mathscr{F}_t) = \mathbb{E}\left(\lim_{u\downarrow t} Y_u|\mathscr{F}_t\right) = \lim_{u\downarrow t} \mathbb{E}\left(Y_u|\mathscr{F}_t\right) \leq Y_t, \text{ a.s..} \tag{2.3.15}$$

since $Y_u \downarrow Y_t$ in $\mathscr{L}^1$. Since X_t is adapted to $\mathscr{F}_t$, 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 $\mathscr{L}^1$-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.

2.4 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 $(\Omega, \mathscr{G} : \mathbb{P}, (\mathscr{G}_t, t \in \mathbb{R}_+))$.

Definition 2.21. A map $T : \Omega \to [0,\infty]$ is called a $\mathscr{G}_t$-stopping time if

$$\{T \leq t\} \equiv \{\omega \in \Omega : T(\omega) \leq t\} \in \mathscr{G}_t, \forall t \leq \infty. \tag{2.4.1}$$

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

$$\Lambda \cap \{T \leq t\} \in \mathscr{G}_t, \forall t \leq \infty. \tag{2.4.2}$$

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

Lemma 2.22. *Let S, T be stopping times. Then:*

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

Proof. Exercise.

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

Definition 2.23. A map $T : \Omega \to [0,\infty]$ is called a $\mathscr{G}_{t+}$-stopping time if

$$\{T < t\} \equiv \{\omega \in \Omega : T(\omega) < t\} \in \mathscr{G}_t, \forall t \leq \infty. \tag{2.4.3}$$

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

$$\Lambda \cap \{T < t\} \in \mathscr{G}_t, \forall t \leq \infty. \tag{2.4.4}$$

Lemma 2.24. *Let S_n be a sequence of $\mathscr{G}_t$-stopping times. Then:*

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

Proof. Consider case (i). Since S_n is increasing, the sequence of sets $\{S_n \leq t\} \in \mathscr{G}_t$ is decreasing, and its limit is also in $\mathscr{G}_t$. 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, $\{S < t\} = \bigcup_n \{S_n \leq t\} \in \mathscr{G}_t$.

Definition 2.25. A process $X_t, t \in \mathbb{R}_+$ is called $\mathscr{G}_t$*-progressive* if, for every $t \geq 0$, the restriction of the map $(s,\omega) \to X_s(\omega)$ to $[0,t]\times\Omega$ is $\mathscr{B}([0,t]\times\mathscr{G}_t$-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 σ-algebra.

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

Lemma 2.26. *An adapted càdlàg process with values in a metrisable space, $(S,\mathscr{B}(S))$, 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 2.27. *If X is progressive with respect to the filtration $\mathscr{G}_t$ and T is a $\mathscr{G}_t$-stopping time, then X_T is $\mathscr{G}_T$ measurable.*

Proof. For $t \geq 0$ let $\widehat{\Omega}_t \equiv \{\omega : T(\omega) \leq t\}$. Define $\widehat{\mathscr{G}_t}$ to be the sub-σ-algebra of $\mathscr{G}_t$ such that any set $A \in \widehat{\mathscr{G}_t}$ 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 $\mathscr{B}([0,t]) \times \mathscr{G}_t$ due to the progressiveness of X. ρ is measurable with respect to $\mathscr{G}_t$ 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 $\mathscr{G}_t$- 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 Γ

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

which proves the measurability of X_T.

2.5 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 2.28. Let X be a stochastic process with values in a measurable space $(E, \mathscr{E})$. Let $\Gamma \in \mathscr{E}$. We call

$$\tau_\Gamma(\omega) \equiv \inf\{t > 0 : X_t(\omega) \in \Gamma\} \tag{2.5.1}$$

the *first hitting time* of the set Γ; we call

$$\Delta_\Gamma(\omega) \equiv \inf\{t \geq 0 : X_t(\omega) \in \Gamma\} \tag{2.5.2}$$

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

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

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

Lemma 2.29. *Consider the case when E is a metric space and let F be a closed set. Let X be a continuous adapted process. Then Δ_F is a $\mathscr{G}_t$-stopping time and τ_F is a $\mathscr{G}_{t+}$-stopping time.*

Proof. Let ρ 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 $\mathscr{G}_q$ measurable, for $q \in \mathbb{Q}_+$. Since the paths $X_t(\omega)$ are continuous, $\Delta_F(\omega) \leq t$ if and only if

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

and so Δ_F is measurable w.r.t. $\mathscr{G}_t$. For τ_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 $\mathscr{G}_t$-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 $\mathscr{G}_\delta$-measurable, and so the event

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

is $\mathscr{G}_{0+}$-measurable and so τ_F is a $\mathscr{G}_{t+}$-stopping time.

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, τ_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!

2.6 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 $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_t,t\in\mathbb{R}_+))$ on which all processes will be defined and adapted.

Our aim is the following *optional sampling theorem*:

Theorem 2.30. *Let X be a càdlàg submartingale and let T,S be $\mathscr{F}_t$- stopping times. Then for each $M < \infty$,*

$$\mathbb{E}\left(X(T\wedge M)|\mathscr{F}_S\right)\geq X(S\wedge T\wedge M),\text{ a.s..}\tag{2.6.1}$$

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

$$\mathbb{E}\left(X(T)|\mathscr{F}_S\right)\geq X(S\wedge T),\text{ a.s..}\tag{2.6.2}$$

Equality holds in the case of martingales.

Proof. In order to prove Theorem 2.30 we frst prove a result for stopping times taking finitely many values.

Lemma 2.31. *Let S,T be $\mathscr{F}_t$ 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 $\mathscr{F}_t$-submartingale, then*

$$\mathbb{E}\left(X(T)|\mathscr{F}_S\right)\geq X(S\wedge T),\text{ a.s..}\tag{2.6.3}$$

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

$$\mathbb{E}\left(\mathbb{1}_A X(T)\right)\geq\mathbb{E}\left(\mathbb{1}_A X(T\wedge S)\right).\tag{2.6.4}$$

Now we can decompose $A=\cup_{i=1}^m A\cap\{S=t_i\}$. Hence we just have to prove (2.6.4) with A replaced by $A\cap\{S=t_i\}$, for any $i=1,\dots m$. Now, since $A\in\mathscr{F}_S$, we have that $A\cap\{S=t_i\}\in\mathscr{F}_{t_i}$. We will first show that

$$\mathbb{E}\left(X(T)|\mathscr{F}_{t_i}\right)\geq X(T\wedge t_i).\tag{2.6.5}$$

To do this, note that

$$\begin{aligned}\mathbb{E}\left(X(T\wedge t_{k+1})|\mathscr{F}_{t_k}\right)&=\mathbb{E}\left(X(t_{k+1})\mathbb{1}_{T>t_k}+X(T)\mathbb{1}_{T\leq t_k}|\mathscr{F}_{t_k}\right)\\&=\mathbb{E}\left(X(t_{k+1})|\mathscr{F}_{t_k}\right)\mathbb{1}_{T>t_k}+X(T)\mathbb{1}_{T\leq t_k}\\&\leq X(t_{k+1})\mathbb{1}_{T>t_k}+X(T)\mathbb{1}_{T\leq t_k}\\&=X(t_k\wedge T),\text{ a.s..}\end{aligned}\tag{2.6.6}$$

Since $S=S\wedge t_m$, this gives (2.6.5) for $i=m-1$. Then we can iterate (2.6.6) to get (2.6.5) for general i.

Using (2.6.4), we can now deduce that

$$\begin{aligned}\mathbb{E}\left(\mathbb{1}_{A\cap\{S=t_i\}}X(T)\right)&=\mathbb{E}\left(\mathbb{1}_{A\cap\{S=t_i\}}\mathbb{E}(X(T)|\mathscr{F}_{t_i})\right)\\&\geq\mathbb{E}\left(\mathbb{1}_A X(T\wedge t_i)\right)\\&=\mathbb{E}\left(\mathbb{1}_A X(T\wedge S)\right)\end{aligned}\tag{2.6.7}$$

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

$$\mathbb{E}\left(X(T^{(n)} \wedge M) \vee \alpha | \mathscr{F}_{S^{(n)}}\right) \geq X(T^{(n)} \wedge S^{(n)} \wedge M) \vee \alpha, \text{ a.s..} \tag{2.6.8}$$

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

$$\mathbb{E}\left(X(T^{(n)} \wedge M) \vee \alpha | \mathscr{F}_S\right) \geq \mathbb{E}\left(X(T^{(n)} \wedge S^{(n)} \wedge M) \vee \alpha | \mathscr{F}_S\right), \text{ a.s..} \tag{2.6.9}$$

Again from using Lemma 2.31, we get that

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

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 (2.6.9) and obtain, using that X is right-continuous,

$$\mathbb{E}\left(X(T \wedge M) \vee \alpha | \mathscr{F}_S\right) \geq \mathbb{E}\left(X(T \wedge S \wedge M) \vee \alpha | \mathscr{F}_S\right), \text{ a.s..} \tag{2.6.10}$$

Since this relation holds for all α, we may let $\alpha \downarrow -\infty$ to get (2.6.1). Using the additional assumptions on T; we can pass to the limit $M \uparrow \infty$ and get (2.6.2) 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

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

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

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

and

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

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 2.32. *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 2.33. *Let X be a uniformly integrable or a non-negative càdlàg supermartingale. Let S and T be stopping times with $S \leq T$. Then $X_T \in \mathscr{L}^1$ and*

$$\mathbb{E}(X_\infty | \mathscr{F}_T) \leq X_T, \text{ a.s.} \tag{2.6.11}$$

and

$$\mathbb{E}(X_T | \mathscr{F}_S)) \leq X_S, \text{ a.s.,} \tag{2.6.12}$$

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 [15], Thms (59.1,59.5)).

Chapter 3
Markov processes in continuous time

In this chapter we develop the theory of Markov processes in continuous time with general state space. We would expect that much that is true in discrete time carries over, 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. You have already seen this in the case of Brownian motion, and 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. Markov processes in continuous time are define analogously to those in discrete time. The following definition is provisional.

Definition 3.1. A stochastic process X with state space S and index set $\mathbb{R}_+$ is a continuous time Markov process with stationary transition kernel P_t if, for all $A \in \mathscr{B}$, $t \in \mathbb{N}$,

$$\mathbb{P}(X_{t+s} \in A|\mathscr{F}_t)(\omega) = P_s(X_t(\omega), A), \mathbb{P} - \text{a.s.}. \tag{3.0.1}$$

Here $\{F_t\}_{t\in\mathbb{N}_0}$ denotes the σ-algebra generated by the random variables $X_0, \ldots, X_t$.

The specific requirements on transition kernels will be discussed in detail below.

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 $C_0(S) \subset C_b(S) \subset C(S) \subset B(S)$.

3.1 Markov jump processes

The simplest 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, $Y_n \in S$, $n \in \mathbb{N}$, be some discrete time Markov process with transition kernel P and initial distribution μ. Let $m(x) : S \to \mathbb{R}_+$ be a uniformly bounded, measurable function. To avoid complications, we will assume that $0 < \inf_{x\in S} m(x) \leq \sup_{x\in S} m(x) < \infty$. Let e_i, $i \in \mathbb{N}$, be a family of independent exponential random variables with mean 1, defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ as Y_n, and let Y_n and the e_x be mutually independent. Then define the process

$$S(n) \equiv \sum_{i=0}^{n-1} e_i m(Y_i). \tag{3.1.1}$$

$S(n)$ 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

$$S^{-1}(t) \equiv \sup\{n : S(n) \leq t\}. \tag{3.1.2}$$

Then set

$$X(t) \equiv Y_{S^{-1}(t)}. \tag{3.1.3}$$

Theorem 3.2. *The process $X(t)$ defined through (3.1.3) is a continuous time Markov process with càdlàg paths.*

Proof. We can express what we would expect to play the role of a transition kernel as follows:

$$P_t(x,A) \equiv \mathbb{P}_x(X_t \in A) = \sum_{n=0}^{\infty} \mathbb{P}_x(Y_n \in A, S(n) \leq t < S(n+1)). \tag{3.1.4}$$

This is just saying that the event $X_t \in A$ can be realized by the process making exactly n jumps before time t and the jump-chain Y being in A at discrete time n. Now let us consider

$$\mathbb{P}_x(X_{t+s} \in A|\mathscr{F}_t) \tag{3.1.5}$$

It is clear that $\mathscr{F}_t$ contains the information on what X_t is and on when the last (say the k-th) jump before t occurred, say at time $t-r$. Since that time, $Y_k = X_t$. Since Y is Markov, the event $\{X_{t+s} \in A\}$ can only depend on this information, and is in fact given by

$$\mathbb{P}_x(X_{t+s} \in A|\mathscr{F}_t) = \sum_{n=0}^{\infty} \mathbb{P}_{X_t}(Y_n \in A, S(n) \leq t+r+s < S(n+1)|S(1) > r). \tag{3.1.6}$$

But due to the fact that the random variable e_0 is exponentially distributed,

$$\mathbb{P}(e_1 m(X_t) - r \geq a | e_1 m(X_t) - r \geq 0) = \mathbb{P}(e_1 m(X_t) \geq a), \tag{3.1.7}$$

so that

$$\begin{aligned} &\mathbb{P}_{X_t}(Y_n \in A, S(n) \leq t + r + s < S(n+1) | S(1) > r) \\ &= \mathbb{P}_{X_t}(Y_n \in A, S(n) \leq t + s < S(n+1)) \end{aligned} \tag{3.1.8}$$

so that indeed the conditional probability depends only on X_t, proving that X is a Markov process. The fact that X has càdlàg paths is obvious from the construction. □

It is clear from the construction that the transition probability kernel P and the function m determine the transition kernels P_t completely. We will now make this connection more explicit. kernel. First we observe that

$$\lim_{t \downarrow 0} P_t(x, A) = \mathbb{1}_{x \in A}. \tag{3.1.9}$$

This follows simply from the fact that

$$\begin{aligned} \mathbb{P}_x(Y_n \in A, S(n) \leq t < S(n+1)) \leq \mathbb{P}[S_n \leq t] \leq \mathbb{P}\left[\bar{m} \sum_{i=0}^{n-1} e_i \leq t\right] \\ = \sum_{k=n}^{\infty} \frac{(t/\bar{m})^k}{k} e^{-t/\bar{m}}, \end{aligned} \tag{3.1.10}$$

where $\bar{m} \equiv \inf_{x \in S} m(x)$. Similarly we see that

$$\begin{aligned} &\lim_{t \downarrow 0} t^{-1} (\mathbb{P}_x(X_t \in A) - \mathbb{1}_{x \in A}) \\ &= \lim_{t \downarrow 0} t^{-1} (\mathbb{1}_{x \in A}(\mathbb{P}(m(x)e_1 > t) - 1) + \mathbb{P}_x(Y_1 \in A, S(1) \leq t)) \\ &= (P(x,A) - \mathbb{1}_{x \in A}) \lim_{t \downarrow 0} t^{-1} \mathbb{P}(m(x)e_1 \leq t) \\ &= (P(x,A) - \mathbb{1}_{x \in A}) \lim_{t \downarrow 0} t^{-1}(1 - e^{-t/m(x)}) = \frac{1}{m(x)} L(x,A). \end{aligned} \tag{3.1.11}$$

We will denote the right-hand side of (3.1.11) by G and call it the *generator* of the Markov jump process X. By the Markov property, it follows that we get a more general result:

Lemma 3.3. *For any $t \geq 0$,*

$$\frac{d}{dt} P_t(x,A) = (P_t G)(x,A) = (G P_t)(x,A). \tag{3.1.12}$$

Proof. Using the Markov property we get the *Chapman-Kolmogorov equation*,

$$\mathbb{P}_x(X_{t+h} \in A) = \mathbb{P}_x(\mathbb{P}_x(X_{t+h} \in A | \mathscr{F}_t)) = \int_S P_h(y,A) P_t(x,dy). \tag{3.1.13}$$

This implies that

$$\lim_{h\downarrow 0} h^{-1}\left(\mathbb{P}_{t+h}(x,A)-P_t(x,A)\right) = \int_S \lim_{h\downarrow 0} h^{-1}\left(P_h(y,A)-\mathbb{1}_{y\in A}\right)P_t(x,dy) \tag{3.1.14}$$
$$= \int_S G(y,A)P_t(x,dy)m(y)P(y,A) \equiv (P_tG)(x,A).$$

Alternatively, we can write

$$\lim_{h\downarrow 0} h^{-1}\left(\mathbb{P}_{t+h}(x,A)-P_t(x,A)\right) = \int_S \lim_{h\downarrow 0} h^{-1}P_t(y,A)\left(P_h(x,dy)-\mathbb{1}_{x\in dy}\right) \tag{3.1.15}$$
$$= \int_S P_t(y,A)G(x,dy) \equiv (GP_t)(x,A).$$

This proves the lemma. □

We can view Eq. (3.1.12) as a differential equation for P_t,

$$\frac{d}{dt}P_t(x,A) = GP_t(x,A), \tag{3.1.16}$$

which has the solution

$$P_t = \exp(tG) \equiv \sum_{n=0}^{\infty}\frac{t^n}{n!}G^n, \tag{3.1.17}$$

where G^n is defined as the n-fold application of G from the right. This can be made rigorous if we think of P as an operator acting on bounded measurable functions. G is a bounded operator on this space,

$$\|Gf\|_\infty = \sup_{x\in S}\left|\int_S \frac{1}{m(x)}P(x,dy)f(y)\right| \leq \|f\|_\infty, \tag{3.1.18}$$

so

$$\|L\| \leq \|1/m\|_\infty < \infty, \tag{3.1.19}$$

where the last inequality holds by assumption. Then the series $\sum_{n=0}^{\infty}\frac{t^n}{n!}G^n$ is absolutely convergent in norm and defines a bounded operator, $\exp(tG)$. This operator solves the differential equation, which has a unique solution with initial condition $P_0 = \mathbb{1}$.

So for Markov jump processes we have a nice picture: the process is uniquely determined by the initial condition and a single operator G, the *generator* of the process. The transition kernel is given by $\exp(tG)$.

The bad news is that this construction relied on the boundedness of the operator G, which in turn relied on the fact that the jump rates, m, where uniformly bounded. Many Markov processes do not fall into this class: Brownian motion, Lévy jump processes with infinite Lévy measure, etc.. In the next sections we will investigate what can be salvaged from this nice picture in the general case.

3.2 Semi-groups, resolvents, generators

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

3.2.1 Transition functions and semi-groups

We now give the precise definition of continuous time Markov processes. In the sequel we will always assume that we are dealing with stochastic processes on càdlàg spaces that satisfy the *usual assumptions* (see Definition 2.17. In particular, all filtrations are assumed to be right-continuous.

Definition 3.4. A *Markov transition function*, P_t is a family of kernels $P_t : S \times \mathscr{B}(S) \to [0,1]$ with the following properties:

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

$$P_{s+t}(x,A)) = \int P_t(y,A)P_s(x,dy). \tag{3.2.1}$$

We can now make the definition of continuous time Markov processes more precise.

Definition 3.5. A stochastic process X with state space S and index set $\mathbb{R}$ is a continuous time homogeneous Markov process with law $\mathscr{P}$ on a filtered space $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_t,t \in \mathbb{R}_+))$ with transition function P_t, if it is adapted to $\mathscr{F}_t$ and, for all bounded $\mathscr{B}$-measurable functions f, $t,s \in \mathbb{R}_+$,

$$\mathbb{E}\left[f(X_{t+s})|\mathscr{F}_s\right](\omega) = (P_t f)(X_s(\omega)), \text{ a.s.}. \tag{3.2.2}$$

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

$$(P_t f)(x) \equiv \int_S P_t(x,dy) f(y). \tag{3.2.3}$$

The Chapman-Kolmogorov equations (iii) then take the simple form $P_sP_t = P_{t+s}$. 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

$$(\mu P_t)(A) \equiv \int_S \mu(dx)P_t(x,A). \tag{3.2.4}$$

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 3.6. 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 ∂. 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^∂. To be precise, we will set

(i) $P_t^\partial(x,A) \equiv P_t(x,A)$, for $x \in S, A \in \mathscr{B}(S)$,
(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^∂-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. (3.2.1) allows us to think of Markov kernels as operators on the Banach space of bounded measurable functions.

Definition 3.7. 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 [1]) 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 $A \in \mathscr{B}$, is $\mathscr{B}(S) \times \mathscr{B}(\mathbb{R}_+)$-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_λ, by

$$(R_\lambda f)(x) \equiv \int_0^\infty e^{-\lambda t}(P_t f)(x)dt = \int_S R_\lambda(x,dy) f(y), \tag{3.2.5}$$

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

$$R_\lambda(x,A) \equiv \int_0^\infty e^{-\lambda t} P_t(x,A)dt. \tag{3.2.6}$$

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

(i) For all $\lambda > 0$, R_λ 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 $\lambda, \mu > 0$,

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu; \tag{3.2.7}$$

(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. (3.2.7) 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. To motivate the following, let us look at this in the case of jump processes, i.e. when the generator is a bounded operator. In this case we search an operator G such that $P_t = \exp(tG)$. Then, formally, we see that

$$R_\lambda = \int_0^\infty e^{-\lambda t} e^{Gt} dt = \frac{1}{\lambda - G}. \tag{3.2.8}$$

This should make sense, because e^{Gt} is bounded (by one), so that the integral converges at infinity for any $\lambda > 0$.

Finally, we can recover G from R_λ: 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_λ is bounded, and we will see that (under certain circumstances, $\exp(tG_\lambda) \to P_t$, as $\lambda \uparrow \infty$.

3.2.2 Strongly continuous contraction semi-groups

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

Definition 3.8. 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 3.9. *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 of such an R_λ motivate the *Definition* of a strongly continuous contraction resolvent (SCCR).

Definition 3.10. Let B be a Banach space, and let R_λ, $\lambda > 0$, be a family of bounded linear operators on B. Then R_λ is called a *contraction resolvent*, if

(i) $\lambda \|R_\lambda\| \leq 1$, for all $\lambda > 0$;
(ii)the resolvent identify (3.2.7) 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 3.11. *Let R_λ be a contraction resolvent on B_0. Then the the range of R_λ is independent of λ, and 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_μ, then it is also in the range of R_λ: 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_λ by $\mathscr{R}$.

Moreover, if $h \in \mathscr{R}$, 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 λR_λ is bounded, it follows that the right-hand side tends to zero, as $\lambda \uparrow \infty$. Also, if h is in the closure of $\mathscr{R}$, the there exist $h_n \in \mathscr{R}$, 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 λR_λ is a contraction, the right hand side can be made as small as desired by letting n and λ tend to infinity. Finally, it is clear that if $h = \lim_{\lambda\uparrow\infty} \lambda R_\lambda h$, then h must be in the closure of $\mathscr{R}$. □

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_λ is dense in B_0.

We now come to the definition of an infinitesimal generator.

Definition 3.12. 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, $\mathscr{D}(G)$, if there exists a function $g \in B_0$, such that

$$\lim_{t\downarrow 0} \|t^{-1}(P_t f - f) - g\| = 0. \tag{3.2.9}$$

For such f we set $Gf = g$ if g is the function that satisfies (3.2.9).

Remark 3.13. 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 (3.2.9).

The crucial fact is that the resolvent is related to the generator in the way anticipated in (3.2.8).

Lemma 3.14. *Let P_t be a strongly continuous contraction semigroup on B_0. Then the operators R_λ and $(\lambda - G)$ are inverses.*

Proof. Let $g \in B_0$ and let $f = R_\lambda g$. We want to show first that $(\lambda - G)f = g$, i.e. that if f is in the range of $\mathscr{R}_\lambda$, then it is in the domain of G and $Gf = \lambda f + g$. 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

$$e^{-\lambda t} P_t R_\lambda g = \int_0^\infty e^{-\lambda(t+s} P_{t+s} g ds = \int_t^\infty e^{-\lambda s} P_s g ds,$$

and so

$$t^{-1}(R_\lambda g - e^{-\lambda t} P_t R_\lambda g) = t^{-1} \int_0^t e^{-\lambda s} P_s g ds.$$

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 $R_\lambda g \in \mathscr{D}(G)$.

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

$$\frac{d}{dt} P_t f = P_t G f.$$

Integrating this relation gives that

$$P_t f - f = \int_0^t P_s G f ds.$$

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

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

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

3.2.3 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 3.15. *Let R_λ 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 $f \in B_0$,*

$$\int_0^\infty e^{-\lambda t} P_t f dt = R_\lambda f. \tag{3.2.10}$$

Moreover, if

$$G_\lambda \equiv \lambda(\lambda R_\lambda - 1) \tag{3.2.11}$$

and

$$P_{t,\lambda} \equiv \exp(tG_\lambda), \tag{3.2.12}$$

then

$$P_t f = \lim_{\lambda \uparrow \infty} P_{t,\lambda} f. \tag{3.2.13}$$

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, $\mathscr{R}$, of R_λ is independent of λ and dense in B_0, due to the assumption of strong continuity. Now we want to show that R_λ is a bijection. Note that we cannot use Lemma 3.14 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 μ. 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_λ^{-1}, of R_λ, with domain equal to $\mathscr{R}$, such that for all $h \in B_0$, $R_\lambda^{-1} R_\lambda h = h$, and for $g \in \mathscr{R}$, $R_\lambda R_\lambda^{-1} g = g$. 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

$$R_\mu^{-1} - (\mu - \lambda) = R_\lambda^{-1}, \tag{3.2.14}$$

which we may rewrite as

$$R_\lambda^{-1} - \lambda = R_\mu^{-1} - \mu \equiv -G \tag{3.2.15}$$

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

$$\frac{1}{\lambda - G} = R_\lambda. \tag{3.2.16}$$

We now show the following lemma:

Lemma 3.16. *Let G_λ be defined in (3.2.11). Then, $f \in \mathscr{D}(G)$ if and only if*

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

exists. Then $Gf = g$.

Proof. Let first $f \in \mathscr{D}(G)$. 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, $f \in \mathscr{R}$, and

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

□

We now continue the proof of the theorem. Note that G_λ 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_λ and R_μ commute for all $\lambda, \mu > 0$, and so all derived operators commute. Thus we have the telescopic expansion

$$\begin{aligned} P_{t,\lambda} - P_{t,\mu} &= P_{t,\lambda} P_{0,\mu} - P_{0,\lambda} P_{t,\mu} \\ &= \sum_{k=1}^n \left(P_{kt/n,\lambda} P_{(n-k)t/n,\mu} - P_{(k-1)t/n,\lambda} P_{(n-k+1)t/n,\mu}\right) \\ &= \sum_{k=1}^n P_{(k-1)t/n,\lambda} P_{(n-k)t/n,\mu} \left(P_{t/n,\lambda} - P_{t/n,\mu}\right). \end{aligned} \tag{3.2.17}$$

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

$$\|P_{t,\lambda} f - P_{t,\mu} f\| \leq t \|G_\lambda f - G_\mu f\|. \tag{3.2.18}$$

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 3.16 for all $f \in \mathscr{D}(G)$. Moreover, the convergence is uniform in t on compact sets, so the map $t \to P_t f$ is continuous. Since $\mathscr{D}(G) = \mathscr{R}$ is dense in B_0, and $P_{t,\lambda}$ are uniformly bounded in norm, these results in fact extends to all functions $f \in B_0$. The family P_t inherits all properties of a SCCSG from the properties of $P_{,\lambda}$.

It remains to show that (3.2.10) 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 μ 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 $R_\lambda f$. Namely,

$$\frac{1}{\lambda - G_\mu} = \frac{1}{\lambda + \mu - \mu^2 R_\mu} = \frac{1}{\lambda + \mu} + \frac{\mu^2 R_\mu}{(\lambda + \mu)(\lambda + \mu - \mu^2 R_\mu)}. \tag{3.2.19}$$

The first term converges to zero. For the second, we write

$$\frac{\mu^2 R_\mu}{(\lambda + \mu)(\lambda + \mu - \mu^2 R_\mu)} = \frac{\mu^2}{(\lambda + \mu)^2} \frac{R_\mu}{1 - \frac{\mu^2}{\lambda + \mu} R_\mu}. \tag{3.2.20}$$

For $\gamma > 0$, to be choose later, we continue

$$\begin{aligned} \frac{R_\mu}{1 - \frac{\mu^2}{\lambda+\mu} R_\mu} &= \frac{R_\mu R_\gamma}{R_\gamma - \frac{\mu^2}{\lambda+\mu} R_\mu R_\gamma} \\ &= \frac{R_\mu R_\gamma}{R_\gamma - \frac{\mu^2}{\lambda+\mu} \frac{R_\gamma - R_\mu}{\mu - \gamma}} \\ &= \frac{R_\mu R_\gamma}{R_\gamma \left(1 - \frac{\mu^2}{(\lambda+\mu)(\mu-\gamma)}\right) + \frac{\mu^2}{\lambda+\mu} \frac{R_\mu}{\mu-\gamma}}. \end{aligned} \tag{3.2.21}$$

Now choose γ such that $\frac{\mu^2}{(\lambda+\mu)(\mu-\gamma)} = 1$, that is $\gamma = \lambda\mu/(\lambda+\mu)$. We get

$$\frac{1}{\lambda - G_\mu} = \frac{1}{\lambda + \mu} + \frac{\mu^2}{(\lambda + \mu)^2} R_\gamma. \tag{3.2.22}$$

As $\mu \uparrow \infty$, $\gamma \to \lambda$, and hence $R_\gamma \to R_\lambda$, and so the claim follows. 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_λ, then there exists a unique operator, G, such that $R_\lambda = (\lambda - G)^{-1}$, and a strongly continuous contraction semigroup, P_t, such that R_λ is its resolvent. Then the operator G will in fact have to be the generator of P_t, from what we already know.

One might rightly ask if we can *start* from a generator: of course, the answer is yes: if we have linear operator, G, with $\mathscr{D}(G) \subset B_0$, 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 3.17. *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, $\mathscr{D}(G)$, is dense in B_0.
(ii)G is dissipative, *i.e. for all $\lambda > 0$ and all $f \in \mathscr{D}(G)$,*

$$\|(\lambda - G)f\| \geq \lambda \|f\|. \tag{3.2.23}$$

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

Proof. By theorem 3.15, 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. 3.15, (ii) is a consequence of the bound $\|\lambda R_\lambda\| \leq 1$: Note that

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

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

$$\Gamma(G) \equiv \{(f, Gf) : f \in \mathscr{D}(G)\} \subset B_0 \times B_0 \tag{3.2.24}$$

is closed in the product topology. Equivalently, G is closed if for any sequence $f_n \in \mathscr{D}(G)$ such that $f_n \to f$ and $G f_n \to g$, $f \in \mathscr{D}(G)$ and $g = Gf$.

Lemma 3.19. *If G is the generator of a strongly continuous contraction semi-group on a Banach space B_0, then G is closed.*

Proof. The proof relies on the fact that for any $f_n \in \mathscr{D}(G)$,

$$P_t f_n - f_n = \int_0^t P_s G f_n ds. \tag{3.2.25}$$

Now take a sequence $f_n \in \mathscr{D}(G)$ such that f_n converges to $f \in B_0$, such that $G f_n \to g \in B_0$. Since P_t is bounded, it follows that

$$P_t f - f = \int_0^t P_s g ds. \tag{3.2.26}$$

By the continuity of P_t,

$$\lim_{t \downarrow 0} t^{-1}(P_t f - f) = g, \tag{3.2.27}$$

so $f \in \mathscr{D}(G)$ and $Gf = g$. Thus G is closed. □

Definition 3.20. 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) **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_λ, R_μ 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 λ 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 $g \in \mathscr{D}(G)$, 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 $\widehat{R}_\mu = (\mu - G)^{-1}$, **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 3.21. *Let G be a dissipative operator and let $\lambda > 0$ be fixed. Then G is closed if and only if* **range**$(\lambda - G)$ *is closed.*

Proof. Let us first show that the range of $(\lambda - G)$ is closed if G is closed. Take $f_n \in \mathscr{D}(G)$ 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. Therefore, the sequence f_n has a limit, $f \in B_0$. But then

$$Gf_n = (G - \lambda) f_n + \lambda f_n \to \lambda f - h,$$

so Gf_n converges, and since G is closed, it holds that $f \in \mathscr{D}(G)$ and $Gf = \lambda f - h$, i.e. $(\lambda - G) f = h$, so that $h \in \mathbf{range}(\lambda - G)$. But this means that any sequence in the range of $\lambda - G$ that converges has its limit in $\mathbf{range}(\lambda - G)$, so this range is closed.

On the other hand, if $\mathbf{range}(\lambda - G)$ is closed, then take some $\mathscr{D}(G) \ni f_n \to f$ and $Gf_n \to g$. Then $(\lambda - G) f_n \to \lambda f - g$ in the range of $(\lambda - G)$. Thus there exists $f_0 \in \mathscr{D}(G)$, 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. $\square$

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 3.22. *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 $\rho^+(G)$ 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 $\mathbf{range}(\lambda - G) = B_0$. 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 $\mathscr{D}(G)$. Then dissipativity shows that

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

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

We now continue with the proof of the theorem. We know from (ii) and (iii) and Lemma 3.21 that G is closed and $\mathbf{range}(\lambda - G) = B_0$ for all $\lambda > 0$. Moreover, (iii) asserts that for some $\lambda > 0$, the range of $\lambda - G$ is B_0. This λ is then also in the resolvent set, and so we know by Lemma 3.22 that $\rho^+(G) = (0,\infty)$. In the proof of that lemma we have also shown that $\lambda\|R_\lambda\| \leq 1$. As we have already explained, the resolvent identity holds for all $\lambda > 0$, so R_λ is a contraction resolvent family.

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

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

Since $f \in \mathscr{D}(G)$, $Gf \in B_0$, 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 $\mathscr{D}(G)$. For general f, since $\mathscr{D}(G)$ is dense in B_0, take a sequence $f_n \in \mathscr{D}(G)$ 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 3.17 are just as difficult to verify then those of Theorem 3.15. In particular, it does not seem easy to check whether an operator is dissipative.

The following lemma, however, can be very helpful.

Lemma 3.23. *Let S be a complete metric space. A linear operator, G, on $C_0(S)$ is dissipative, if for any $f \in \mathscr{D}(G)$, 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 $f(y) \geq 0$, so that $f(y)$ is a maximum. For $\lambda > 0$, 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. □

Definition 3.24. A linear operator satisfying the hypothesis of Lemma 3.23 is said to satisfy the *positive maximum principle*.

Examples

We can verify the conditions of Theorem 3.17 in some simple examples.

- Let $S=[0,1]$, $G=\frac{1}{2}\frac{d^2}{dx^2}$, $B_0=C([0,1])$, equipped with the sup-norm, and let $\mathscr{D}(G)=\{f\in C^2([0,1]:f'(0)=f'(1)=0\}$. Since here S is compact, clearly any continuous function takes on its maximum at some point $y\in[0,1]$. If $y\in(0,1)$, then clearly $\frac{1}{2}\frac{d^2}{dx^2}f(y)\leq 0$; if $y=0$, for 0 to be a maximum, 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 $\mathscr{D}(G)$ is dense is clear from the definition. To show that the range of $\lambda-G$ is $B([0,1])$, we must show that the equation

$$\lambda f-\frac{1}{2}f''=g \tag{3.2.28}$$

  with boundary conditions $f'(0)=f'(1)=0$ has a solution in $C^2([0,1])$ for all $g\in C([0,1])$. Such a solution can be written down explicitly. In fact, (we just consider the case $\lambda=1$, which is enough)

$$f(x)=e^{\sqrt{2}x}\int_0^x e^{-\sqrt{2}t}\int_0^t g(s)dsdt+K\cosh(\sqrt{2}x), \tag{3.2.29}$$

  with

$$K=-\frac{\sqrt{2}e^{\sqrt{2}}\int_0^1 e^{-\sqrt{2}t}\int_0^t g(s)dsdt+\int_0^1 g(s)ds}{\sqrt{2}\sinh(\sqrt{2})}.$$

  is easily verified to solve this problem uniquely.

(ii)The same operator as above, but replace $[0,1]$ with $\mathbb{R}$, $B_0=C_0(\mathbb{R})$, and $\mathscr{D}(G)=C_0^2(\mathbb{R})$. We first show that the range of R_λ is contained in $C_b^2(\mathbb{R})$. Let f be given by $f=R_\lambda g$ with $g\in C_0(\mathbb{R})$. R_λ 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

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

Now one can show that

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

and so

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

Hence

$$f'(x) = -\int_{-\infty}^{x} e^{-\sqrt{2\lambda}|x-y|} g(y)dy + \int_{x}^{\infty} e^{-\sqrt{2\lambda}|x-y|} g(y)dy. \tag{3.2.30}$$

Thus, differentiating once more,

$$\begin{aligned} f''(x) &= -2g(x) + \sqrt{2\lambda} \int_{-\infty}^{x} e^{-\sqrt{2\lambda}|x-y|} g(y)dy \\ &\quad + \sqrt{2\lambda} \int_{x}^{\infty} e^{-\sqrt{2\lambda}|x-y|} g(y) \\ &= -2g(x) + 2\lambda f(x). \end{aligned} \tag{3.2.31}$$

Hence $f'' \in C_0(\mathbb{R})$, so $\mathbf{range}(R_\lambda) \subset C_0^2(\mathbb{R})$. Moreover, f solves (3.2.28) and thus $G - \lambda = \Delta/2 - \lambda$ is the inverse of R_λ. Since this operator maps $C_0^2(\mathbb{R})$ into $C_0(\mathbb{R})$, we see that $C_0^2(\mathbb{R}) \subset \mathscr{D}(G)$. Hence $C_0^2(\mathbb{R}) = \mathscr{D}(G)$, $\Delta/2$ 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, Δ with domain $C_b^2(\mathbb{R}^d)$ is not a closed operator in $\mathbb{R}^d$ if $d \geq 2$. Namely, if we get a solution of the equation

$$(\lambda - \Delta/2) f = g$$

with $g \in C_0(\mathbb{R}^d)$, then we only know that $\Delta f \in C_0(\mathbb{R}^d)$, which does not imply that $f \in C_0^2(\mathbb{R}^d)$. 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.

3.3 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", ∂, ("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 3.25. A *Feller-Dynkin semigroup* is a strongly continuous sub-Markov semigroup, P_t, acting on the space $C_0(S)$, that is:

(i) for all $t \geq 0$,

$$P_t : C_0(S) \to C_0(S); \tag{3.3.1}$$

(ii)For all $f \in C_0(S)$ such that $0 \leq f \leq 1$, $0 \leq P_t f \leq 1$;
(iii)For all $t, s \geq 0$, $P_{t+s} = P_t P_s$;
(iv)For all $f \in C_0(S)$, $\lim_{t\downarrow 0} \mathbb{P}_f f - f\| = 0$.

Note that the condition (iii) which is a property of sub-Markov semi-groups (see Definition 3.7 is now added again. This is because we want Feller-Dynkin semi-groups to be associated to Markov processes.

It is very convenient that the sufficient criterion for dissipativity, the *posivitive maximum principle*, also ensures positivity.

Lemma 3.26. *Let G be a linear operator with domain and range in $C_0(S)$ that satisfies the Conditions (i) and (iii) of the Hille-Yosida theorem 3.17 with $B_0 = C_0(S)$, and that satisfies the positive maximum principle (Definition 3.24). Then G is the generator of a Feller-Dynkin semi-group.*

Proof. In view of what we know, we only have to show that the semi-group generated by G maps positive functions to positive functions. Notice first that if $f \in \mathscr{D}(G)$ is a function such that $\inf_{y\in S} f(y) = f(x) < 0$, then

$$\inf_{y\in S}(\lambda - G)f(y) \leq (\lambda - G)f(x) \leq \lambda f(x) < 0. \tag{3.3.2}$$

But that means that $(\lambda - G)f \geq 0$ only if $f \geq 0$, or that R_λ maps positive functions to positive functions. From this it follows easily that the same is true for the semi-group. Next,

$$e^{tG_\lambda} = e^{-\lambda t} e^{t\lambda^2 \mathbb{R}_\lambda} = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(t\lambda^2)^n}{n} R_\lambda^n, \tag{3.3.3}$$

so also this operator maps positive functions to positive functions. Finally we know that $P_t = \lim_{\lambda\uparrow\infty} e^{tG_\lambda}$, and taking the limit preserves the positivity property. □

Remark 3.27. In fact, less is necessary. It is easy to see that if G satisfies the positive maximum principle on a dense subset of its domain, then the conclusions of Lemma 3.26 remain valid. (Exercise!) This is important in applications, since often we can do explicit computations with generators only on such sets.

We can now connect back to Markov transition kernels.

Theorem 3.28. *Let P_t be a Feller-Dynkin semigroup. Then there exists a sub-Markov transition kernel $P_t(x, dy)$, such that for all $f \in C_0(S)$,*

$$(P_t f)(x) = \int_S P_t(x, dy) f(y). \tag{3.3.4}$$

The semigroup can be naturally extended to $B_0(S)$ by the right-hand side of Eq. (3.3.4)

Proof. 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, μ, 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 $\mathscr{D}(G) \subset C_0(S)$. Note that then we have for $f \in \mathscr{D}(G)$ the formula

$$Gf(x) \equiv \lim_{t\downarrow 0} t^{-1}\left(\int_S P_t(x,dy)f(y) - f(x)\right) \tag{3.3.5}$$

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

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

Lemma 3.30. *Let G be a generator of a Feller-Dynkin semigroup and let C be an extension of G. Assume that if $f \in \mathscr{D}(C)$ 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 $R_1 g \in \mathscr{D}(G)$ and thus

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

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

Now let $f\in\mathscr{D}(C)$ 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 3.31. *Let P_t be a Feller-Dynkin semigroup on $C_0(S)$. Then there exists a strong Markov process with values in S^∂ and càdlàg paths and transition kernel P_t.*

Remark 3.32. 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 3.33. *Let $g\in C_0(S)$ and $g\geq 0$. Set $h=R_1g$. Then*

$$0\leq e^{-t}P_th\leq h. \tag{3.3.6}$$

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

Proof. Let us first prove (3.3.6). The lower bound is clear since P_t and hence R_λ map positive function to positive functions. Next

$$\begin{aligned} e^{-s}P_sh=e^{-s}P_sR_1g &= e^{-s}P_s\int_0^\infty e^{-u}P_ugdu \\ &= \int_s^\infty e^{-u}P_ugdu\leq R_1g=h. \end{aligned} \tag{3.3.7}$$

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

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

where of course we used (3.3.6) 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 $h_i=R_1g_i$. The set $\mathscr{H}=\{h_i\}_{i\in\mathbb{N}}$ separates points in S^∂, 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^∂ (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).

3.4 The strong Markov property

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

$$\mathbb{E}[\zeta f(X_{t+s})] = \mathbb{E}[\zeta P_s f(X_t)]. \tag{3.4.1}$$

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 θ_t the shift acting on Ω, via

$$X(\theta_t \omega)_s \equiv (\theta_t X)(\omega)_s \equiv X(\omega)_{s+t}. \tag{3.4.2}$$

We then have the following *strong Markov property*:

Theorem 3.34. *Let T be a $\mathscr{F}_{t+}$stopping time, and let $\mathbb{P}$ be the law of a Feller-Dynkin Markov process, X. Then, for all bounded random variables η, if T is a stopping time, then*

$$\mathbb{E}[\theta_T \eta | \mathscr{F}_{T+}] = \mathbb{E}_{X_T}[\eta], \tag{3.4.3}$$

or equivalently, for all $\mathscr{F}_{T_+}$-measurable bounded random variables ξ,

$$\mathbb{E}[\xi \theta_T \eta] = \mathbb{E}[\xi \mathbb{E}_{X_T}[\eta]], \tag{3.4.4}$$

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

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

Let f be a continuous function on S. Then

$$\begin{aligned} \mathbb{E}\left[f(X_{T^{(n)}+s})\mathbb{1}_{\Lambda}\right] &= \sum_{k\in\mathbb{N}\cup\{+\infty\}} \mathbb{E}\left[f(X_{k2^{-n}+s})\mathbb{1}_{\Lambda_{n,k}}\right] \\ &= \sum_{k\in\mathbb{N}\cup\{+\infty\}} \mathbb{E}\left[P_s f(X_{k2^{-n}})\mathbb{1}_{\Lambda_{n,k}}\right] \\ &= \mathbb{E}\left[P_s f(X_{T^{(n)}})\mathbb{1}_{\Lambda}\right] \end{aligned} \tag{3.4.5}$$

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,

$$\mathbb{E}[f(X_{T+s})\mathbb{1}_{\Lambda}] = \mathbb{E}[P_s f(X_T \mathbb{1}_{\Lambda}]$$

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 $\mathscr{F}_{T+}$-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. □

3.5 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 3.35. *A dissipative linear operator, G, on B_0 whose domain, $\mathscr{D}(G)$, 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 $f_n \in \mathscr{D}(G)$ 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 $g_n \in \mathscr{D}(G)$ such that $g_n \to g$. Such a sequence exists because $\mathscr{D}(G)$ 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 λ and taking the limit $\lambda \uparrow \infty$ implies that

$$\|g_n\| \leq \|g_n - g\|.$$

Since $g_n - g \to 0$, this implies $g_n \to 0$ and hence $g = 0$.

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 3.36. Let G be a linear operator on a Banach space B_0. A subspace $D \subset \mathscr{D}(G)$ is called a *core* for G, if the closure of the restriction of G to D is equal to G.

Lemma 3.37. *Let G be the generator of a strongly continuous contraction semigroup on B_0. Then a subspace $D \subset \mathscr{D}(G)$ 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 3.38. *Let G be the generator of a strongly continuous contraction semigroup, P_t, on B_0. Let D be a dense subset of $\mathscr{D}(G)$. 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,

$$\begin{aligned} \lim_{n\uparrow\infty} (\lambda - G) f_n &= \lim_{n\uparrow\infty} \frac{1}{n} \sum_{k=0}^{n^2} e^{-\lambda k/n} P_{k/n} (\lambda - G) f \\ &= \int_0^\infty e^{-\lambda t} P_t (\lambda - G) f \\ &= R_\lambda (\lambda - G) f = f \end{aligned} \tag{3.5.1}$$

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^∞ is a core, since obviously C^∞ is dense in the space of continuous functions, by the preceding lemma we need only to show that P_t maps C^∞ to C^∞. 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^∞ is $\frac{1}{2}\Delta$, which is a simple calculation (we essentially did that in [1]). 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 **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 3.39. *Let X be a Feller-Dynkin process with transition function P_t and generator G. Define, for $f, g \in B(S)$,*

$$M_t \equiv f(X_t) - \int_0^t g(X_s)ds. \tag{3.5.2}$$

Then, if $f \in \mathscr{D}(G)$ and $g = Gf$, M_t is a $\mathscr{F}_t$-martingale.

Proof. The proof goes exactly as in the discrete time case.

$$\begin{aligned}
\mathbb{E}[M_{t+u}|\mathscr{F}_t] &= \mathbb{E}[f(X_{t+u})|\mathscr{F}_t] - \int_0^t (Gf)(X_s)ds - \int_t^{t+u} \mathbb{E}[Gf(X_s)|\mathscr{F}_t]ds \qquad (3.5.3)\\
&= (P_u f)(X_t) - \int_0^t (Gf)(X_s)ds - \int_0^u (P_s Gf)(X_t)ds\\
&= f(X_t) - \int_0^t (Gf)(X_s)ds\\
&\quad +(P_u f)(X_t) - f(X_t) - \int_0^u (P_s Gf)(X_t)ds\\
&= M_t + (P_u f)(X_t) - f(X_t) - \int_0^u (P_s Gf)(X_t)ds.
\end{aligned}$$

But

$$(P_s Gf)(z) = \frac{d}{ds}(P_s f)(z),$$

and so

$$(P_u f)(X_t) - f(X_t) - \int_0^u (P_s Gf)(X_t)ds = 0,$$

from which the claim follows. □

By "the martingale problem" we will consider the inverse problem associated to this observation.

Definition 3.40. Given a linear operator G with domain $\mathscr{D}(G)$, a S-valued process defined on a filtered càdlàg space $(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t, t \in \mathbb{R}_+))$, is called a solution of the martingale problem associated to the operator G, if for any $f \in \mathscr{D}(G)$, M_t defined by (3.5.2) is a $\mathscr{F}_t$-martingale.

Remark 3.41. 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 3.42. 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 3.43. *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, since the fact that $\sup_n \| f_n\| < \infty$ and $\sup_n \|G f_n\| < \infty$ allows to use dominated convergence. □

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 3.44. *Let G_1 be an operator with $\mathscr{D}(G_1)$ and* **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 3.45. *Let $\mathscr{F}_t$ be a filtration and X an adapted process. Let $f,g \in B(S)$. Then, for $\lambda \in \mathbb{R}$, (3.5.2) is a martingale if and only if*

$$e^{-\lambda t} f(X_t) + \int_0^t e^{-\lambda s} (\lambda f(X_s) - g(X_s))\, ds \tag{3.5.4}$$

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 (3.5.4) should be a martingale whenever (3.5.2) is, and vice versa. □

Lemma 3.46. *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 $f \in \mathscr{D}(G)$ and $g = Gf$. Now use that (3.5.4) is a martingale with $\lambda > 0$. Taking expectations and sending t to infinity gives thus

$$f(X_0) = f(x) = \mathbb{E}\left[\int_0^\infty e^{-\lambda s}(\lambda f(X_s) - g(X_s))\,ds\right]$$

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 3.47. *Assume that S is separable, and that $\mathscr{D}(G) \subset C_b(S)$. Suppose moreover that $\mathscr{D}(G)$ 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*

$$\mathbb{P}(\forall t \in [0,T] \cap \mathbb{Q} : X_t \in K_{\varepsilon,T}) > 1 - \varepsilon, \tag{3.5.5}$$

then X has càdlàg modification.

Proof. By assumption there exists a sequence $f_i \in \mathscr{D}(G)$ 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 (3.5.5), the set $\{X_t(\omega), t \in [0,T]\}$ has compact closure for almost all ω for all T. Let Ω' 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 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

$$f_i(Y_t) = \mathbb{E} f_i(Y_t)|\mathscr{F}_t]\mathbb{E} f_i(Y_t)|\mathscr{F}_t] = \lim_{s \downarrow t} \mathbb{E}(f_i(X_s)|\mathscr{F}_t) = f_i(X_t), \text{ a.s.}$$

by the fact that $M_t^{(i)}$ is a martingale. $\square$

3.5.1 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, μ_0, is given. The data for the martingale problem is thus a pair (G,μ), where G is a linear operator with its domain $\mathscr{D}(G)$ and μ is a probability measure on S.

The following first result is not terribly surprising.

Theorem 3.48. *Let S be separable and let G be a linear dissipative operator on $B(S)$ with $\mathscr{D}(G) \subset B(S)$. Suppose there exists G' with $\mathscr{D}(G') \subset \mathscr{D}(G)$ such that G is an extension of G'. Let $\overline{\mathscr{D}(G')} = \overline{\mathbf{range}(\lambda - G')} \equiv L$, and let L be separating. Let X be a solution for the martingale problem for (G,μ). Then X is a Markov process whose semigroup on L is generated by the closure of G', and the martingale problem for (G,μ) 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 $f \in \mathscr{D}(G')$ and $\lambda > 0$. Then, by Lemma 3.45,

$$e^{-\lambda t} f(X_t) + \int_0^t e^{-\lambda s}(\lambda f(X_s) - G' f(X_s))ds$$

is a martingale,

$$f(X_t) = \mathbb{E}\left[\int_0^\infty e^{-\lambda s}\left(\lambda f(X_{t+s} - G' f(X_{t+s}))\right) ds \middle| \mathscr{F}_t\right]. \tag{3.5.6}$$

To see this note that for any $T > 0$, by simple algebra,

$$\begin{aligned}
&\int_0^T e^{-\lambda s}\left(\lambda f(X_{t+s} - G' f(X_{t+s}))\right) ds \qquad (3.5.7)\\
&= e^{\lambda t}\int_0^{t+T} e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds - e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds\\
&= e^{\lambda t}\left[\int_0^{t+T} e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds + e^{-(t+T} f(X_{t+T})\right] - e^{-T\lambda} f(X_{t+T})\\
&-e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds
\end{aligned}$$

Hence,

$$\mathbb{E}\left[\int_0^T e^{-\lambda s}\left(\lambda f(X_{t+s}) - G' f(X_{t+s}))\right) ds \middle| \mathscr{F}_t\right] \tag{3.5.8}$$
$$= f(X_t) + e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds$$
$$-e^{-\lambda T}\mathbb{E}\left[f(X_{t+T})\middle|\mathscr{F}_t\right] - e^{\lambda t}\int_0^t e^{-\lambda s}\left(\lambda f(X_s) - G' f(X_s))\right) ds$$
$$= f(X_t) - e^{-\lambda T}\mathbb{E}\left[f(X_{t+T})\middle|\mathscr{F}_t\right].$$

Letting T tend to infinity, we get (3.5.6).

We will use the following lemma.

Lemma 3.49. *Let P_t be a SCCSG on B_0 and G its generator. Then, for any $f \in B_0$,*

$$\lim_{n\uparrow\infty}(1 - n^{-1}G)^{-[nt]} f = P_t f. \tag{3.5.9}$$

Proof. Set $V(t) \equiv (1 - tG]^{-1}$. We want to show that $V(1/n)^[tn] \to P_t$. But

$$n[V(1/n)f - f] = n\left[(1 - n^{-1}G)^{-1} f - f\right] = n\left[(n - G)^{-1} f - f\right] = G_n f,$$

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 $B = \frac{1}{n}G_n + 1$ (check that this is a contraction since G is dissipative). 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 (3.5.6) with $\lambda = n$ and $f \in L$,

$$(1 - n^{-1}G')^{-1} f(X_t) = n\frac{1}{n - G'} f(X_t) \tag{3.5.10}$$
$$= \mathbb{E}\left[n\int_0^\infty e^{-ns} f(X_{t+s}) ds \middle| \mathscr{F}_t\right]$$
$$= \mathbb{E}\left[\int_0^\infty e^{-s} f(X_{t+n^{-1}s}) ds \middle| \mathscr{F}_t\right]$$

Iterating this formula and re-arranging the resulting multiple integrals, and using the formula for the area of the k-dimensional simplex, gives

$$\begin{aligned}
&(1-n^{-1}G')^{-[nu]}f(X_t) \qquad (3.5.11)\\
&= \mathbb{E}\left[\int_0^\infty e^{-s_1-s_2\cdots-s_{[un]}} f(X_{t+n^{-1}(s_1+\cdots+s_{[un]})}) ds_1 \dots ds_{[un]} \Big| \mathscr{F}_t\right]\\
&= \mathbb{E}\left[\int_0^\infty e^{-s} \frac{s^{[un]-1}}{\Gamma([un])} f(X_{t+n^{-1}s}) ds \Big| \mathscr{F}_t\right]
\end{aligned}$$

We write, for $f \in \mathscr{D}(G')$,

$$\mathbb{E}\left[f(X_{t+n^{-1}s})|\mathscr{F}_t\right] = \mathbb{E}\left[f(X_{t+u})|\mathscr{F}_t\right] + \mathbb{E}\left[\int_u^{s/n} G'f(X_{t+v})dv \Big| \mathscr{F}_t\right]$$

and insert this into (3.5.11). Finally, since

$$\int_0^\infty e^{-s} \frac{s^{[un]-1}}{\Gamma([un])} ds = 1,$$

we arrive at

$$\begin{aligned}
(1-n^{-1}G')^{-[nu]}f(X_t) &= \mathbb{E}\left[f(X_{t+u})\Big|\mathscr{F}_t\right] \qquad (3.5.12)\\
&\quad + \mathbb{E}\left[\int_0^\infty e^{-s} \frac{s^{[un]-1}}{\Gamma([un])} \int_u^{s/n} G'f(X_{t+v)})dvds \Big| \mathscr{F}_t\right]
\end{aligned}$$

We are finished if the second term tends to zero. But, re-expressing the volume of the sphere through multiply integrals, we see that

$$\begin{aligned}
&\left|\mathbb{E}\left[\int_0^\infty e^{-s} \frac{s^{[un]-1}}{\Gamma([un])} \int_u^{s/n} G'f(X_{t+v)})dvds \Big| \mathscr{F}_t\right]\right| \qquad (3.5.13)\\
&\leq \|G'f\|_\infty \int_0^\infty ds_1 \dots ds_{[un]} \left|n^{-1}(s_1+\cdots+s_{[un]})-u\right| e^{-s_1-\cdots-s_{[un]}}
\end{aligned}$$

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

$$P_u f(X_t) = \mathbb{E}[f(X_{t+u})|\mathscr{F}_t]$$

for all $f \in \mathscr{D}(G')$. In the usual way, this relation extends to the closure of $\mathscr{D}(G')$ which by assumption is L. □

Finally we establish an important uniqueness criterion and the strong Markov property for solutions of uniquely posed martingale problems.

Theorem 3.50. *Let S be a separable space and let G be a linear operator on $B(S)$. Suppose that the for any initial distribution, μ, any two solutions, X, Y, of the martingale problem for (G, μ) 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 $\mathscr{D}(G) \subset C_b(S)$ and X is a solution of the martingale problem with càdlàg sample paths, then for any a.s. finite stopping time, τ,

$$\mathbb{E}[f(X_{t+\tau})|\mathscr{F}_\tau] = \mathbb{E}[f(X_{t+\tau})|X_\tau], \tag{3.5.14}$$

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 δ_x, $x \in S$, then the strong Markov property holds, i.e.

$$\mathbb{E}[f(X_{t+\tau})|\mathscr{F}_\tau] = P_t f(X_\tau). \tag{3.5.15}$$

Proof. Let X be the solution of the martingale problem with respect to some filtration $\mathscr{G}_t$. We want to prove that it is a Markov process. Let $F \in \mathscr{G}_r$ have positive probability. The, for any measurable set B let

$$P_1(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[\mathbb{1}_B|\mathscr{G}_r]\right]}{\mathbb{P}(F)} \tag{3.5.16}$$

and

$$P_2(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[\mathbb{1}_B|X_r]\right]}{\mathbb{P}(F)}. \tag{3.5.17}$$

Let $Y_s \equiv X_{r+s}$. We see that, since $\mathbb{E}[f(X_r)|X_r] = f(X_r) = \mathbb{E}[f(X_r)|G_r]$,

$$P_1(Y_0 \in \Gamma) = P_2(Y_0 \in \Gamma) = \mathbb{P}[X_r \in \Gamma|F] \tag{3.5.18}$$

Now chose any $0 \leq t_1 < t_2 < \cdots < t_{n+1}$, $f \in \mathscr{D}(G)$, $g = Gf$, and $h_k \in B(S)$, ($k \in \mathbb{N}$. Define

$$\eta(Y) \equiv \left(f(Y_{t_{n+1}}) - f(Y_{t_n}) - \int_{t_n}^{t_{n+1}} g(Y_s)ds \right) \prod_{k=1}^{n} h_k(Y_{t_k}). \tag{3.5.19}$$

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 $\mathbb{E}\left[\eta(X_{r+\cdot})|\mathscr{G}_r\right] = 0$, 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 (3.5.18),

$$E_1[f(Y_t)] = E_2[f(Y_t)],$$

for any bounded measurable function. Thus, for any $F \in \mathscr{G}_r$,

$$\mathbb{E}\left[\mathbb{1}_F \mathbb{E}[f(X_{r+s})|\mathscr{G}_r]\right] = \mathbb{E}\left[\mathbb{1}_F \mathbb{E}[f(X_{r+s})|X_r]\right],$$

and hence

$$\mathbb{E}[f(X_{r+s})|\mathscr{G}_r] = \mathbb{E}[f(X_{r+s})|X_r].$$

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,μ). We want to show that

$$\mathbb{E}\left[\prod_{k=1}^{n} h_k(X_{t_k})\right] = \mathbb{E}\left[\prod_{k=1}^{n} h_k(Y_{t_k})\right]. \tag{3.5.20}$$

By hypothesis, this holds for $n = 1$, so we will proceed by induction, assuming (3.5.20) for all $m \leq n$. For with we define two new measures

$$\widetilde{P}(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_B \prod_{k=1}^{n} h_k(X_{t_k})\right]}{\mathbb{E}\left[\prod_{k=1}^{n} h_k(X_{t_k})\right]}, \tag{3.5.21}$$

$$\widetilde{Q}(B) \equiv \frac{\mathbb{E}\left[\mathbb{1}_B \prod_{k=1}^{n} h_k(Y_{t_k})\right]}{\mathbb{E}\left[\prod_{k=1}^{n} h_k(Y_{t_k})\right]}. \tag{3.5.22}$$

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 τ instead of r, and optional sampling theorem for bounded continuous functions of càdlàg martingales. E.g., to get (ii), note that

$$\mathbb{E}[\eta(X_{\tau+s})|\mathscr{G}_\tau] = 0.$$

For part (iii) we construct the measures P_i replacing r by τ 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 $\mathscr{D}(G)$ (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 3.51. Consider two separable metric spaces (S,ρ) and (E,r). Let G_1,G_2 be two linear operators on $B(S)$, resp. $B(E)$. Let μ,ν 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,μ) and (G_2,ν) are *dual* with respect to (f,α,β), of for any solution, X, of the martingale problem for (G_1,μ) and any solution Y of (G_2,ν), the following hold:

(i) $\int_0^t (|\alpha(X_s)|+|\beta(Y_s)|)ds < \infty$, a.s.,
(ii)

$$\int \mathbb{E}\left[\left| f(X_t,y)\exp\left(\int_0^t \alpha(X_s)ds\right)\right|\right]\nu(dy) < \infty, \tag{3.5.23}$$

$$\int \mathbb{E}\left[\left| f(x,Y_t)\exp\left(\int_0^t \beta(Y_s)ds\right)\right|\right]\mu(dx) < \infty, \tag{3.5.24}$$

(iii)and,

$$\int \mathbb{E}\left[\left| f(X_t,y)\exp\left(\int_0^t \alpha(X_s)ds\right)\right|\right]\nu(dy) \tag{3.5.25}$$
$$= \int \mathbb{E}\left[\left| f(x,Y_t)\exp\left(\int_0^t \beta(Y_s)ds\right)\right|\right]\mu(dx)$$

for any $t \geq 0$.

Proposition 3.52. *With the notation of the definition, let $\mathscr{M}\subset\mathscr{M}_1(S)$ 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,μ) and (G_2,δ_y) are dual with respect to $(f,0,\beta)$ for every μ with compact support and any $y\in E$. Assume further that the set $\{f(\cdot,y) : y\in E\}$ is separating on $\mathscr{M}$. If for every $y\in E$ there exists a solution of the martingale problem (G_2,δ_y), then uniqueness holds for each μ in the martingale problem (G_1,μ).*

Proof. Let X and $\widetilde{X}$ be solutions for the martingale problem for (G_1,μ) where μ has compact support, and let Y^y be a solution to the martingale problem (G_2,δ_y). By duality we have then that

$$\mathbb{E}[f(X_t,y)] = \int \mathbb{E}\left[f(x,Y_t^y)\exp\left(\int_0^t \beta(Y_s^y)ds\right)\right]\mu(dx) = \mathbb{E}[f(\widetilde{X}_t,y)] \qquad (3.5.26)$$

Now we assumed that the class of functions $\{f(\cdot,y) : y\in E\}$ is separating on $\mathscr{M}$, so the one-dimensional marginals of X and $\widetilde{X}$ coincide.

If μ 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 μ is inner regular, the equality of the one dimensional distributions and thus uniqueness by Theorem 3.50. □

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

$$f(X_t,y) - \int_0^t g(X_s,y)ds \qquad (3.5.27)$$

and

$$f(x,Y_t) - \int_0^t h(x,Y_s)ds \qquad (3.5.28)$$

are martingales with respect to the natural filtrations for X, respectively Y. Then (3.5.25) is the integral of

$$\frac{d}{ds}\mathbb{E}\left[f(X_s,Y_{t-s})\exp\left(\int_0^s \alpha(X_u)du + \int_0^{t-s}\beta(Y_u)du\right)\right]. \qquad (3.5.29)$$

Computing (assuming that we can pull the derivative into the expectation) gives that (3.5.29) equals

$$\begin{aligned}\mathbb{E}\ &\Bigg[\Big(g(X_s,Y_{t-s}) - h(X_s,Y_{t-s}) + (\alpha(X_s)-\beta(Y_{t-s}))\,f(X_s,Y_{t-s})\Big)\\ &\times\exp\left(\int_0^s \alpha(X_u)du + \int_0^{t-s}\beta(Y_u)du\right)\Bigg]. \qquad (3.5.30)\end{aligned}$$

This latter quantity is equal to zero, if

$$g(x,y)+\alpha(x)f(x,y)=h(x,y)+\beta(y)f(x,y). \tag{3.5.31}$$

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

$$X_t^y - \int_0^t \left(y(y-1)X_s^{y-2} - yX_s^y\right) ds \tag{3.5.32}$$

are martingales. Of course, this suggest to choose

$$g(x,y) = y(y-1)x^{y-2} - yx^y, \tag{3.5.33}$$

Similarly,

$$x^{Y_t} - \int_0^t Y_s(Y_s-1)\left(x^{Y_s-2} - x^{Y_s}\right) ds \tag{3.5.34}$$

is a martingale and hence

$$h(x,y) = y(y-1)\left(x^{y-2} - x^y\right). \tag{3.5.35}$$

Now we may set $\alpha = 0$ and see that we can satisfy (3.5.31) by putting

$$\beta(y) = y^2 - 2y. \tag{3.5.36}$$

Thus we get

$$\mathbb{E}\left[X_t^{Y_0}\right] = \mathbb{E}\left[X_0^{Y_t} \exp\left(\int_0^t \left(Y_u^2 - 2Y_u\right) du\right)\right]. \tag{3.5.37}$$

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.

In fact, we can use (3.5.37) to compute the moments of the limititing distribution of X_t as $t \uparrow \infty$. Let first $Y_0 = 2k+1$. Then the process Y_t is absorbed in the state $+1$, and hence,

$$\begin{aligned}\lim_{t\uparrow\infty}\mathbb{E}\left[X_t^{2k+1}\right] &= \mathbb{E}\left[X_0^{Y_{\tau_1}} \exp\left(\int_0^{\tau_1} \left(Y_u^2 - 2Y_u\right) du\right)\right] \\ &= \mathbb{E}\left[X_0 \exp\left(\int_0^{\tau_1} \left(Y_u^2 - 2Y_u\right) du\right) \exp\left(-\int_{\tau_1}^{\infty} du\right)\right] = 0.\end{aligned} \tag{3.5.38}$$

The last equality holds since τ_1 is finite almost surely, and even the $\mathbb{E}\exp\left(\int_0^{\tau_1}\left(Y_u^2 - 2Y_u\right)du\right) < \infty$.

If $Y_0 = 2k$, then the process Y is absorbed in the state zero, and we get

$$\lim_{t\uparrow\infty}\mathbb{E}\left[X_t^{2k}\right] = \mathbb{E}\left[\exp\left(\int_0^{\tau_0}\left(Y_u^2 - 2Y_u\right)du\right)\right]. \tag{3.5.39}$$

An elementary computation shows that this expectation equals $\prod_{\ell=1}^{k}(2\ell-1)$. This implies that the limiting distribution is the standard normal distribution.

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.

We still need to verify when the formal computation of the derivative in (3.5.29) can be justified. Basically, we need integrabiliy conditions that justify the use of the Leibnitz rule.

Lemma 3.53. *Let $f:\mathbb{R}_+^2 \to \mathbb{R}$ be absolutely contiuous in each variable for almost all values of the other variable, and assume that the partial derivatives are absolutely integrable, i.e., for all $T>0$,*

$$\int_0^T\int_0^T\left|\frac{\partial}{\partial x_i}f(x_1,x_2)\right|dx_1dx_2 < \infty. \tag{3.5.40}$$

Then, for almost all $t \geq 0$,

$$f(t,0) - f(0,t) = \int_0^t\left(\frac{\partial}{\partial x_1}f(s,t-s) - \frac{\partial}{\partial x_2}f(s,t-s)\right)ds. \tag{3.5.41}$$

Proof. To prove the result, integrate the right-hand side over t and use Fubini's theorem. This yields the integral over the left-hand side. Differentiating give the claim of the lemma.

We see that in order to justify the formula (3.5.37), we have to provide conditions (3.5.40) for the rather complicated fuctions appearing there. The following theorem provides such conditions.

Theorem 3.54. *Let X,Y be independent processes on S_1,S_2. Let f,g,h,α,β be as in Definition 3.51. Assume that for all $T>0$, there exist an integrable random variables V_T and a constant C_T, such that*

$$\sup_{r,s,t\leq T}\left(|\alpha(X_r)|+1\right)|f(X_s,Y_t)| \leq V_T \tag{3.5.42}$$
$$\sup_{r,s,t\leq T}\left(|\beta(X_r)|+1\right)|f(X_s,Y_t)| \leq V_T$$
$$\sup_{r,s,t\leq T}\left(|\alpha(X_r)|+1\right)|g(X_s,Y_t)| \leq V_T$$
$$\sup_{r,s,t\leq T}\left(|\beta(X_r)|+1\right)|g(X_s,Y_t)| \leq V_T,$$

and

$$\int_0^T (|\alpha(X_u)| + |\beta(Y_u)|)\, du \le C_T. \tag{3.5.43}$$

Assume that $f(X_t, y)$ and $f(x, Y_t)$ are martingales as above. Then

$$\mathbb{E}\left[f(X_t, Y_0)\exp\left(\int_0^t \alpha(X_u)du\right)\right] - E\left[f(X_0, Y_t)\exp\left(\int_0^t \beta(Y_u)du\right)\right] \tag{3.5.44}$$

$$\mathbb{E}\Bigg[\Big(g(X_s, Y_{t-s}) - h(X_s, Y_{t-s}) + (\alpha(X_s) - \beta(Y_{t-s}))\, f(X_s, Y_{t-s})\Big)$$

$$\times \exp\left(\int_0^s \alpha(X_u)du + \int_0^{t-s} \beta(Y_u)du\right)\Bigg]. \tag{3.5.45}$$

Proof. The proof of this theorem is fairly technical and can be found in [6].

The tricky part in the use of duality 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.

3.5.2 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 3.55. *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 $f \in \mathscr{D}(A)$, there exists a sequence, $f_n \in \mathscr{D}(G_n)$, such that*

$$\lim_{n\uparrow\infty} \|f_n - f\| = 0, and \lim_{n\uparrow\infty} \|G_n f_n - Gf\| = 0. \tag{3.5.46}$$

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.

Remark 3.56. By weak convergence of processes we understand more precisely the weak convergence of the law of the process defined on the Skorokhod space $D_S[0,\infty)$. See Section 7.

Proof. Let $0 \le t_i \le t < s$ be elements of the set $\mathscr{C}(X) \equiv \{u \in \mathbb{R}_+ : \mathbb{P}[X_u = X_{u-}] = 1\}$. Let $h_i \in C_b(S)$, $i \in \mathbb{N}$. Let f, f_n be as in the hypothesis of the lemma. Then

$$\mathbb{E}\left[\left(f(X_s) - f(X_t) - \int_t^s Gf(X_u)du\right)\prod_{i=1}^k h_i(X_{t_i})\right] \tag{3.5.47}$$
$$= \lim_{n\uparrow\infty}\mathbb{E}\left[\left(f_n(X_s^n) - f_n(X_t^n) - \int_t^s Gf_n(X_u^n)du\right)\prod_{i=1}^k h_i(X_{t_i}^n)\right]$$
$$= 0$$

Here we used that the Now the complement of the set $\mathscr{C}(X)$ is at most countable, and then the relation (3.5.47) 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 3.57. *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 $f \in \mathscr{D}(G)$,*

$$\lim_{n\uparrow\infty} n(T_n - 1)f = Gf. \tag{3.5.48}$$

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 $n(T_n - 1)f = \frac{nG}{n-G}f = T_nGf$ converges to Gf, for $f \in \mathscr{D}(G)$. To do so, we only need to show that $T_nf \to f$. For this let $f \in \mathscr{D}(G)$. Then

$$T_nf = f + (n-G)^{-1}Gf = f + n^{-1}n(n-G)^{-1}Gf. \tag{3.5.49}$$

Since G is dissipative, $\|n(n-G)^{-1}Gf\| \leq \|Gf\| < \infty$, and so the second term tends to zero in norm. Since T_n is bounded, the result extends to the closure of the domain of G. This concludes the proof. □

The point of the lemma is that it shows that the martingale problem for G can be approximated by martingale problems with *bounded* generators $G_n \equiv n(T_n - 1)$ that act like

$$G_nf(x) = n\int (f(y) - f(x))\mu_n(x,dy).$$

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 *Markov jump processes* which we have already encountered in Section 3.1. There we have seen that they can be constructed explicitly as a time change of a discrete time Markov chain. Thus existence is no problem.

Lemma (3.57) the suitably converging generators. To use Lemma (3.55), we only need that the associated processes converge weakly. The standard way to proceed here is to establish *tightness* of the sequences X_n. This can actually be established under the previous hypothesis, by showing that for any $T > 0$ and $\varepsilon > 0$, there exists a compact set $K_{\varepsilon,T} \subset S$, such that

$$\inf_n \mathbb{P}\left(\forall_{0\leq t\leq T} X_n(t) \in K_{\varepsilon,T}\right) \geq 1-\varepsilon. \tag{3.5.50}$$

This uses that the processes X_n are solutions of martingale problems. For details and proofs, see [6], in particular Chapter 3.

Tightness implies the existence of a convergent subsequence whose limit, X, will be a solution of the martingale problem for G. *If* uniqueness can be shown for this martingale problem, then this also implies that there can be only one limit point and hence the sequence X_n converges.

Chapter 4
Stochastic differential equations

4.1 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

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{4.1.1}$$

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

$$X_t = x_0 + \int_0^t b(s, X(s))ds + \int_0^t \sigma(s, X(s))dB_s, \tag{4.1.2}$$

where the integral with respect to B is understood as the Itô stochastic integral defined in the last chapter. The functions b, σ 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.

4.2 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 $\mathscr{H}$ the corresponding Borel-σ-algebra, and by $\mathscr{H}_t \equiv \sigma\{x_s, s \leq t\}$ 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

$$(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}, \xi, B), \tag{4.2.1}$$

where

(i) $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\})$ is a filtered space satisfying the usual conditions;
(ii)B is a Brownian motion (on $\mathbb{R}^d$), adapted to $\mathscr{F}_t$,
(iii)ξ is a $\mathscr{F}_0$-measurable random variable.

The minimal or *canonical* set-up has $\Omega = \mathbb{R}^n \times W$, $\mathbb{P} = \mu \times \mathbb{Q}$, where μ is the law of ξ and $\mathbb{Q}$ is Wiener measure and $\mathscr{F}_t$ the usual augmentation of $\mathscr{F}_t^0 \equiv \sigma\{\xi, B_s, s \leq t\}$.

The precise definition of *path-wise uniqueness* of a SDE is as follows:

Definition 4.1. For a SDE, path-wise uniqueness holds, if the following holds: For any set-up $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}, \xi, B)$, and any two continuous semi-martingales X and X', such that

$$\int_0^t (|b(s, X_s)| + |\sigma(s, X_s)|^2) ds < \infty, \tag{4.2.2}$$

and the same condition for X' hold and both processes solve the SDE with this initial condition ξ and this Brownian motion B,

$$\mathbb{P}[X_t = X_t', \quad \forall t] = 1. \tag{4.2.3}$$

If a SDE admits for any setup $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}, \xi, B)$ 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 4.2. A strong solution of a SDE is a function,

$$F : \mathbb{R}^n \times W \to W, \tag{4.2.4}$$

such that

$$F^{-1}(\mathscr{H}_t) \subset \mathscr{B}(\mathbb{R}^n) \times \overline{\mathscr{H}_t}, \forall t \geq 0, \tag{4.2.5}$$

and on any set-up $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}, \xi, B)$, the process

$$X = F(\xi, B)$$

solves the SDE. $\overline{\mathscr{H}_t}$ is the augmentation of $\mathscr{H}_t$ 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 4.3. *Assume that σ 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.*

$$|\sigma(t,x) - \sigma(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y|, \tag{4.2.6}$$

for all $x,y \in U, t < T$. Let X,Y be two solutions of (4.1.2) (with the same Brownian motion B), and set

$$\tau \equiv \inf\{t \geq 0 : X_t \notin U or Y_t \notin U\}. \tag{4.2.7}$$

Then, if $\mathbb{E}[X_0 - Y_0]^2 = 0$, it follows that

$$\mathbb{P}[X(t\wedge\tau) = Y(t\wedge\tau), \forall 0 \leq t \leq T] = 1. \tag{4.2.8}$$

Proof. The proof is based on Gronwall's lemma and very much like the deterministic analog. We compute

$$\begin{aligned}
&\mathbb{E}\left[\max_{0\leq s\leq t}(X(s\wedge\tau) - Y(s\wedge\tau))^2\right] \qquad (4.2.9)\\
&\leq 2\mathbb{E}\left[\max_{0\leq s\leq t}\left(\int_0^{s\wedge\tau}(\sigma(u,X(u)) - \sigma(u,Y(u)))dB_u\right)^2\right]\\
&+2\mathbb{E}\left[\max_{0\leq s\leq t}\left(\int_0^{s\wedge\tau}(b(u,X(u)) - b(u,Y(u)))du\right)^2\right]\\
&\leq 8\mathbb{E}\left[\int_0^{t\wedge\tau}(\sigma(u,X(u)) - \sigma(u,Y(u)))^2 du\right]\\
&+2t\mathbb{E}\left[\int_0^{t\wedge\tau}(b(u,X(u)) - b(u,Y(u)))^2 du\right]\\
&\leq 2K^2(t+4)\mathbb{E}\left[\int_0^{t\wedge\tau}(X(u) - Y(u))^2 du\right]\\
&\leq 2K^2(4+t)\int_0^t \mathbb{E}\left[\max_{0\leq u\leq s}(X(u\wedge\tau) - Y(u\wedge\tau))^2 ds\right].
\end{aligned}$$

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

$$\mathbb{E}\left[\max_{0\leq t\leq T}(X(t\wedge\tau) - Y(t\wedge\tau))^2\right] = 0.$$

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

$$0 \le \frac{d}{dx}\left(e^{-tK}F(t)\right) \le e^{-Kt}\left(-KF(t)+f(t)\right) \le 0,$$

and hence $e^{-tK}F(t) \le 0$, meaning that $F(t) \le 0$. But since F is the integral of the non-negative function f, this means that $f(t) = 0$.

Thus we have in particular that $\mathbb{P}[\max_{0\le t\le T}|X_t - Y_t| = 0] = 1$ as claimed. □

Finally, existence of solutions (for finite times) can be proven by the usual Picard iteration scheme under Lipschitz and growth conditions.

Theorem 4.4. *Let b,σ satisfy the Lipshitz conditions (4.2.6) and assume that*

$$|b(t,x)|^2 + |\sigma(t,x)|^2 \le K^2(1+|x|^2). \tag{4.2.10}$$

Let ξ be a random vector with finite second moment, independent of B_t, and let $\mathscr{F}_t$ be the usual augmentation, $\mathscr{F}_t$, of the filtration associated with B and ξ. Then there exists a continuous, $\mathscr{F}_t$-adapted process X which is a strong solution *of the SDE with initial condition ξ. Moreover, X is square integrable, i.e. for any $T > 0$, there exists $C(T,K)$, such that, for all $t \le T$,*

$$\mathbb{E}|X_t|^2 \le C(K,T)(1+\mathbb{E}|\xi|^2)e^{C(K,T)t}. \tag{4.2.11}$$

Proof. We define a map, F, from the space of continuous adapted processed X, uniformly square integrable on $[0,T]$, to itself, via

$$F(X)_t \equiv \xi + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s. \tag{4.2.12}$$

Note that the square integrability of $F(X)$ needs the growth conditions (4.2.10)

Exercise: Prove this!

As in (4.2.9)

$$\begin{aligned} &\mathbb{E}\left(\sup_{0\le t\le T}(F(X)_t - F(Y)_t)\right)^2 \qquad (4.2.13)\\ &\le 2E\left(\sup_{0\le t\le T}\left(\int_0^t(\sigma(X_s)-\sigma(Y_s))dB_s\right)^2\right)\\ &+2E\left(\sup_{0\le t\le T}\left(\int_0^t(b(X_s)-b(Y_s))d_s\right)^2\right)\\ &\le 2K^2(1+T)\int_0^T \mathbb{E}\sup_{0\le s\le t}(X_s-Y_s)^2\,dt \end{aligned}$$

and hence

$$\mathbb{E}\left(\sup_{0\le t\le T}(F^k(X)_t - F^k(Y)_t)\right)^2 \le \frac{C^kT^{2k}}{k!}\mathbb{E}\left(\sup_{0\le t\le T}(X_t-Y_t)\right)^2. \tag{4.2.14}$$

Thus, for n sufficiently large, F^n is a contraction, and hence has a unique fixed point which solves the SDE. □

Remark 4.5. 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 [11], 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 4.6. A stochastic integral equation

$$X_t = X_0 + \int_0^t \sigma(s,X_s)dB_s + \int_0^t b(s,X_s)ds \tag{4.2.15}$$

has a *weak solution* with initial distribution μ, if there exists a filtered space $(\Omega,\mathscr{F},\mathbb{P},\{\mathscr{F}_t\})$, satisfying the usual conditions, and continuous martingales X and B, such that

(i) B is an $\mathscr{F}_t$-Brownian motion;
(ii)X_0 has law μ;
(iii)$\int_0^t(|\sigma(s,X_s)|^2 + |b(s,X_s)|)ds < \infty$, a.s., for all t;
(iv)(4.2.15) holds.

Definition 4.7. A solution of (4.2.15) 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

$$X_t = X_0 + \int_0^t \text{sign}(X_s)dB_s. \tag{4.2.16}$$

Here we define $\text{sign}(x) = -1$, if $x \leq 0$, and $\text{sign}(x) = +1$, 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

$$B_t \equiv \int_0^t \text{sign}(X_s)dX_s. \tag{4.2.17}$$

Then $dB_s = \text{sign}(X_s)dX_s$, and hence

$$\int_0^t \operatorname{sign}(X_s)dB_s = \int_0^t \operatorname{sign}(X_s)^2 dX_s = \int_0^t dX_s = X_t - X_0,$$

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 (4.2.17) 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 4.8. 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 4.9. *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*

$$|X_t| - |X_0| = \int_0^t \operatorname{sign}(X_s)dX_s + \ell_t. \tag{4.2.18}$$

ℓ_t grows only when X is zero, i.e.

$$\int_0^t 1_{X_s \neq 0} d\ell_s = 0. \tag{4.2.19}$$

Proof. The proof uses Itô's formula and an applroximation of the absolute value by C^∞ functions. Chose some non-decreasing smooth funtion ϕ 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

$$f_n(X_t) - f_n(X_0) = \int_0^t f_n'(X_s)dX_s + \frac{1}{2}\int_0^t f_n''(X_s)d[X]_s. \tag{4.2.20}$$

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

$$\int_0^t \mathbb{1}_{X_s \notin [0,1/n]} dC_s^n = 0. \tag{4.2.21}$$

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 prove 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

$$\left\| \int_0^\infty \left(\mathrm{sign}(X_s) - f_n'(X_s)\right) dM_s \right\|_2^2 \leq \mathbb{E} \int_0^\infty \left(\mathrm{sign}(X_s) - f_n'(X_s)\right)^2 d[M]_s. \tag{4.2.22}$$

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

$$\mathbb{P}\left(\sup_{t\leq\infty} \left| \int_0^\infty \left(\mathrm{sign}(X_s) - f_n'(X_s)\right) dM_s \right| > \varepsilon \right] \leq \varepsilon^{-2} \mathbb{E} \int_0^\infty \left(\mathrm{sign}(X_s) - f_n'(X_s)\right)^2 d[M]_s, \tag{4.2.23}$$

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 ℓ_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 4.10. *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]

4.3 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 (4.1.2) has a solution, then it is a solution of a martingale problem.

Lemma 4.11. *Assume that X solves (4.1.2). Define the family of operator G_t on the space of C^∞-functions $f : \mathbb{R} \to \mathbb{R}$, as*

$$G_t \equiv \frac{1}{2}\sigma^2(t,x)\frac{d^2}{dx^2} + b(t,x)\frac{d}{dx}. \tag{4.3.1}$$

Then X is a solution of the martingale problem for G_t.

Remark 4.12. 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 $\mathscr{D}$, we say that a process X_t is a solution of the martingale problem, if for all $f : S \to \mathbb{R}$ in $\mathscr{D}$,

$$f(X_t) - \int_0^t (G_s f)(X_s)ds \tag{4.3.2}$$

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

$$M_t \equiv f(t,X_t) - f(0,X_0) - \int_0^t (\partial_s f(s,X_s) + (G_s f)(s,X_s))ds. \tag{4.3.3}$$

Restricting the domain of $\widetilde{G}$ to functions of the form $f(t,x) = \gamma(t)g(x)$, this reduces to

$$M_t \equiv g(X_t)\gamma(t) - g(X_0)g(0) - \int_0^t (\partial_s\gamma(s)g(X_s) + (G_s g)(X_s,s)\gamma(s))\, ds. \tag{4.3.4}$$

We see immediately, by setting $\gamma(t) \equiv 1$, that is (t, X_t) makes (4.3.4) a martingale, then X_t solves the time dependent martingale problem (4.3.2). On the other hand it is also easy to see that if X_t makes (4.3.2) a martingale then (t, X_t) makes (4.3.4) 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

$$\begin{aligned} f(t,X_t) - f(0,X_0) = & \int_0^t \partial_s f(s,X_s)ds + \int_0^t \partial_x f(s,X_s)dX_s \\ & + \frac{1}{2}\int_0^t \partial_x^2 f(s,X_s)d[X]_s. \end{aligned} \tag{4.3.5}$$

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 (4.1.2) 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

$$f(t,X_t)-f(0,X_0)-\int_0^t [\partial_s f(s,X_s+(Gf)(s,X_s)]ds = -\int_0^t \partial_x f(s,X_s)dM_s, \quad (4.3.6)$$

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 4.13. *Assume that b and σ are locally bounded as above and assume that in addition σ^{-1} is locally bounded. Let G_t be given by (4.3.1). Assume that X is a continuous solution to the martingale problem for (G,δ_{x_0}), then there exists a Brownian motion, B, such that (X,B) is a solution to the stochastic integral equation (4.1.2).*

Proof. We know that for every $f\in C^\infty(\mathbb{R})$,

$$f(X_t)-f(X_0)-\int_0^t (G_s f)(s,X_s)ds \quad (4.3.7)$$

is a continuous martingale. Choosing $f(x)=x$, it follows that

$$X_t - X_0 - \int_0^t b(s,X_s)ds \equiv M_t \quad (4.3.8)$$

is a continuous martingale. Essentially we want to show that this martingale is precisely the stochastic integral term in (4.1.2). To do this, we need to compute the bracket of M. For this we consider naturally (4.3.7) with $f(x)=x^2$. To simplify the notation, let us assume without loss of generality that $X_0=0$. This gives

$$X_t^2 - 2\int_0^t X_s b(s,X_s)ds - \int_0^t \sigma^2(s,X_s)ds = \widehat{M}_t, \quad (4.3.9)$$

where $\widehat{M}$ is a martingale. Thus

$$M_t^2 = X_t^2 - 2X_t \int_0^t b(s,X_s)ds + \left(\int_0^t b(s,X_s)ds\right)^2 \qquad (4.3.10)$$
$$= 2\int_0^t X_s b(s,X_s)ds + \int_0^t \sigma^2(s,X_s)ds + \widehat{M}_t$$
$$- 2X_t \int_0^t b(s,X_s)ds + \left(\int_0^t b(s,X_s)ds\right)^2.$$

I claim that

$$2\int_0^t X_s b(s,X_s)ds - 2X_t \int_0^t b(s,X_s)ds + \left(\int_0^t b(s,X_s)ds\right)^2 \qquad (4.3.11)$$

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 (4.3.11) 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

$$M_t^2 - \int_0^t \sigma^2(s,X_s)ds \qquad (4.3.12)$$

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 (4.1.2) with this particular realization of Brownian motion. □

We can summarize these findings in the following theorem.

Theorem 4.14. *Let $\mathbb{P}^y$ be a solution of the martingale problem associated to the operator G defined in (4.3.1) starting in y. Then there exists a weak solution of the SDE (4.1.2) with law $\mathbb{P}^y$. Conversely, if there is a weak solution of (4.1.2), then there* *exists a solution of the martingale problem for (4.3.1). 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 (4.3.1). 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 σ^{-1} from Theorem 4.13, which in fact can be avoided with some extra work.

As a consequence, we sketch two existence and uniqueness results for weak solutions.

Theorem 4.15. *Consider the SDE with time-independent coefficients,*

$$dX_t = b(X_t) + \sigma(X_t)dB_t, \tag{4.3.13}$$

in $\mathbb{R}^d$ where the coefficients b_i and σ_{ij} are bounded and continuous. Then for any measure μ such that

$$\int \|x\|^{2m}\mu(dx) < \infty, \tag{4.3.14}$$

for some $m > 0$, there exists a weak solution to (4.3.13) with initial measure μ.

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

$$\begin{aligned} X_0^{(n)} &= \xi \\ X_t^{(n)} &= X_{t_j^{(n)}}^{(n)} + b(X_{t_j^{(n)}}^{(n)})(t - t_j^{(n)}) + \sigma(X_{t_j^{(n)}}^{(n)})(B_t - B_{t_j^{(n)}}), t \in (t_j^{(n)}, t_{j+1}^{(n)}]. \end{aligned} \tag{4.3.15}$$

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

$$X_t^{(n)} = \xi + \int_0^t b^{(n)}(s, X^{(n)})ds + \int_0^t \sigma^{(n)}(s, X^{(n)})dB_s. \tag{4.3.16}$$

But then $X^{(n)}$ solves the martingale problem for the (time dependent) operator

$$(G_t^{(n)} f)(y) \equiv \sum_i b_i^{(n)}(t,y)\partial_i f(y(t)) + \frac{1}{2}\sum_{i,j,k} \sigma_{ik}^{(n)}(t,y)\sigma_{kj}^{(n)}(t,y)\partial_i\partial_j f(y(t)). \quad (4.3.17)$$

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 7.21. The basic ingredient is the following:

$$\mathbb{E}\left\|X_t^{(n)} - X_s^{(n)}\right\|^{2m} \leq C_m(t-s)^m \quad (4.3.18)$$

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,

$$\mathbb{E}\|X_0^{(n)}\|^{2m} \leq C_m' < \infty \quad (4.3.19)$$

by assumption. To prove (4.3.18), we write

$$\begin{aligned}
\mathbb{E}\left\|X_t^{(n)} - X_s^{(n)}\right\|^{2m} &\leq \mathbb{E}\left\|\int_s^t b_n(u,X_u^{(n)})du\right\|^{2m} && (4.3.20)\\
&\quad + \mathbb{E}\left\|\int_s^t \sigma_n(u,X_u^{(n)})dB_u\right\|^{2m} && (4.3.21)\\
&\leq (t-s)^{2m}\mathbb{E}\sup_{u\in[s,t]}\left\|b_n(u,X_u^{(n)})\right\|^{2m} && (4.3.22)\\
&\quad + K_m\mathbb{E}\left(\int_s^t \left\|\sigma_n(u,X_u^{(n)})\right\|^2 du\right)^m && (4.3.23)\\
&\leq C(m)(t-s)^m && (4.3.24)
\end{aligned}$$

Here we used the inequality (valid for local martingales

$$\mathbb{E}|M_t|^{2m} \leq K_m\mathbb{E}[M]_t^m, \quad (4.3.25)$$

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 (3.55) implies that P^* is a solution of the martingale problem and hence a weak solution of the sde exists. □

Remark 4.16. 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 $\mathscr{F}_t$.

Remark 4.17. The preceeding theorem can be extended rather easily to the case when b and σ are time-dependent, and even to the case when they are bounded, continuous progressive functionals.

Remark 4.18. The boundedness conditions on the coefficients can be replaced by the condition

$$\|b(y)\|^2 + \|\sigma(y)\|^2 \leq K\left(1+\|y\|^2\right), \tag{4.3.26}$$

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 4.19. *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,*

$$k_m \mathbb{E}[M]_T^m \leq \mathbb{E}\left(\sup_{0\leq s\leq T} |M_s|\right)^{2m} \leq K_m \mathbb{E}[M]_T^m. \tag{4.3.27}$$

Proof. The following proof (which is taken from [14]) is based on the following simple lemma, called the "good λ inequality". It is a nice eexample of how to use the margingale property to prove powerful inequalities.

Lemma 4.20. *Let X,Y be non-negative random variables. Assume that there exists $\beta > 1$, such that for all $\lambda >, \delta > 0$,*

$$\mathbb{P}(X > \beta\lambda, Y \leq \delta\lambda) \leq \psi(\delta)\mathbb{P}(X > \lambda), \tag{4.3.28}$$

where $\psi(\delta) \downarrow 0$, as $\delta \downarrow 0$. Then for any function positive, increlasing function $F : \mathbb{R}_+ \to \mathbb{R}_+$, such that $F(0) = 0$ and $\sup_{x>0} \frac{F(\alpha x)}{F(x)} < \infty$, there exists a constant C such that

$$\mathbb{E}F(X) \leq C\mathbb{E}F(Y). \tag{4.3.29}$$

Remark 4.21. Clearly, any $F(x) = x^m$, for any $m > 0$, satisfies the hypothesis of the lemma trivially, since $(\alpha x)^m/x^m = \alpha^m$ does not depend on x.

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 γ 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 (4.3.28) w.r.t. $F(d\lambda)$ and get, using partial integration,

$$\begin{aligned}\psi(\delta)\mathbb{E}F(X) &\geq \int_0^\infty F(d\lambda)\mathbb{E}\mathbb{1}_{Y/\delta\leq\lambda<X/\beta} \\ &= \mathbb{E}\left(\int_0^{X/\beta} F(d\lambda) - \int_0^{Y/\delta} F(d\lambda)\right)_+ \\ &= \mathbb{E}F(X/\beta) - \mathbb{E}F(Y/\delta) \\ &\geq \gamma\mathbb{E}F(X) - \mathbb{E}F(Y/\delta).\end{aligned} \tag{4.3.30}$$

Now we solve this for $\mathbb{E}F(X)$ to get

$$\mathbb{E}F(X) \leq \frac{\mathbb{E}F(Y/\delta)}{\gamma - \psi(\delta)} \tag{4.3.31}$$

We can choose δ so small that $\psi(\delta) \leq \gamma/2$. Then there exists μ such that $F(x/\delta) \leq \mu F(x)$, for all $x > 0$. This proves the inequality with $C = 2\mu/\gamma$. $\square$

We have to establish the inquality (4.3.28) 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 $a < 0 < b$,

$$\mathbb{P}(\tau_b < \tau_a) \leq -a/(b-a). \tag{4.3.32}$$

Now fix $\beta > 1, \lambda > 0$, and $0 < \delta < (\beta - 1)$. Set $\tau \equiv \inf(t : |M_t| > \lambda)$. Define

$$N_t \equiv (M_{t+\tau} - M_\tau)^2 - ([M]_{t+\tau} - [M]_t). \tag{4.3.33}$$

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

$$\sup_{t\leq T} N_t \geq (\beta-1)^2\lambda^2 - \delta^2\lambda^2, \tag{4.3.34}$$

and

$$\inf_{t\leq T} N_t \geq -\delta^2\lambda^2. \tag{4.3.35}$$

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 (4.3.32),

$$\mathbb{P}\left(M_T^* \geq \beta\lambda, [M]_T^{1/2} \leq \delta\lambda | \mathscr{F}_\tau\right) \leq \delta^2/(\beta-1)^2. \tag{4.3.36}$$

From this it follows that

$$\mathbb{P}\left(M_T^* \geq \beta\lambda, [M]_T^{1/2} \leq \delta\lambda\right) \leq \delta^2/(\beta-1)^2\mathbb{P}(\tau < T) = d^2/(\beta-1)^2\mathbb{P}(|M_T^*| > \lambda). \tag{4.3.37}$$

This proves (4.3.28) and hence

$$\mathbb{E}F(M_T^*) \leq C\mathbb{E}F([M]_T^{1/2}). \tag{4.3.38}$$

The converse inequality is obtained by the same procedure but chosing of $Y = M_T^*$ and $X = [M]_T^{1/2}$. $\square$

A uniqueness result is interestingly tied to a Cauchy problem.

Lemma 4.22. *If for every $f \in \mathbb{C}_0^\infty(\mathbb{R}^d)$ the Cauchy problem*

$$\begin{aligned} \frac{\partial u(t,x)}{\partial t} &= (Gu)(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d \\ u(0,x) &= f(x), \quad x \in \mathbb{R}^d \end{aligned} \tag{4.3.39}$$

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 $0 \le t \le T$,

$$\begin{aligned} \frac{\partial g(t,x)}{\partial t} + (G_s g)(t,x) &= 0, \quad (t,x) \in (0,\infty) \times \mathbb{R}^d \\ g(T,x) &= f(x), \quad x \in \mathbb{R}^d \end{aligned} \tag{4.3.40}$$

Then it follows from (4.3.6) that $g(t,X_t)$ is a local martingale for any solution of the martingale problem. Hence

$$\mathbb{E}_x f(X_T) = \mathbb{E}_x g(T,X_T) = \mathbb{E}_x g(0,X_0) = g(0,x), \tag{4.3.41}$$

is the same for any solution. This implies uniqueness of the one-dimensional distributions. □

Now Theorem 3.50 implies immediately the following corollary:

Corollary 4.23. *Under the assumptions of the preceeding lemma, weak uniqueness holds for the SDE corresponding to the generator G.*

4.4 Weak solutions from Girsanov's theorem

Girsanov's theorem provides a very efficient and explicite way of constructing weak solutions of certain SDE's.

Theorem 4.24. *Consider the stochastic differential equation*

$$dX_t = b(t,X_t) + dB_t, \quad 0 \le t \le T, \tag{4.4.1}$$

for fixed T. Assume that $b : [0,T]\times\mathbb{R}^d$ is measurable and satisfies, for some $K < \infty$,

$$\|b(t,x)\| \le K(1+\|x\|). \tag{4.4.2}$$

Then for any probability measure μ on $\mathbb{R}^d$ there exists a weak solution of (4.4.1) with initial law μ.

Proof. Let X be a family of Brownian motions starting in $x \in \mathbb{R}$ under laws P_x. Then

$$Z_t \equiv \exp\left(\int_0^t b(s,X_s)\cdot dX_s - \frac{1}{2}\int_0^t \|b(s,X_s)\|^2 ds\right) \tag{4.4.3}$$

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

$$W_t \equiv X_t - X_0 - \int_0^t b(s,X_s)ds \tag{4.4.4}$$

for $0 \leq t \leq T$ is a Brownian motion starting in 0. Thus we have a pair (X_t, W_t) such that

$$X_t = X_0 + \int_0^t b(s,X_s)ds + W_t, \tag{4.4.5}$$

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 (4.4.1). □

A complementary result also provided criteria for uniqueness in law.

Theorem 4.25. *Assume that we have weak solutions $(X^{(i)}, W^{(i)})$, $i = 1,2$, on filtered spaces $(\Omega^{(i)}, \mathscr{F}^{(i)}, \mathbb{P}^{(i)}, \mathscr{F}_t^{(i)})$, of the SDE (4.24) with the same initial distribution. If*

$$\mathbb{P}^{(i)}\left[\int_0^T \|b(t,X_t^{(i)})\|^2 dt < \infty\right] = 1, \tag{4.4.6}$$

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

$$\tau_k^{(i)} \equiv T \wedge \inf\left\{0 \leq t \leq T : \int_0^t \|b(t,X_t^{(i)})\|^2 dt = k\right\}. \tag{4.4.7}$$

We define the martingales

$$\xi_t^{(k)}(X^{(i)}) \equiv \exp\left(-\int_0^{t\wedge\tau_k^{(i)}} b(s,X_s^{(i)})dW_s^{(i)} - \frac{1}{2}\int_0^{t\wedge\tau_k^{(i)}} \|b(s,X_s^{(i)})\|^2 ds\right), \tag{4.4.8}$$

and the corresponding transformed measures $\widetilde{\mathbb{P}}_k^{(i)}$. Then by Girsanov's theorem,

$$X^{(i)}_{t\wedge\tau_k^{(i)}} \equiv X_0^{(i)} + \int_0^{\tau_k^{(i)}} b(s,X_s^{(i)})ds + W^{(i)}_{t\wedge\tau_k^{(i)}} \tag{4.4.9}$$

is a Brownian motion with unital distribution μ, 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 $[0,T]$. $\square$

4.5 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 [3, 4, 8].

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

$$\mathbb{P}\left[\sup_{0\leq s\leq 1} \|T^{-1}B_{sT} - \gamma(s)\| \leq \varepsilon\right]. \tag{4.5.1}$$

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 (4.5.1).

Lemma 4.26. *Let B be Brownian motion, set $B_s^T \equiv T^{-1}B_{Ts}$, and let γ be a smooth path in $\mathbb{R}^d$ starting in the origin. Then*

$$\lim_{\varepsilon\downarrow 0}\lim_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[\|B^T - \gamma\|_\infty \leq \varepsilon\right] \geq -I(\gamma) \equiv -\frac{1}{2}\int_0^1 \|\dot\gamma(s)\|^2 ds. \tag{4.5.2}$$

Proof. For notational simplicity we consider the case $d = 1$ only. Note that $B_s^T = T^{-1}B_{sT}$ has the same distribution as $T^{-1/2}B_s$. Thus we must estimate the probabilities

$$\mathbb{P}\left[\sup_{t\leq 1}\|B_t - \sqrt{T}\gamma(t)\| \leq \sqrt{T}\varepsilon\right]. \tag{4.5.3}$$

To do this, we observe that by Girsanov's theorem, the process

$$\widehat{B}_t \equiv B_t - \sqrt{T}\gamma(t) \tag{4.5.4}$$

is a Brownian motion under the measure $\mathbb{Q}$ defined through

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\sqrt{T}\int_0^t \dot\gamma(s)dB_s - \frac{T}{2}\int_0^t \|\dot\gamma(s)\|^2 ds\right). \tag{4.5.5}$$

Hence

$$\begin{aligned}
&\mathbb{P}\left[\|B-\sqrt{T}\gamma\|_\infty \le \sqrt{T}\varepsilon\right] \qquad (4.5.6)\\
&= \mathbb{P}\left[\|\widetilde{B}\|_\infty \le \sqrt{T}\varepsilon\right]\\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)dB_s+\frac{T}{2}\int_0^1\|\dot\gamma(s)\|^2ds}\mathbb{1}_{\|\widetilde{B}\|_\infty\le\sqrt{T}\varepsilon}\right]\\
&= \mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1 \dot\gamma(s)d\widetilde{B}_s-\frac{T}{2}\int_0^1\|\dot\gamma(s)\|^2ds}\mathbb{1}_{\|\widetilde{B}\|_\infty\le\sqrt{T}\varepsilon}\right]\\
&= e^{-\frac{T}{2}\int_0^1\|\dot\gamma\|^2(s)ds}\mathbb{Q}\left[\|\widetilde{B}\|_\infty\le\sqrt{T}\varepsilon\right]\mathbb{E}_{\mathbb{Q}}\left[e^{-\sqrt{T}\int_0^1\dot\gamma(s)d\widetilde{B}_s}\middle|\|\widetilde{B}\|_\infty\le\sqrt{T}\varepsilon\right]\\
&= e^{-\frac{T}{2}\int_0^1\|\dot\gamma(s)\|^2ds}\mathbb{P}\left[\|B\|_\infty\le\sqrt{T}\varepsilon\right]\mathbb{E}_{\mathbb{P}}\left[e^{-\sqrt{T}\int_0^1\dot\gamma(s)dB_s}\middle|\|B\|_\infty\le\sqrt{T}\varepsilon\right].
\end{aligned}$$

Now we may use Jensen's inequality to get that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left[e^{-\sqrt{T}\int_0^1\dot\gamma(s)dB_s}\middle|\|B\|_\infty\le\sqrt{T}\varepsilon\right] \qquad (4.5.7)\\
&\ge \exp\left(-\sqrt{T}\mathbb{E}_{\mathbb{P}}\left[\int_0^1\dot\gamma(s)dB_s\middle|\|B\|_\infty\le\sqrt{T}\varepsilon\right]\right)=1.
\end{aligned}$$

On the other hand, it is easy to see, using e.g. the maximum inequality, that, for any $\varepsilon>0$,

$$\lim_{T\uparrow\infty}\mathbb{P}\left[\|B\|_\infty\le\sqrt{T}\varepsilon\right]=1. \qquad (4.5.8)$$

Hence,

$$\liminf_{T\uparrow\infty}T^{-1}\ln\mathbb{P}\left[\|B-\sqrt{T}\gamma\|_\infty\le\sqrt{T}\varepsilon\right]\ge-\frac{1}{2}\int_0^t\|\dot\gamma(s)\|^2ds, \qquad (4.5.9)$$

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_s^T such that for all t_k, $B_{t_k}^T=L_{t_k}$. 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)

$$\mathbb{P}[\sup_{0\leq s\leq t} |B_s| > xt] \leq 2\exp\left(-\frac{x^2 t}{2}\right) \tag{4.5.10}$$

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 [1]).

This gives us

$$\begin{aligned}\mathbb{P}\left[\max_{0\leq t\leq \alpha} \|B_t\| > \delta\sqrt{T}/2\right] &\leq d\mathbb{P}\left[\max_{0\leq t\leq \alpha} |B_t| > \delta\sqrt{T}/2\sqrt{d}\right] \\ &\leq = 2de^{-\frac{\delta^2 n^2 T}{8d}}\end{aligned} \tag{4.5.11}$$

and so

$$\mathbb{P}\left[\|B^T - L\|_\infty > \delta\right] \leq n2e^{-\frac{\delta^2 n^2 T}{8d}} \tag{4.5.12}$$

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,

$$\limsup_{T\uparrow\infty} T^{-1}\ln\mathbb{P}[I(L)\geq\lambda] \leq \lambda. \tag{4.5.13}$$

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 η_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

$$\mathbb{P}\left[\frac{1}{2T}\sum_{i=1}^{dn}\eta_i^2 > \lambda\right] \leq e^{-\rho\lambda T}\mathbb{E}e^{\rho\sum_{i=1}^{nd}\eta_i^2/2} \\ \leq e^{-\rho\lambda T}C_\rho^{nd} \tag{4.5.14}$$

for all $\rho < 1$, and so (4.5.13) follows, for any n.

We can deduce from the two estimates the following version of the upper bound:

Proposition 4.27. *Let $K_\lambda \equiv \{\phi : I(\phi) \leq \lambda\}$. Then*

$$\limsup_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[\mathrm{dist}(\mathrm{B}^\mathrm{T},\mathrm{K}_\lambda) \geq \delta\right] \leq -\lambda. \tag{4.5.15}$$

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 λ has probability less than $\exp(-\lambda T)$.

The two bounds, together with the fact that the levels sets K_λ (of I are compact (a fact we will not prove), imply the usual formulation of a *large deviation principle*:

Theorem 4.28. *For any Borel set $A \subset W$,*

$$-\inf_{\gamma\in\mathrm{int}A} I(\phi) \leq \liminf_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[B^T \in A\right] \\ \leq \limsup_{T\uparrow\infty} T^{-1}\ln\mathbb{P}\left[B^T \in A\right] \leq -\inf_{\gamma\in\bar{A}} I(\phi), \tag{4.5.16}$$

where 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 (4.4.1) with a scaled down Brownian term. i.e. we want to consider the equation

$$X_t = T^{-1/2}B_t + \int_0^t b(X_s)ds. \tag{4.5.17}$$

(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

$$F(\gamma) = f, \tag{4.5.18}$$

where f is the solution of the integral equation

$$f(t) = \int_0^t b(f(s))ds + \gamma(t). \tag{4.5.19}$$

We may use Gronwall's lemma to show that this mapping is continuous. Then $X = F(B^T)$, and

$$\mathbb{P}[X \in A] = \mathbb{P}[B^T \in F^{-1}(A)]. \tag{4.5.20}$$

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

$$\mathbb{P}[X \in A] \leq \sup_{\gamma \in F^{-1}(\bar{A})} I(\gamma) = \sup_{F(\gamma) \in \bar{A}} I(\gamma) = \sup_{\gamma \in \bar{A}} I(F^{-1}(\gamma)), \tag{4.5.21}$$

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

$$F^{-1}(\gamma)(t) = \gamma(t) - \int_0^t b(\gamma_s) ds,$$

$$\widetilde{I}(\gamma) = \frac{1}{2} \int_0^1 \|\dot{\gamma}_s - b(\gamma_s)\|^2 ds \tag{4.5.22}$$

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

$$I(\gamma) = \int_0^t \mathscr{L}(\gamma(s), \dot{\gamma}(s), s) ds, \tag{4.5.23}$$

where the Lagrangian, $\mathscr{L}$, takes the special form

$$\mathscr{L}(\gamma(s), \dot{\gamma}(s), s) = \|\dot{\gamma}(s) - b(\gamma(s), s)\|_2^2. \tag{4.5.24}$$

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,

$$\frac{d}{dt} \frac{\partial}{\partial \dot{\gamma}} \mathscr{L}(\gamma, \dot{\gamma}, s) = \frac{\partial}{\partial \gamma} \mathscr{L}(\gamma, \dot{\gamma}, s). \tag{4.5.25}$$

In our case, these take the form

$$\frac{d^2}{dt^2} \gamma(t) = \frac{\partial}{\partial t} b(\gamma(t), t) + b(\gamma(t), t) \frac{\partial}{\partial \gamma(t)} b(\gamma(t), t). \tag{4.5.26}$$

One can readily identify a special class of solution of this second order equation, namely solutions of the first order equations

$$\dot{\gamma}(t) = b(\gamma(t), t), \tag{4.5.27}$$

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

$$\mathbb{P}[|X_0 - a| \leq d, |X_t - b| \leq \delta] \sim \exp\left(-\varepsilon^{-1} \inf_{\gamma:\gamma(0)=a,\gamma(t)=b} I(\gamma)\right), \tag{4.5.28}$$

which leads us to solve (4.5.26) subject to boundary conditions $\gamma(0) = a, \gamma(t) = b$. In general this will not solve (4.5.27), 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:

$$\begin{aligned}&\mathbb{P}[|X_0 - a| \leq d, |X_t - b| \leq \delta, \text{for some}\, t < \infty] \\ &\sim \exp\left(-\varepsilon^{-1} \inf_{\gamma:\gamma(0)=a,\gamma(t)=b,\text{for some}\, t\infty} I(\gamma)\right).\end{aligned} \tag{4.5.29}$$

Clearly the infimum will be zero, if the solution of the initial value problem (4.5.27) 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 (4.5.28) 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 (4.5.29) that admits an zero-action path γ, 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

$$I(\hat{\gamma}) - I(\gamma) = 2\int_0^T b(\gamma(s)) \cdot \dot{\gamma}(s) ds = \int_\gamma b(\gamma) d\gamma. \tag{4.5.30}$$

Let us now specialize to the case when the vector field b is the gradient of a potentia, $b(x) = \nabla F(x)$. Then

$$\int_\gamma b(\gamma) d\gamma = F(\gamma(T)) - F(\gamma(0)) = F(b) - F(a). \tag{4.5.31}$$

Hence

$$I(\hat{\gamma}) = I(\gamma) + F(b) - F(a), \tag{4.5.32}$$

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 [8].

4.6 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 4.29. *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 $\mathscr{F}_t$ measurable random variable,*

$$\mathbb{E}^h_x[Y] = \frac{1}{h(x)}\mathbb{E}_x[h(X_t)Y]. \tag{4.6.1}$$

Then $\mathbb{P}^h$ is the law of a solution of the martingale problem for the operator G^h defined by

$$(G^h f)(x) \equiv \frac{1}{h(x)}(Lhf)(x). \tag{4.6.2}$$

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 τ_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

$$dX_t = \frac{\nabla h(X_t)}{h(X_t)}dt + dB_t. \tag{4.6.3}$$

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

$$\mathbb{P}^h[\cdot] = \mathbb{P}[\cdot | X_{\tau_D} \in A] \tag{4.6.4}$$

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

$$dX_t = \frac{1}{X_t} dt + dB_t. \tag{4.6.5}$$

Note that we can take $R \uparrow \infty$ without changing the SDE. Thus, the solution of (4.6.5) 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 (4.6.5) 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 4.30. *The Bessel process in dimension d is a weak solution of*

$$dR_t = \frac{d-1}{2R_t} + dB_t. \tag{4.6.6}$$

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 (4.6.6) 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. [12], 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 4.31. *Let R_t be the solution of the SDE (4.6.6) with $d \geq 2$ and initial condition $R_0 = r \geq 0$. Then*

$$\mathbb{P}[\forall t > 0 : R_t > 0] = 1. \tag{4.6.7}$$

Proof. Let first $r > 0$. Let

$$\tau_k \equiv \inf\left\{t \geq 0 : R_t = k^{-k}\right\},$$

$$\sigma_k \equiv \inf\{t \geq 0 : R_t = k\}$$

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

$$\mathbb{E}_r\left[h\left(R_{T_k}\right)\right] = h(r). \tag{4.6.8}$$

Finally,

$$\mathbb{E}_r\left[h\left(R_{T_k}\right)\right] = h(k)\mathbb{P}_r[T_k = \sigma_k] + h(k^{-k})\mathbb{P}_r[T_k = \tau_k] + h(B_n)\mathbb{P}_r[T_k = n]. \tag{4.6.9}$$

Hence

$$\mathbb{P}_r[T_k = \tau_k] \leq \frac{h(r)}{h(k^{-k})} \leq \begin{cases} k^{-(\alpha-1)k} r^{-\alpha+1}, & \text{if } d \neq 2, \\ \frac{\ln r}{k \ln k}, & \text{if } d = 2. \end{cases} \tag{4.6.10}$$

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.,

$$\mathbb{P}[\tau < \infty] \leq \lim_{k \uparrow \infty} \mathbb{P}_r[\tau < \sigma_k] \leq \lim_{k \uparrow \infty} \mathbb{P}_r[\tau_k < \sigma_k] = 0. \tag{4.6.11}$$

This proves the case $r > 0$. For $r = 0$, just use that, by the strong Markov property, for any $\varepsilon > 0$,

$$\mathbb{P}_0\left[R_t > 0, \forall \varepsilon < t < \infty\right] = \mathbb{E}_0 \mathbb{P}_{B_\varepsilon}\left[\left[R_t > 0, \forall 0 < t < \infty\right] = 1, \tag{4.6.12}$$

since $\mathbb{P}_0[R_\varepsilon > 0] = 1$. Finally let $\varepsilon \downarrow 0$ to complete the proof. □

Remark 4.32. 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.

Chapter 5
SDE's and partial differential equations

Already in the context of discrete time Markov processes we have seen that the martingale problem formulation of Markov processes leads to an interesting connection between probability theory and linear boundary value problems. In the case of stochastic differential equations, this connections become even more profound leads to the connection between diffusion processes and potential theory which can be seen as one of the mathematical highlights of stochastic analysis. The classical case relates only to Brownian motion, but the extension to more general second order stochastic differential equations is quite straight-forward. Note that we will henceforth switch notation and denote generators by $\mathscr{L}$ rather than G, as the letter G will be needed to denote Green's functions. For analytic background on elliptic partial differential equations the standard reference is the textbook [10] by Gilbarg and Trudiger.

5.1 The Dirichlet problem

We consider the stochastic differential equation of the previous chapter with time-independent drift and dispersion matrix

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{5.1.1}$$

in $\mathbb{R}^d$. We have seen that the (weak) solutions of this equation are a strong Markov process with generator whose restriction to $C^2(\mathbb{R}^d)$ is given by

$$\mathscr{L} = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\frac{\partial}{\partial x_i}, \tag{5.1.2}$$

where the *diffusion matrix* a is given by

$$a_{ij}(x) = \sum_{k=1}^{d} \sigma_{ik}(x)\sigma_{kj}(x). \tag{5.1.3}$$

In the sequel we will always assume that the dispersion matric σ is non-degenerate and hence the diffusion matrix is strictly positive, i.e. for all $x \in \mathbb{R}^d$, $a(x)$ defines a strictly positive quadratic form. If a is strictly positive, then the operator $\mathscr{L}$ is called *elliptic*. If in some domain $D \subset \mathbb{R}^d$,

$$\sum_{i,j} a_{ij}(x)\xi_i\xi_j \geq \delta \|\xi\|^2,$$

for all $x \in D$, then we call $\mathscr{L}$ *uniformly elliptic* in D.

The classical *Dirichlet problem* associated with an elliptic operator $\mathscr{L}$ and a domain, D, is described as follows (we assume here that D is bounded). Let $D \subset \mathbb{R}$ and continuous functions $g\overline{D} \to \mathbb{R}$, $k : \overline{D} \to [0,\infty)$, and $u : \partial D \to \mathbb{R}$ be given. can we find a continuous function $f : \overline{D} \to \mathbb{R}$, such that

$$-(\mathscr{L} f)(x) + k(x) f(x) = g(x), \forall x \in D \tag{5.1.4}$$

$$f(x) = u(x), \forall x \in \partial D. \tag{5.1.5}$$

Remark 5.1. The Dirichlet problem can also be posed if u is not a continuous function on the boundary of D. In that case the condition that f be continuous on $\overline{D}$ must be replaced by that condition that, for all $x \in \partial D$, whenever a sequence $x_n \in D$ converges to x, then $f(x_n) \to u(x)$.

It is rather straightforward to see that the existence of a solution of such a problem implies a stochastic representation. Namely:

Theorem 5.2. *Assume that f solves the Dirichlet problem above, and let X be a weak solution of the SDE (5.1.1). Let $\tau_D \equiv \inf\{t \geq 0 : X_t \notin D\}$. If*

$$\mathbb{E}_x \tau_D < \infty, \quad \forall x \in D, \tag{5.1.6}$$

then

$$f(x) = \mathbb{E}_x \left[f(X_{\tau_D}) \exp\left(-\int_0^{\tau_D} k(X_s) ds \right) + \int_0^{\tau_D} g(X_s) \exp\left(-\int_0^t k(X_s) ds \right) dt \right] \tag{5.1.7}$$

Proof. The key to this result is the following lemma:

Lemma 5.3. *Let $\mathscr{F}_t$ be a filtration and X an adapted process. Let $f,g,k \in B(S)$. Then $f(X_t) - \int_0^t (\mathscr{L} f)(X_s) ds$ is a martingale if and only if*

$$M_t \equiv e^{-\int_0^t k(X_s)ds} f(X_t) + \int_0^t e^{-\int_0^s k(X_r)dr} \left(k(X_s) f(X_s) - (\mathscr{L} f)(X_s) \right) ds \tag{5.1.8}$$

is a martingale.

Proof. The proof of this lemma follows from Proposition 4.1.1 in [1] . Just choose for $M(t)$ the martingale $f(X_t) - \int_0^t (\mathscr{L} f)(X_s) ds$ and for $V(t)$ the process $\exp\left(-\int_0^t k(X_s) ds\right)$. Then it is a slightly tedious but straightforward computation (that uses Fubini's theorem at the right moment) to show that the expression in (5.1.8) is of the form $V(t)M(t) - \int_0^t M(t) dV(t)$ and hence a martingale. □

We use Lemma 5.3 with $g = \mathscr{L} f$ where f solves the Dirichlet problem. Now since $\mathbb{E}_x \tau_D < \infty$ by assumption, we get from the optional sampling theorem that

$$\mathbb{E}_x M_{\tau_D} = \mathbb{E}_x M_0. \tag{5.1.9}$$

But $\mathbb{E}_x M_0 = f(x)$, while, due to the fact that

$$\begin{aligned}\mathbb{E}_x M_{\tau_D} &= \mathbb{E}_x \Big[e^{-\int_0^{\tau_D} k(X_s) ds} f(X_D) \\ &\qquad + \int_0^{\tau_D} e^{-\int_0^s k(X_r) dr} \left(k(X_s) f(X_s) - \mathscr{L} f(X_s)\right) ds \Big] \\ &= \mathbb{E}_x \left[e^{-\int_0^{\tau_D} k(X_s) ds} f(X_D) + \int_0^{\tau_D} e^{-\int_0^s k(X_r) dr} g(X_s) ds \right]\end{aligned} \tag{5.1.10}$$

which is what we claimed. □

It is interesting to note that the finiteness of the expectation of the exit time τ_D is quite easily ensured (for bounded domains) from a rather weak ellipticity condition.

Lemma 5.4. *Let D be open and bounded in $\mathbb{R}^d$, and assume that for some $1 \leq \ell \leq d$,*

$$\min_{x \in \bar{D}} a_{\ell\ell}(x) > 0. \tag{5.1.11}$$

Then $\mathbb{E}_x \tau_D < \infty$, for all $x \in D$.

Proof. Set $a = \min_{x \in \bar{D}} a_{\ell\ell}(x)$, $b \equiv \max_{x \in \bar{D}} \|b(x)\|$, and $q \equiv min_{x \in \bar{D}} x_\ell$. Let $\nu > 2b/a$. Consider the smooth function $h(x) = -\mu e^{\nu x_\ell}$, with $\mu > 0$ to be chosen later. Clearly

$$-\mathscr{L} h(x) = \mu e^{\nu x_\ell} \left(\frac{1}{2} \nu^2 a_{\ell\ell}(x) + \nu b_\ell(x) \right) \geq \frac{1}{2} \mu \nu a e^{\nu b} \left(\nu - 2b/a\right).$$

Now we can chose μ such that the right-hand side is larger than 1, and so $\mathscr{L} h(x) \leq -1$, for all $x \in D$. But

$$h(X_{t \wedge \tau_D}) - \int_0^{t \wedge \tau_D} \mathscr{L} h(X_s) ds$$

is a martingale, and so

$$-\mathbb{E}_x \int_0^{t \wedge \tau_D} \mathscr{L} h(X_s) ds = h(x) - \mathbb{E}_x h(X_{t \wedge \tau_D}),$$

or

$$h(x) - \mathbb{E}_x h(X_{t\wedge\tau_D}) \geq \mathbb{E}_x(t \wedge \tau_D)$$

and hence

$$\mathbb{E}_x(t \wedge \tau_D) \leq \max_{y\in D} |h(y)| < \infty.$$

Passing to the limit $t \uparrow \infty$ implies $\mathbb{E}_x \tau_D < \infty$. $\square$

The previous results give a stochastic representation formula for solutions of the Dirichlet problem, assuming that a solution to the Dirichlet problem and a weak solution of the SDE exist. One may ask whether one can use this representation to prove the existence of solutions of the Dirichlet problem? We will address this question in the simpler context of Brownian motion.

Brownian motion and potential theory.

Let us now consider the setting where $\mathscr{L} = \frac{1}{2}\Delta$ and X_t is Brownian motion in $\mathbb{R}^d$. Let us begin with the simplest boundary value problem

$$\begin{aligned} \Delta f(x) &= 0, \quad x \in D, \\ f(x) &= u(x), \quad x \in \partial D. \end{aligned} \tag{5.1.12}$$

We assume that u is bounded and continuous. From the theorem above, an obvious candidate solution is

$$f(x) = \mathbb{E}_x u(B_{\tau_D}). \tag{5.1.13}$$

Now f clearly satisfies the boundary conditions, and it is also not hard to show hat $\Delta f(x) = 0$ for $x \in D$. There are various ways ti show this. Note first that we can write, with P_t the semi-group corresponding to the Brownian motion starting at x that

$$\begin{aligned} (P_t f)(x) &= \mathbb{E}_x\left[(\mathbb{1}_{\tau_D>t} + \mathbb{1}_{\tau_D\leq t})\,\mathbb{E}_{X_t}[u(B_{\tau_D})]\right] \\ &= \mathbb{E}_x\left[\mathbb{1}_{\tau_D>t}\mathbb{E}_{X_t}[u(B_{\tau_D})]\right] + \mathbb{E}_x\left[\mathbb{1}_{\tau_D\leq t}\mathbb{E}_{X_t}[u(B_{\tau_D})]\right]. \end{aligned} \tag{5.1.14}$$

Now in the first term we can use the Markov property to see that

$$\mathbb{E}_x\left[\mathbb{1}_{\tau_D>t}\mathbb{E}_{X_t}[u(B_{\tau_D})]\right] = \mathbb{E}_x[u(B_{\tau_D})] = f(x) \tag{5.1.15}$$

while the second satisfies the bound

$$\left|\mathbb{E}_x\left[\mathbb{1}_{\tau_D\leq t}\mathbb{E}_{X_t}[u(B_{\tau_D})]\right]\right| \leq \max_{x\in\partial D} u(x)\mathbb{P}_x\left[\tau_D \leq t\right]. \tag{5.1.16}$$

If $\mathrm{dist}(x, D^c) = r > 0$, then is follows easily that

$$\mathbb{P}_x\left[\tau_{D^c} \leq t\right] \leq \mathbb{P}_x\left[\sup_{0\leq s\leq t} |B_s| \leq r\right] \leq 2de^{-r^2/2t}, \tag{5.1.17}$$

and in particular

$$\lim_{t\downarrow 0} t^{-1}\mathbb{P}_x\left[\tau_D \le t\right] = 0, \tag{5.1.18}$$

for any $x \in D$. This then implies that

$$\frac{1}{2}\Delta f(x) = \lim_{t\downarrow 0} t^{-1}([P_t - \mathbb{1}]f)(x) = 0. \tag{5.1.19}$$

We see that all that remains to show to establish that f solves the Dirichlet problem is the continuity of f at the boundary of D. [1]

As we will see, the continuity property is linked to regularity properties of the boundary of D.

Definition 5.5. Define the stopping time $\sigma_D \equiv \inf\{t > 0 : B_t \in D^c\}$ (note the difference to τ_D when we start the process in the boundary of D!). A point, $z \in \partial D$, is called *regular* , if $\mathbb{P}_z[\sigma_D = 0] = 1$.

Thus a regular point has the property that the Brownian motion starting at it will essentially immediately return to the boundary. An irregular point is one from which Brownian motion can immediately escape into D.

Remark 5.6. It follows from the so-called *Blumenthal-Getoor* $0-1$-*law* (Lemma 5.7 below) that If a point z is not regular, then $\mathbb{P}_z[\sigma_D = 0] = 0$.

Lemma 5.7. *[Blumenthal-Getoor* 0-1-*law] Let B_t be a d-dimensional Brownian motion, starting in x, on a filtered space $(\Omega, \widetilde{\mathscr{F}}, \mathbb{P}_x, \widetilde{\mathscr{F}_t})$ where $\widetilde{\mathscr{F}}$ is the usual augmentation of the natural filtration, $\mathscr{F}_t$, generated be the Brownian motion. Then, if $F \in \widetilde{\mathscr{F}_0}$, $\mathbb{P}_x[F] \in \{0,1\}$.*

Proof. If $F \in \widetilde{\mathscr{F}_0}$, then F differs from some set $G \in \mathscr{F}_0$ only by a $\mathbb{P}_x$-null set. But since G must be of the form $G = \{B_0 \in A\}$ for some Borel set A, it follows that

$$\mathbb{P}_x[F] = \mathbb{P}_x[G] = \mathbb{1}_A(x) \in \{0,1\}.$$

□

The following theorem establishes that the Dirichlet problem is solvable (uniquely) for bounded regular domains.

Theorem 5.8. *Let $d \ge 2$ and let $z \in \partial D$ be fixed. Then the following statements are equivalent:*

(i) For any bounded measurable function $u : \partial D \to \mathbb{R}$ which is continuous at z,

$$\lim_{D \ni x \to z} \mathbb{E}_x u(B_{\tau_D}) = u(z). \tag{5.1.20}$$

---

[1] Clearly continuity is essential: without asking it there is no point in the problem, since it would admit lots of solutions, e.g. zero in D and u on ∂D.

(ii)z is a regular point for D.
(iii)For all $\varepsilon > 0$,

$$\lim_{D \ni x \to z} \mathbb{P}_x[\tau_D > \varepsilon] = 0. \tag{5.1.21}$$

Proof. We first proof that (i) implies (ii). From the remark 5.6, we know that if the origin is irregular, then $\mathbb{P}_z[\sigma_D = 0] = 0$. We will use the fact that in $d \geq 2$, the probability that Brownian motion visits any given point is zero, and in particular the probability that it returns to its starting point is zero. Thus, if K_r denotes the ball of radius r around z,

$$\lim_{r \downarrow 0} \mathbb{P}_z[B_{\sigma_D} \in K_r] = \mathbb{P}_z[B_{\sigma_D} = z] = 0.$$

Now fix r such that $\mathbb{P}_z[B_{\sigma_D} \in K_r] < 1/4$ and chose a sequence δ_n, $0 < \delta_n < r$, tending to zero. Let $\tau_n \equiv \inf\{t \geq 0 : \|B_t\| \geq \delta_n\}$. Then $\mathbb{P}_z[\tau_n \downarrow 0] = 1$, and so $\lim_n \mathbb{P}_z[\tau_n < \sigma_D] = 1$. Moreover, on $\{\tau_n < \sigma_D\}$ we have that $B_{\tau_n} \in D$. Thus for n so large that $\mathbb{P}_z[\tau_n < \sigma_D] \geq 1/2$, we have then that

$$\begin{aligned}
\frac{1}{4} &\geq \mathbb{P}_z[B_{\sigma_D} \in K_r] \geq \mathbb{P}_z[B_{\sigma_D} \in K_r, \tau_n < \sigma_n] \\
&= \mathbb{E}_z[\mathbb{1}_{\tau_n < \sigma_D} \mathbb{P}_z[B_{\sigma_D} \in K_t | \mathscr{F}_{\tau_n}] \\
&= \int_{D \cap \partial K_{\delta_n}} \mathbb{P}_z[\tau_n < \sigma_D, B_{\tau_n} \in dx] \mathbb{P}_x[B_{\sigma_D} \in K_r] \\
&\geq \inf_{x \in D \cap K_{\delta_n}} \mathbb{P}_x[B_{\tau_D} \in K_r] \int_{D \cap K_{\delta_n}} \mathbb{P}_z[\tau_n < \sigma_D, B_{\tau_n} \in dx] \\
&= \inf_{x \in D \cap \partial K_{\delta_n}} \mathbb{P}_x[B_{\tau_D} \in K_r] \mathbb{P}_z[\tau_n < \sigma_D] \\
&\geq \frac{1}{2} \inf_{x \in D \cap \partial K_{\delta_n}} \mathbb{P}_x[B_{\tau_D} \in K_r].
\end{aligned} \tag{5.1.22}$$

Hence $\mathbb{P}_{x_n}[B_{\tau_D} \in K_r] \leq \frac{1}{2}$ for some $x_n \in D \cap K_{\delta_n}$. Now choose a continuous bounded function, f, with $f(z) = 1$, $f(x) \leq 1$, $x \in K_r$, and $f(x) = 0$, $x \notin K_r$. For such functions we get

$$\limsup_n \mathbb{E}_{x_n} f(B_{\tau_D}) \leq \limsup_n \mathbb{P}_{x_n}[B_{\tau_D} \in K_r] \leq \frac{1}{2} < 1 = f(z),$$

so that (i) cannot hold. Therefore (i) implies (ii).

Let us now show that (ii) implies (iii). Notice that the function

$$\begin{aligned}
g_\delta(x) &\equiv \mathbb{P}_x[B_s \in D; \delta \leq s \leq \varepsilon] = \mathbb{E}_x[\mathbb{P}_{B_\delta}[\tau_D > \varepsilon - \delta]] \\
&= \int \mathbb{P}_y[\tau_D > \varepsilon - \delta] \mathbb{P}_x[B_\delta \in dy]
\end{aligned} \tag{5.1.23}$$

is continuous in x. But

$$g_\delta(x) \downarrow g(x) \equiv \mathbb{P}_x[B_s \in D; 0 < s \leq \varepsilon] = \mathbb{P}_x[\sigma_D > \varepsilon],$$

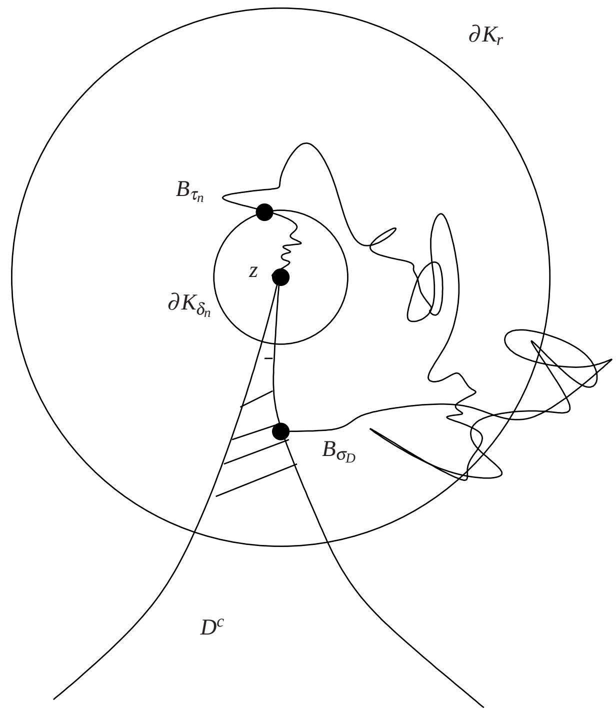

Fig. 5.1 Setting in the proof of (i) implies (ii)

as $\delta \downarrow 0$, so that g is upper semi-continuous. This implies that

$$\limsup_{x\to z} P_x[\tau_D > \varepsilon] \le \limsup_{x\to z} P_x[\sigma_D > \varepsilon] = \limsup_{x\to z} g(x) \le g(z) = 0,$$

where the last inequality comes from the regularity of z, i.e. (ii). Thus we have (iii) from (ii).

Finally we show that (iii) implies (i). We start from the observation that $\mathbb{P}_x[\max_{0\le t\le\varepsilon} \|B_t - B_0\| < r]$ is independent of x and converges to one as $\varepsilon \downarrow 0$. Now

$$\begin{aligned}\mathbb{P}_x[\|B_{\tau_D} - B_0\| < r] &\ge \mathbb{P}_x\left[\left\{\max_{0\le t\le\varepsilon} \|B_t - B_0\| < r\right\} \cap \{\tau_D \le \varepsilon\}\right] \\ &\ge \mathbb{P}_0\left[\left\{\max_{0\le t\le\varepsilon} \|B_t\| < r\right\}\right] - \mathbb{P}_x[\tau_D \le \varepsilon].\end{aligned}$$

When $x \to z$, by (iii) the second term vanishes for all ε, and letting $\varepsilon \downarrow 0$, the first term tends to one. Thus we get that

$$\lim_{D\ni x\to z} \mathbb{P}_x[\|B_{\tau_D} - x\| < r] = 1.$$

Thus

$$\begin{aligned} |\mathbb{E}_x f(B_{\tau_D}) - f(z)| &\leq |\mathbb{E}_x f(B_{\tau_D}) - f(z)| \\ &\leq |\mathbb{E}_x \left[\mathbb{1}_{\|B_{\tau_D} - x\| < r} (f(B_{\tau_D}) - f(z)) \right] | \\ &\quad + 2 \max_{y \in \partial D} |f(y)| \mathbb{P}_x [\|B_{\tau_D} - x\| \geq r] \end{aligned} \tag{5.1.24}$$

Clearly all three terms vanish as $x \to z$ and $r \downarrow 0$ by the continuity at zero and boundedness of f. □

The preceeding theorems imply that the Dirichlet problem has a unique solution if and only if any point in the boundary of D is regular. Otherwise, no solution exists. Moreover, the solution has the stochastic representation (5.1.14).

Th following proposition gives a sufficient verifiable criteria for regularity.

Proposition 5.9. *A point, $z \in \partial D$ is regular if there exists a cone, A, with vertex z, such that, for some $r > 0$, $A \cap K_r(z) \subset D^c$.*

Proof. Let $C > 0$ denote the fraction of the surface of $K_r(z)$ that lies within A. Let $K^{(n)} \equiv K_{r/n}(z)$, and $A_n \equiv A \cap \partial K^{(n)}$. Now $\tau_D = 0$ if $B_{\tau_{K^{(n)}}} \in A_n$ for arbitrary large n, i.e. $\{\tau_D = 0\} \supset \limsup_n \{B_{\tau_{K^{(n)}}} \in A_n\}$. Thus

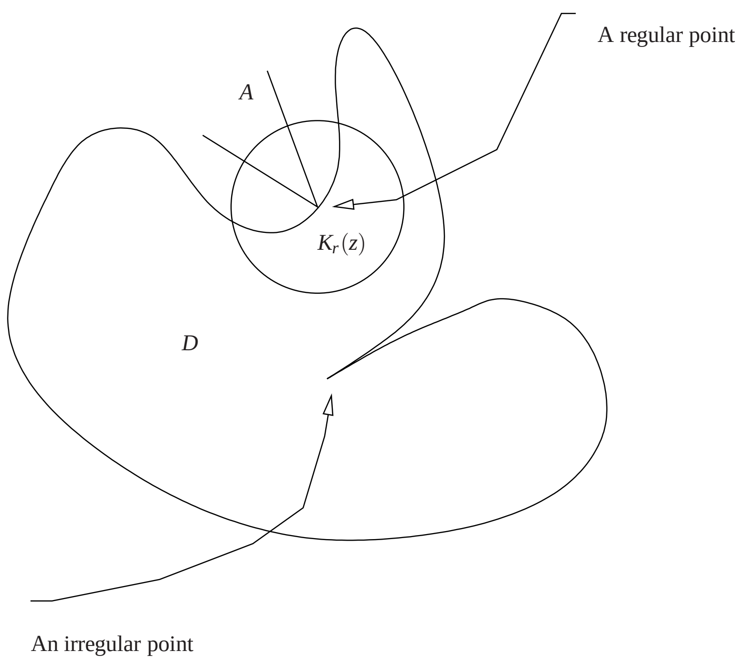

Fig. 5.2 A domain with one irregular point violating the cone-condition

$$\mathbb{P}_z[\tau_D = 0] \geq \mathbb{P}_z[\limsup_n \{B_{\tau_{K^{(n)}}} \in A_n\}] \qquad (5.1.25)$$
$$\geq \limsup_n \mathbb{P}_z[B_{\tau_{K^{(n)}}} \in A_n] = C > 0.$$

The fact that then $\mathbb{P}_z[\tau_D = 0] = 1$ follows from the fact that the event in question is in $\mathscr{F}_{0+}$ and the Blumenthal-Getoor zero-one law. $\square$

A slightly more abstract criterion is interesting because it involves the notion of a *barrier*.

Definition 5.10. Let $D \subset \mathbb{R}^d$ be open and $a \in \partial D$. A continuous function $v : \bar{D} \to \mathbb{R}$ that is harmonic in D, positive in $\bar{D}\backslash\{a\}$, and zero at a is called a *barrier*.

Proposition 5.11. *Let D be bounded and $a \in \partial D$. If there exists a barrier at a, then a is regular.*

Proof. Let v be a barrier. Let $f : \partial D \to \mathbb{R}$ and define $M \equiv \sup_{x\in\partial D} |f(x)|$. For any $\varepsilon > 0$, we can find $\delta > 0$, such that for $x \in \partial D$ and $|x-a| \leq \delta$, $|f(x) - f(a)| \leq \varepsilon$. Choose k such that $kv(x) \geq 2M(x)$, for $x \in \bar{D}$ and $|x-a| \geq \delta$. Then $|f(x) - f(a)| \leq \varepsilon + kv(x)$, for all $x \in \partial D$. Thus

$$|\mathbb{E}_x f(B_{\tau_D}) - f(a) \leq \varepsilon + k\mathbb{E}_x v(B_{\tau_D}) \leq \varepsilon + kv(x),$$

for all $x \in D$. Now since v is continuous and $v(a) = 0$, it follows that

$$\limsup_{d\ni x\to a} |\mathbb{E}_x f(B_{\tau_d}) - f(a)| \leq \varepsilon,$$

for all $\varepsilon > 0$, hence condition (i) of Theorem 5.8 holds and a is regular. $\square$

To show that the discussion of regular points is not empty, let us look at a classical example of a point that is not regular. This is called *Lebesgue's thorn*. Let $d = 3$, and define, for ε_n, $n \in \mathbb{N}$ such that $\varepsilon_n \downarrow 0$, the sets

$$E \equiv \left\{(x_1,x_2,x_3) : -1 < x_1 < 1; x_2^2 + x_3^2 < 1\right\}; \qquad (5.1.26)$$

$$F_n \equiv \left\{(x_1,x_2,x_3) : 2^{-n} \leq x_1 \leq 2^{-n+1}; x_2^2 + x_3^2 \leq \varepsilon_n\right\}; \qquad (5.1.27)$$

$$D \equiv E \backslash \bigcup_{n\in\mathbb{N}} F_n. \qquad (5.1.28)$$

Let $B_t \equiv (B_t^{(1)}, B_t^{(s)}, B_t^{(3)})$ be three dimensional Brownian motion. We know form our discussion of the Bessel-processes that $(B_t^{(2)}, B_t^{(3)})$ will never hit the point $(0,0)$, i.e.

$$\mathbb{P}\left[\exists t > 0 : (B_t^{(2)}, B_t^{(3)}) = (0,0)\right] = 0.$$

Thus B_t will never hit the compact set

$$K_n \equiv \left\{(x_1,x_2,x_3) : 2^{-2} \leq x_1 \leq 2^{-n+1}; x_2^2 + x_3^2 = 0\right\};$$

Since moreover $\|B_t\| \to \infty$, a.s., almost all paths remain some positive distance away form the set K_n, and hence, the probability that a path enters an ε-neighborhood of it can be made as small as desired by choosing ε small enough. In particular, one can choose ε_n so small that

$$\mathbb{P}\left[\exists t > 0 : B_t \in F_n\right] \leq 3^{-n}.$$

But unless B_t (starting at 0) immediately returns to D, i.e. if $\sigma_D = 0$, B_t must enter the set $\bigcup_{n\in\mathbb{N}} F_n$, so that

$$\begin{aligned}\mathbb{P}_0[\sigma_D = 0] &\leq \mathbb{P}\left[\exists t > 0, \exists_n : B_t \in F_n\right] \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}\left[\exists t > 0 : B_t \in F_n\right] \leq \sum_{n=1}^{\infty} 3^{-n} < 1.\end{aligned} \tag{5.1.29}$$

Hence 0 is not regular.

5.2 Maximum principle and Harnack-inequalities

The relation between harmonic functions and martingales has a number of further implications.

The first of these is the *mean value property*.

Lemma 5.12. *Let D be a bounded domain, τ_D the first hitting time of a Markov process with generator $\mathscr{L}$. Let $z \in D$ be fixed. Assume that $\mathbb{E}_z \tau_D < \infty$. Define μ_D as the probability measure on ∂D given by*

$$\mu_D(dx) = \mathbb{P}_z\left[X_{\tau_D} \in dx\right]. \tag{5.2.1}$$

Then, a if a function $h : D \to \mathbb{R}$ is harmonic in D, it holds that,

$$\int_{\partial D} \mu_D(dx) h(x) = h(z). \tag{5.2.2}$$

Proof. Use the fact that $h(x_{\tau_D})$ is a martingale. □

The measure $\mu_D(dx)$ is called the *exit distribution*. It is absolutely continuous with respect to the Euclidean surface measure, $n_D(dx)$ on ∂D.

An immediate consequence of the mean value property is the *maximum principle:*

Theorem 5.13. *Let h be harmonic in an open, connected domain D. If h achieves its supremum in D, then it is constant.*

Proof. Let $h(x) = \sup_{y\in D} h(y) = M$. Let $D_M \equiv \{y \in D : h(y) = M\}$. Since h is continuous, this set is closed. Moreover, by the mean value property, for any $y \in D_M$, for any ball $B_r(y) \subset D$

$$M = h(y) = \int_{\partial B_r(y)} \mu_{B_r(y)}(dz) h(z),$$

which implies that for $\mu_{B_r(y)}$-almost all $z \in \partial B_r(y)$, $h(z) = M$. Since both h is continuous (and $\mu_{B_r(y)}$ is is absolutely continuous with respect to the surface measure, $h(y) = M$ for all $y \in \partial B_r(y)$. But then D_M is open, and being open and closed, it must coincide with D. $\square$

A more subtle consequence of the martingale property for harmonic functions are the *Harnack-inequalities*.

We consider first the case of Brownian motion. Let $R > 0$ and let $B_R(x)$ the ball of radius R centered at x. By symmetry, we have the $\mathbb{P}_x\left[\tau_{B_R(x)} \in dz\right]$ is the uniform distribution on $\partial B_R(x)$. Hence, by the mean value property,

$$\begin{aligned} h(x) &= \frac{1}{\int_{B_R(x)} d^d y} \int_0^R dr \int_{\partial B_r(x)} h(z) \sigma_{B_r(x)}(dz) \\ &= \frac{1}{V(B_R)} \int_{B_R(x)} h(y) d^d y. \end{aligned} \tag{5.2.3}$$

I.e., the value of $h(x)$ equals to its spatial average over the ball of radius R. Now let $y \in B_R(x)$ and r sucht that $B_r(y) \subset B_R(x)$. Clearly we have again that

$$h(y) = \frac{1}{V(B_r)} \int_{B_r(y)} h(z) d^d z. \tag{5.2.4}$$

Now let h be a *non-negative* harmonic function. Then it follows that

$$h(x) \geq \frac{1}{V(B_R)} \int_{B_r(y)} h(z) d^d z = \frac{V(B_r)}{V(B_R)} h(y) = \left(\frac{r}{R}\right)^d h(y). \tag{5.2.5}$$

From these basic estimates one can now derive the the Harnack inequality.

Theorem 5.14. *Let $D' \subset D$ be two connected open sets. Let h be a non-negative harmonic fucntion with respect to Brownian motion on $D \subset \mathbb{R}^d$. Then there exists a constant K, depending only on D, D', such that*

$$\sup_{x \in D'} h(x) \leq K \inf_{x \in D'} h(x). \tag{5.2.6}$$

Proof. For $y \in D'$ choose R such that $B_{4R}(y) \subset D$. Then for any two points, $x_1, x_2 \in B_R(y)$, the previous inequalities imply that

$$\begin{aligned} h(x_1) &= \frac{1}{V(B_R)} \int_{B_R(x_1)} h(z) d^d z \leq \frac{1}{V(B_R)} \int_{B_{2R}(y)} h(z) d^d z, \\ h(x_2) &= \frac{1}{V(B_{3R})} \int_{B_{3R}(x_2)} h(z) d^d z \geq \frac{1}{V(B_{3R})} \int_{B_{2R}(y)} h(z) d^d z, \end{aligned} \tag{5.2.7}$$

Hence

$$\sup_{x \in B_R(y)} h(x) \leq 3^d \inf_{x \in B_R(y)} h(x). \tag{5.2.8}$$

Now let x_1 and x_2 in $\bar{D'}$ be such that $h(x_1) = \sup_{x \in D'} h(x)$, $h(x_2) = \inf_{x \in D'} h(x)$. Now let γ be a closed arc joining x_1 and x_2 in D. Choose R such that $4R <$ $\text{dist}(\gamma, \mathrm{D}^c)$. This arc can, by the Heine-Borel theorem, be covered by a finite number, N, of balls of radius R, where N depends only on D and D'. Then we an compare $h(x_1)$ and $h(x_2)$ by using he estimate (5.2.8) not more than N times, hence

$$h(x_1) \leq 3^{dN} h(x_2). \tag{5.2.9}$$

This proves the theorem. □

There are obvious extensions of the Harnack inequality beyond Brownian motion (for analytic proofs in the general case of elliptic SDE's, see [10]). In fact, inspecting the proof all we used on Brownian motion beyond the martingale property of harmonic functions was the uniformity of the exit distribution on balls. Moreover, it is clear that to get a Harnack inequalty, we do not really need uniformity, but upper and lower bounds on the density of the exit distribution are sufficient.

Theorem 5.15. *Let X be a continuous strong solution of an SDE. Let $D \subset \mathbb{R}^d$ be a bounded open domain. Assume that there exist constants, $0 < c < C < \infty$, depending only on D, such that, for any ball $B_R(x) \subset D$,*

$$c \leq \frac{\mathbb{P}_x \left(X_{\tau_{B_R}} \in dy \right)}{\sigma_{B_R}(dy)} \leq C. \tag{5.2.10}$$

Then any harmonic function h satisfies a Harnack inequality in D, in the sense that for any $D' \subset D$, there exists a constant K, such that (5.2.6) holds.

The proof is on the exact same lines as that of the previous theorem and will be left as an exercise.

Chapter 6
Reversible diffusions

In this Chapter we turn to more explicit computations in the context of diffusion processes with small diffusivity. We will exploit the some special structures in the context of reversible processes.

6.1 Reversibility

The theory of Markov processes that we have developed so far can be seen as a theory of operators acting either on bounded functions (the semi-group action of functions), or on measures. In special cases we can replace this by a L^2 theory with respect to certain measures.

Let P_t by a strongly continuous contraction semi-group acting on a space $B(\mathscr{S})$. Assume that a measure, μ, on S, is invariant with respect to P_t. Then the action of P_t can be extended to the L^2 space $L^2(S,\mu)$.

Lemma 6.1. *Let f be in $L^2(S,\mu)$ where μ is invariant with respect to P_t. Then $(P_t f) \in L^2(S,\mu)$.*

Proof. We will show that the L^2-norm of $P_t f$ is controlled by that of f. Namely,

$$\begin{aligned}\int \mu(dx)\,[(P_t f)(x)]^2 &= \int \mu(dx)\left[\int P_t(x,dy) f(y)\right]^2 \\ &\leq \int \mu(dx) \int P_t(x,dy) f(y)^2 \int P_t(x,dy) \\ &\leq \int \mu(dx) \int P_t(x,dy) f(y)^2 = \int \mu(dx) f(x)^2\end{aligned} \tag{6.1.1}$$

Note that we used the Cauchy-Schwarz inequality and the invariance of μ. □

Having an L^2-action of P_t, we can naturally define its adjoint, P_t^*, via

$$\int \mu(dx) f(x)(P_t g)(x) = \int \mu(dx)\,(P_t^* f)\,(x) g(x), \tag{6.1.2}$$

for all $f,g \in L^2(S,\mu)$. One may check that P^* is itself a Markov semigroup that generates the time-reversed process to X, in the sense that $(P_t^* f)(X_t) = f(X_0)$.

Definition 6.2. A measure, μ, on S is called reversible with respect to P_t, if, for all functions $f,g \in L^2(S,\mu)$,

$$\int f(x)(P_t g)(x)\mu(dx) = \int g(x)(P_t f)(x)\mu(dx) \tag{6.1.3}$$

Lemma 6.3. *If μ is a reversible probability measure for P_t, then μ is an invariant probability measure for P_t.*

Proof. Clearly $f = 1$ is in $L^2(S,\mu)$. Hence we have

$$\int (\mu P_t)(dx)g(x) = \int (P_t g)(x)\mu(dx) = \int g(x)\mu(dx). \tag{6.1.4}$$

for all bounded measurable functions g, hence μ is invariant. □

Note that the converse is not true in general, i.e. an invariant measure is not necessarily reversible.

Thus, we may also say that that a measure is reversible with respect to P_t, if P_t is *self-adjoint* on the space $L^2(S,\mu)$.

The terminology "reversible measure" is customary, but actually irritating. The reversibility property is one of the Markov process, resp. the semi-group, and not one of the measure. So I prefer to call a Markov semi-group reversible, if there exits a measure, μ, such that P_t is symmetric in the space $L^2(S,\mu)$, i.e. that (6.1.3) holds.

One of the nice things is that a SCCSG that is reversible is a contraction in the L^2-space, by Lemma 6.1.

The notions above introduced through the semi-group extend to the generator of a Markov process. Thus, for an invariant measure μ, we can define the adjoint, $\mathscr{L}^*$ of a generator $\mathscr{L}$, such that

$$\int \mu(dx)(\mathscr{L}^* g)(x) f(x) = \int \mu(dx)g(x)(\mathscr{L} f)(x), \tag{6.1.5}$$

for all $f,g \in \mathscr{D}(\mathscr{L})$ such that $\mathscr{L} f, \mathscr{L} g \in L^2(S,\mu)$. Note that, if μ is a probability measure, the second condition is automatically verified. A reversible Markov process is then characterized by the fact that its generator is self-adjoint in $L^2(S,\mu)$ for some invariant measure μ.

Theorem 6.4. *Let μ be a reversible measure for a Markov process. Then the generator, $\mathscr{L}$, defines a non-negative definite quadratic form,*

$$\mathscr{E}(f,g) \equiv -\int \mu(dx)g(x)(\mathscr{L} f)(x), \tag{6.1.6}$$

called the Dirichlet form .

Proof. First, due to the fact that $\mathscr{L}$ is self-adjoint, $\mathscr{E}(f,f)$ is real for all f in $\mathscr{D}(\mathscr{L})$. Moreover, by definition, we have that for such f and if $\mathscr{E}(f,f) < \infty$,

$$\mathscr{E}(f,f) = \lim_{t\downarrow 0} t^{-1} \int \mu(dx) f(x) \left(f(x) - (P_t f)(x) \right). \tag{6.1.7}$$

But

$$\begin{aligned} \int \mu(dx) f(x) \left(f(x) - (P_t f)(x) \right) &= \|f\|_{2,\mu}^2 - \int \mu(dx) f(x) (P_t f)(x) \\ &\geq \|f\|_{2,\mu}^2 - \|f\|_{2,\mu} \|P_t f\|_{2,\mu} \geq \|f\|_{2,\mu}^2 - \|f\|_{2,\mu} \|f\|_{2,\mu} = 0, \end{aligned} \tag{6.1.8}$$

where we used Cauchy-Schwarz and Lemma 6.1. This implies that the limit, too, is non-negative. □

Remark 6.5. The form $\mathscr{E}$ can be extended to the set $\{f : \mathscr{E}(f,f) < \infty\}$ which mostly is larger than the domain if $\mathscr{L}$. There is an entire theory that allows to use this fact to construct a Markov process from a Dirichlet form. For a detailed treatment, see e.g. the book [9] by Fukushima et al..

Since $\mathscr{L}$ is positive and self-adjoint, it can be written in the form $\mathscr{L} = A^* A$, with A positive, and the Dirichlet form then has the form

$$\mathscr{E}(f,g) = \int \mu(dx) A f(x) A g(x). \tag{6.1.9}$$

6.2 Reversible diffusions

We will now look at reversibility issues in the context of diffusions. The formal adjoint of the operator $\mathscr{L}$ given in (5.1.2) is

$$\begin{aligned} \mathscr{L}^* g(x) &= \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x) g(x) - \sum_i \frac{\partial}{\partial x_i} b_i(x) g(x) \\ &= \frac{1}{2} \sum_{i,j} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} g(x) \\ &\quad + \sum_i \left(\sum_j \frac{\partial a_{ij}(x)}{\partial x_j} - b_i(x) \right) \frac{\partial}{\partial x_i} g(x) \\ &\quad + \left(\frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x) - \sum_i \frac{\partial}{\partial x_i} b_i(x) \right) g(x). \end{aligned} \tag{6.2.1}$$

We can see that this is equal to $\mathscr{L}$ if and only if

$$\sum_j \frac{\partial a_{ij}(x)}{\partial x_j} = 2 b_i(x), \tag{6.2.2}$$

for all $i = 1,\dots d$. Thus (6.2.2) is a condition for the diffusion to be reversible with respect to Lebesgue measure.

Next we may want to look for a reversible measure $\mu(dx) = e^{F(x)}dx$, i.e. a reversible measure that is absolutely continuous w.r.t. Lebesgue measure. This will be the case if

$$(\mathscr{L}^*(ge^F)(x) = e^{F(x)}\mathscr{L}g.$$

But

$$\begin{aligned}(\mathscr{L}^*(ge^F)(x) = & e^{F(x)}\frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 g(x)}{\partial x_i \partial x_j} \\ & + e^{F(x)}\sum_{i,j} a_{ij}(x)\frac{\partial g(x)}{\partial x_i}\frac{\partial F(x)}{\partial x_j} \\ & + e^{F(x)}\frac{1}{2}\sum_{i,j} a_{ij}(x) g(x)\left[\frac{\partial^2 F(x)}{\partial x_i \partial x_j} + \frac{\partial F(x)}{\partial x_i}\frac{\partial F(x)}{\partial x_j}\right] g(x) \\ & + e^{F(x)}\sum_i \left(\sum_j \frac{\partial a_{ij}(x)}{\partial x_j} - b_i(x)\right)\left(\frac{\partial F(x)}{\partial x_i} g(x) + \frac{\partial g(x)}{\partial x_i}\right) \\ & + e^{F(x)}\left(\frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x) - \sum_i \frac{\partial}{\partial x_i} b_i(x)\right) g(x)\end{aligned} \tag{6.2.3}$$

The first condition for reversibility is then that

$$\sum_j \left(a_{ij}(x)\frac{\partial F(x)}{\partial x_j} + \frac{\partial a_{ij}}{\partial x_j}\right) = 2b_i(x). \tag{6.2.4}$$

or

$$2b_i(x) = e^{-F(x)}\sum_j \frac{\partial}{\partial x_j}\left(a_{ij}(x)e^{F(x)}\right). \tag{6.2.5}$$

In particular, in the simples case when $a_{ij}(x) = \delta_{ij}$, we get a necessary and sufficient condition

$$2b_i(x) = \frac{\partial}{\partial x_i}F(x), \tag{6.2.6}$$

i.e. the drift must be the gradient of a potential F (up to the factor 2). In that case the generator takes the very suggestive form

$$\mathscr{L} = \frac{1}{2}e^{-F(x)}\nabla e^{F(x)}\nabla. \tag{6.2.7}$$

The corresponding Dirichlet form then can be written as

$$\mathscr{E}(f,g) = -\int \mu(dx) f(x)(\mathscr{L}g)(x) = \frac{1}{2}\int \mu(dx)\langle \nabla f(x), \nabla g(x)\rangle, \tag{6.2.8}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathbb{R}^d$.

6.3 Equilibrium measure, equilibrium potential, and capacity

In the following we will return to the general case of SDE corresponding to a generator that is a uniformly elliptic differential operator $\mathscr{L}$ with coefficients satisfying Lipschitz conditions (so that unique strong solutions to the SDE exist).

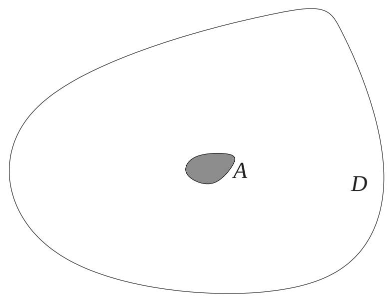

Fig. 6.1 Capacitor

Let D be a open domain in $\mathbb{R}^d$ with $\partial D = A \cup B$, with $A \cap B = 0$. Then the solution of the Dirichlet problem

$$\begin{aligned} \mathscr{L}h(x) &= 0, x \in D \\ h(x) &= 1, x \in A \\ h(x) &= 0, x \in B \end{aligned} \tag{6.3.1}$$

is called the *equilibrium potential* of the *capacitor* (A,B). Recall that, for $x \in D$,

$$h(x) = \mathbb{P}_x[\tau_A < \tau_B]. \tag{6.3.2}$$

Remark 6.6. The boundary conditions here are not continuous, so recall Remark 5.1. We do not assume that D is connected.

Remark 6.7. The names here come from the classical case when $\mathscr{L} = \Delta/2$. Then the Dirichlet problem is a classical problem of electrostatics. The sets A and B correspond to two metal plates attached to a battery that imposes a constant voltage (potential) difference between the plates. The solution of this problem then describes the electrostatic potential (whose gradient is the electrostatic field).

Next we consider the inhomogeneous Dirichlet problem,

$$\begin{aligned} -(\mathscr{L}f)(x) &= g,\ x \in D \\ f(x) &= 0,\ x \in \partial D \end{aligned} \tag{6.3.3}$$

We have seen that, if this problem has a unique solution, then it has the probabilistic representation

$$f(x) = \mathbb{E}_x \int_0^{\tau_D} g(X_t) = \mathbb{E}_x \int_0^{\tau_D} \int_D P_t^D(x,dy) g(y) dt, \tag{6.3.4}$$

where $P_t^D(x,dy)$ is the sub-Markov semi-group associated to the generator $\mathscr{L}^D$ of the process killed when exiting D. Thus we define the *Green kernel*,

$$G_D(x,dy) \equiv \mathbb{E}_x \int_0^{\tau_D} P_t^D(x,dy) dt \tag{6.3.5}$$

in terms of which the solution of (6.3.3) can be written as

$$f(x) = \int_D g(y) G_D(x,dy) \equiv (G_D g)(x). \tag{6.3.6}$$

Note the similarity with the *resolvent* of the semigroup. In fact, one may define

$$G_D^{(\lambda)}(x,dy) \equiv \mathbb{E}_x \int_0^{\tau_D} e^{-\lambda t} P_t^D(x,dy) dt \tag{6.3.7}$$

Then G_D^λ exists for all $\lambda > 0$ even if $\mathbb{E}_x \tau_D = \infty$, and

$$f_\lambda(x) = \int_D g(y) G_D^{(\lambda)}(x,dy) \equiv (G_D^{(\lambda)} g)(x) \tag{6.3.8}$$

solves the Dirichlet problem

$$\begin{aligned} (-\mathscr{L} - \lambda) f_\lambda(x) &= g, x \in D \\ h(x) &= 0, x \in \partial D \end{aligned} \tag{6.3.9}$$

Note that it is of course an interesting question (to which we will return), to ask for which values of λ we can still define $G_D^{(\lambda)}$ for given D.

The Green kernel will often have a density with respect to Lebesgue measure, i.e.

$$G_D(x,dy) = G_D(x,y) dy. \tag{6.3.10}$$

The function $G_D(x,y)$ is then called the *Green function*.

Let us now look at the relation between equilibrium potential and the Dirichlet form in the case of a reversible diffusion. Let us try to compute $\mathscr{E}(h,h)$. One might be tempted to think that $\mathscr{E}(h,h) = 0$, since $\mathscr{L}h(x) = 0$ except on the boundary of the sets A and B. But of course on these, $\mathscr{L}h$ may be singular, since no differentiability assumptions are made on the boundary. So we may interpret $\mathscr{L}h$ as a measure that is concentrated on the boundaries of A and B. Since h vanishes on ∂B, we get that

$$\mathscr{E}(h,h) = -\int_{\partial A} \mu(x) (\mathscr{L}h)(dx). \tag{6.3.11}$$

The measure $(-\mathscr{L}h)(dx)$ is called the *equilibrium measure* associated to the capacitor A,B.

To understand this better, let us return to the case $a_{ij}(x) \equiv \delta_{ij}$. We then have the following *integral formula*, known as the *first Green's formula* .

Theorem 6.8. *Let D be a regular domain and let ϕ, ψ be in $C^2(D)$. Let $\mathscr{L}$ be given by (5.1.2). Then*

$$\int_D dx\, e^{F(x)} \left(\langle \nabla \phi(x), \nabla \psi(x) \rangle - \psi(x)(2\mathscr{L}\phi)(x) \right) = \int_{\partial D} e^{F(x)} \psi(x) \partial_{n(x)} \phi(x)\, d\sigma_D(x), \tag{6.3.12}$$

where $\partial_{n(x)}$ denotes the inner normal derivative at $x \in \partial D$.

Proof. In the case $F = 0$ this formula a classical. The extension to the general case is by a straightforward computation. □

Remark 6.9. An immediate consequence of this identity is the so-called *second Green's formula*,

$$\int_D e^{F(x)} dx \left(\phi(x)(2\mathscr{L}\psi)(x) - \psi(x)(2\mathscr{L}\phi)(x) \right) = \int_{\partial D} e^{F(x)} (\psi(x)\partial_{n(x)}\phi(x) - \phi(x)\partial_{n(x)}\psi(x))\, d\sigma_D(x) \tag{6.3.13}$$

The second Green's formula gives rise to the integral representation of a solution of the Dirichlet boundary value problem,

$$\begin{aligned} -(\mathscr{L}f)(x) &= 0, \quad x \in D, \\ f(x) &= u(x), \quad x \in \partial D, \end{aligned} \tag{6.3.14}$$

in terms of the *Poisson kernel* , namely

$$f(x) = \int_{\partial D} e^{F(y)-F(x)} u(y) \partial_{n(y)} G_D(y,x) d\sigma_D(y). \tag{6.3.15}$$

Using the first Green's formula, we can give a precise relation between equilibrium potential and capacity. Namely, setting $\phi = \psi = h$ in (6.3.14), we see that

$$\int_D dx\, e^{F(x)} \langle \nabla h(x), \nabla h(x) \rangle = \int_A e^{F(x)} \partial_{n(x)} h(x) d\sigma_A(x), \tag{6.3.16}$$

i.e. we have that on A the equilibrium measure, $(-\mathscr{L}h)(x)$ is given by

$$e_{A,B}(dx) \equiv \partial_{n(x)} h(x) d\sigma_A(x). \tag{6.3.17}$$

The quantity

$$\mathrm{cap}(A,B) \equiv \int_A e^{F(x)} \partial_{n(x)} h(x) d\sigma_A(x), \tag{6.3.18}$$

is called the *capacity* of the capacitor A, B. In electrical language, it is the total charge on the plate A. Using relation (6.3.16), we see that alternatively, the capacity is also the total *energy* of the potential h.

Last exit distribution and equilibrium measure.

It will be nice to have a probabilistic interpretation of the equilibrium measure that will at the same time explain why Lh really becomes a surface measure.

We see that, for x in A, we should have something like

$$\begin{aligned} -(\mathscr{L}h)(x) &= \lim_{t\downarrow 0} t^{-1}(1-P_t)(h(x)) \\ &= \lim_{t\downarrow 0} t^{-1}\mathbb{E}_x\left(1-\mathbb{P}_{X_t}\left[\tau_A<\tau_B\right]\right) \\ &= \lim_{t\downarrow 0} t^{-1}\mathbb{E}_x\mathbb{P}_{X_t}\left[\tau_B<\tau_A\right]. \end{aligned} \tag{6.3.19}$$

Let us define the *last exit time*, L_A, from A as

$$L_A \equiv \sup\{0<0<\tau_B : X_t\in A\}, \tag{6.3.20}$$

with the convention $\sup\emptyset = 0$ Note that this is obviously no stopping time and that

$$\mathbb{P}_x[L_A>0] = \mathbb{P}_x[\tau_A<\tau_B] \equiv h(x). \tag{6.3.21}$$

Note that we can write the expression in the last line of (6.3.19) as

$$\mathbb{E}_x\mathbb{P}_{X_t}\left[\tau_B<\tau_A\right] = \mathbb{P}_x[0<L_A<t].$$

Hence we set

$$\psi_t(z) \equiv t^{-1}\mathbb{P}_z\left[0<L_A<t\right]. \tag{6.3.22}$$

Let us also define the *last exit distribution*, $L(x,dy)$, on A, by

$$L(x,dy) \equiv \mathbb{P}_x\left[X_{L_A-}\in dy; L_A>0\right]. \tag{6.3.23}$$

We want to prove the following lemma:

Lemma 6.10. *Let f be a continuous function on $\bar{D}$. Then*

$$\lim_{t\downarrow 0}\int G_D(x,y)\psi_t(y)f(y)dy = \int_A L(x,dy)f(y). \tag{6.3.24}$$

Proof. Without loss let $f\geq 0$. Using the representation of the Green function through the semigroup (6.3.5) we get

$$\begin{aligned}\int G_D(x,y)\psi_t(y)f(y)dy &= \mathbb{E}_x\int_0^{\tau_B}\psi_t(X_s)f(X_s)ds \\ &= t^{-1}\int_0^\infty \mathbb{E}_x\left[f(X_s)\mathbb{P}_{X_s}[0<L<t]\right]ds \\ &= t^{-1}\int_0^\infty \mathbb{E}_x\left[f(X_s)\mathbb{1}_{s<L<s+t}\right]ds \\ &= \mathbb{E}_x\left[0<L_A\le t; t^{-1}\int_0^{L_a} f(X_s)ds\right] \\ &\quad +\mathbb{E}_x\left[t<L_A; t^{-1}\int_{L_A-t}^{L_a} f(X_s)ds\right].\end{aligned} \tag{6.3.25}$$

First, both terms in the last line are obviously uniformly bounded as $t\downarrow 0$. Moreover,

$$\mathbb{E}_x\left[0<L_A\le t; t^{-1}\int_0^{L_a} f(X_s)ds\right]\le C\mathbb{E}_x\left[0<L_A\le t\right]\downarrow 0, \tag{6.3.26}$$

as $t\downarrow 0$. Finally, by continuity of f,

$$\lim_{t\downarrow 0}\mathbb{E}_x\left[t<L_A; t^{-1}\int_{L_A-t}^{L_a} f(X_s)ds\right] = \mathbb{E}_x\left[0<L_A; f(X_{L_A-})ds\right]. \tag{6.3.27}$$

Integrating over A gives the claim of the lemma. □

From Lemma (6.10) one can deduce that the family of measures $\psi_t(y)dy$ converges to a measure $e(dy)$ on A. Moreover, this measure satisfies

$$G_D(x,y)e(dy) = L(x,dy). \tag{6.3.28}$$

Integrating this formula over A, we arrive at the expression

$$\int_A G_D(x,y)e(dy) = \int_A L(x,dy) = h(x). \tag{6.3.29}$$

Hence $e(dy)$ satisfies the defining relation of the equilibrium measure

Thus we have proven a very interesting and useful relation between the equilibrium potential, the equilibrium measure, and the Green function.

Theorem 6.11. *Let as before $A\subset D$ be open sets with smooth boundary. Then, for all $x\in D$,*

$$h(x) = \int_{\partial A} G_D(x,y)e_{A,D}(dy). \tag{6.3.30}$$

Remark 6.12. Is is instructive to think about this result in the following way. We have already seen that we may want to think of $\mathscr{L}h$ as a measure. Then we have that

$$\begin{aligned}-(\mathscr{L}h)(x)dx &= e_{A,D}(dx), \quad x\in D, \\ h(x) &= 0, \quad x\in\partial D.\end{aligned} \tag{6.3.31}$$

Then the solution of this problem in terms of the Green function is precisely the expression (6.3.30). Note that (6.11) holds also in A, with $h(x)=1$.

This formula for the Green function gives of course corresponding formulas for solutions of Dirichlet problems. E.g., if we consider for some function g the Dirichlet problem

$$\begin{aligned} -(\mathscr{L} f)(x) &= g(x), \quad x \in D \\ f(x) &= 0, \quad x \in \partial D, \end{aligned} \tag{6.3.32}$$

then of course $f(x)=\int_D dy G_D(x,y)g(y)$. By symmetry, $G_D(x,y)=e^{F(y)-F(x)}G_D(y,x)$, and so

$$\begin{aligned} \int_D dx e^{F(x)} h(x) g(x) &= \int_D dx h(x) \int_{\partial A} e^{F(x)} g(x) G_D(y,x) e^{F(y)-F(x)} e_{A,D}(dy) \\ &= \int_{\partial A} e^{F(y)} \int_D G_D(y,g) g(x) e_{A,D}(dy) \\ &= \int_{\partial A} e^{F(y)} f(y) e_{A,D}(dy). \end{aligned} \tag{6.3.33}$$

Introducing the probability measure

$$\nu_{A,D}(dy) \equiv \frac{e^{F(y)} e_{A,D}(dy)}{\mathrm{cap}(A,D)}, \tag{6.3.34}$$

on ∂A, this gives

$$\int_{\partial A} \nu_{A,D}(dy) f(y) = \frac{1}{\mathrm{cap}(A,D)} \int_D dx e^{F(x)} h(x) g(x). \tag{6.3.35}$$

As a particular example we get, with $g(x)=1$,

$$\int_{\partial A} \nu_{A,D}(dy) \mathbb{E}_y \tau_D = \frac{1}{\mathrm{cap}(A,D)} \int_D dx e^{F(x)} h(x). \tag{6.3.36}$$

Dirichlet principle

. We have seen that the equilibrium Dirichlet form computed on the equilibrium potential gives the capacity. We will how show that the equilibrium potential is the solution of a variational problem.

Theorem 6.13. *With the notations and assumptions above, the following holds. Let $\mathscr{H}_{\mathrm{A,B}}$ be the space of continuous functions,, f, on $\overline{D}$ such that,*

(i) $\mathscr{E}(f,f,) < \infty$, and
(ii)$f(x) \geq 1$, $x \in A$ and $f(x) \leq 0$, $x \in< D^c$.

$$\text{cap}(A,B) = \inf_{f \in \mathscr{H}_{A,B}} \mathscr{E}(f,f). \qquad (6.3.37)$$

Moreover, if $\mathscr{H}_{A,B} \neq \emptyset$*, the infimum in (6.3.37) is achieved uniquely on the equilibrium potential, i.e.* $\text{cap}(A,B) = \mathscr{E}(h_{A,B}, h_{A,B})$.

Proof. Let us assume that the set $\mathscr{H}_{A,B}$ is not empty. Consider a function g such that $\mathscr{E}(g,g) < \infty$ and that g vanishes on both ∂B and ∂A. Notice that, for $h \in \mathscr{H}_{A,B}$,

$$\mathscr{E}(h+\varepsilon g, h+\varepsilon g) - \mathscr{E}(h,h) = 2\varepsilon \int_{D\setminus \bar{A}\cup B} \mu(dx) g(x)(\mathscr{L}h)(x) + \varepsilon^2 \mathscr{E}(g,g). \qquad (6.3.38)$$

This implies two things: first, if $\mathscr{L}h(x) = 0$, then h is a global minimum of $\mathscr{E}$ in $\mathscr{H}_{A,B}$. We know already that such a function exists, namely the equilibrium potential. Next assume that there is another function, f, such that $\mathscr{E}(f,f)) = \mathscr{E}(h,h)$. Then the identity

$$\mathscr{E}\left(\tfrac{f+h}{2}, \tfrac{f+h}{2}\right) + \mathscr{E}\left(\tfrac{f-h}{2}, \tfrac{f-h}{2}\right) = \frac{1}{2}\mathscr{E}(f,f) + \frac{1}{2}\mathscr{E}(h,h), \qquad (6.3.39)$$

implies that

$$\mathscr{E}\left(\tfrac{f+h}{2}, \tfrac{f+h}{2}\right) \leq \mathscr{E}(h,h) - \mathscr{E}\left(\tfrac{f-h}{2}, \tfrac{f-h}{2}\right). \qquad (6.3.40)$$

Since h is an absolute minimum, this can only hold if

$$\mathscr{E}(f-h, f-h) = 0. \qquad (6.3.41)$$

But this means that $\|\nabla(f-g)(x)\|_2 = 0$ μ-almost surely. $\square$

The Dirichlet principle is a powerful tool for asymptotic computations of capacities, and, hence, as we shall see, much more. To a large extend this is due to the fact that it allows for natural upper and lower bounds. The most immediate one of these is of course given by the elementary observation that

Corollary 6.14. *For any function,* $f \in \mathscr{H}_{A,B}$,

$$\text{cap}(A,B) \leq \mathscr{E}(f,f). \qquad (6.3.42)$$

6.4 The case of dimension one.

The above considerations lead to very explicite answers in the case when $d = 1$. The first observation is that all homogeneous boundary value problems in this case can, by linearity, be reduced to computing the equilibrium potential for on interval (a,b),i.e.

$$\begin{aligned}(\mathscr{L}h)(x) &= 0, \quad x \in (a,b) \\ h(a) &= 0 \\ h(b) &= 1\end{aligned} \tag{6.4.1}$$

Note also that the general uniformly elliptic case, we can be reduced to the problem with generator

$$(\mathscr{L}h)(x) = -\frac{1}{2}f''(x) + b(x)f'(x). \tag{6.4.2}$$

Note also that, in $d = 1$, any bounded function b can be written as a derivative of another function, $F/2$, where

$$F(x) = 2\int_0^x b(x)dx. \tag{6.4.3}$$

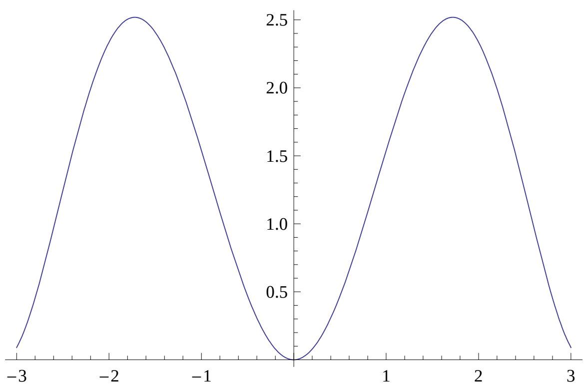

Fig. 6.2 A potential function on $[-3,3]$

Thus we are always in the reversible case. Hence we are reduced to solving the ordinary differential equation

$$\frac{1}{2}h''(x) + b(x)h'(x) = 0, \tag{6.4.4}$$

which in turn reduces to the first order equation

$$\frac{1}{2}u'(x) + b(x)u(x) = 0 \tag{6.4.5}$$

when we set $u = h'$. Clearly (6.4.5) has the general solution

$$u(x) = C_1 e^{-F(x)} \tag{6.4.6}$$

and so the general solution of (6.4.4) is

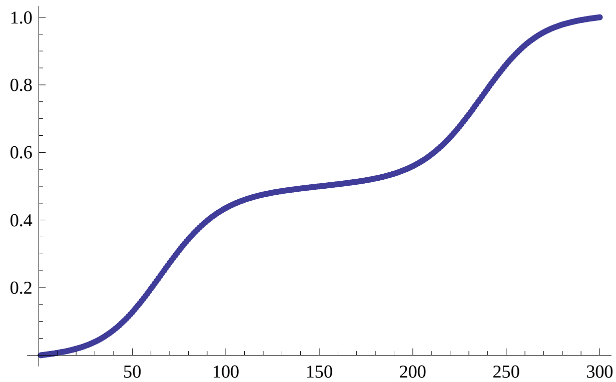

Fig. 6.3 The corresponding equilibrium potential $P_x[\tau_3 < \tau_{-3}]$

$$h(x) = C_1 \int_0^x e^{-F(y)} dy + C_2, \tag{6.4.7}$$

with C_1 and C_2 integration constants to be determined from the boundary conditions. In particular, for the equilibrium potential related to the interval (a,b) we have

$$h(x) = \frac{\int_a^x e^{-F(y)} dy}{\int_a^b e^{-F(y)} dy}. \tag{6.4.8}$$

Hence the capacity $\mathrm{cap}(a,b)$ is readily computed as

$$\mathrm{cap}(a,b) = \mathscr{E}(h,h) = \frac{1}{2\int_a^b e^{-F(y)} dy}. \tag{6.4.9}$$

Some reflection shows that we can get from (6.11) the following formula for the Green function in (a,b): For $x < y$,

$$\begin{aligned} G_{(a,b)}(x,y) &= \frac{h_{x,\{a,b\}}(y)}{e_{x,b}} \\ &= {}^{F(x)} \frac{1 - h_{x,b}(y)}{\mathrm{cap}(x,b)} \\ &= e^{F(x)} \frac{\frac{\int_y^b e^{-F(z)} dz}{\int_x^b e^{-F(z)} dz}}{\frac{1}{2\int_x^b e^{-F(z)} dz}} \\ &= \frac{1}{2} e^{F(x)} \int_y^b e^{-F(z)} dz \end{aligned} \tag{6.4.10}$$

and for $y < x$,

$$G_{(a,b)}(x,y) = \frac{1}{2} e^{F(y)} \int_a^y e^{-F(z)} dz. \tag{6.4.11}$$

6.5 Another view on one-dimensional diffusions

We have seen in the previous section that the computation of the equilibrium potential in one-dimensional case allows to compute the Green function and hence to essentially solve everything that can be expressed in terms of Dirichlet problems.

We will now take a different look at the same issue. The perspective will be more on the level of the process. We will see that the solution of a 1d SDE can be constructed from Brownian motion in a way that will exhibit again the crucial rôle.

Let us recall that the harmonic functions we encountered can be written in the form

$$\mathbb{P}_x(\tau_a < \tau_b) = \frac{s(x) - s(a)}{s(b) - s(a)}, \tag{6.5.1}$$

where $s(x)$ is an increasing function whose derivative is $e^{-F(x)}$. The function s is usually called the *scale function* Recall that in the case of Brownian motion, $s(x) = x$. Now let B be Brownian motion and consider the process $Y_t = s^{-1}(B_t)$. Clearly we have that

$$\mathbb{P}^Y_x(\tau_a < \tau_b) = \mathbb{P}^B_{s(x)}(\tau_{s(a)} < \tau_{s(b)}) = \frac{s(x) - s(a)}{s(b) - s(a)}, \tag{6.5.2}$$

(here the superscripts indicate that the probabilities are w.r.t. the corresponding processes) hence the process has the same harmonic function as the one solving $dX_t = \frac{1}{2}F'(X_t)dt + dB_t$. Is it the same process? No, but using Itô's formula, we see that $Z_t \equiv s(X_t)$ satisfies

$$\begin{aligned} dZ_t &= s'(X_t)dX_t + \frac{1}{2}s''(X_t)dt \\ &= s'(X_t)dB_t + \frac{1}{2}(s'(X_t)F'(X_t) + s''(X_t))dt \\ &= s'\left(s^{-1}(X_t)\right)dB_t, \end{aligned} \tag{6.5.3}$$

which is of the form

$$dZ_t = g(Z_t)dB_t. \tag{6.5.4}$$

We will show that any solution of an SDE of the form (6.5.4) is a time change of Brownian motion.

Theorem 6.15. *Let g be a measurable function such that $g(x) \geq \delta > 0$. Then (6.5.4) has a unique weak solution. Moreover, there exists a Brownian motion, B, such that*

$$Z_t = B(\gamma_t), \tag{6.5.5}$$

where $\gamma_t \equiv \inf\{u : A(u) > t\}$, and

$$A(t) \equiv \int_0^t g(B(u))^{-2} du. \tag{6.5.6}$$

Proof. Clearly we have that th if Z solves (6.5.4), then

$$d[Z]_t = g(Z_t)^2 dt. \tag{6.5.7}$$

On the other hand,

$$[B]_{[Z]_t} = [Z]_t. \tag{6.5.8}$$

Inverting this relation we may write

$$B(t) = Z_{\tau(t)}, \tag{6.5.9}$$

where $\tau(t)$ is the inverse time change, i.e.

$$[Z]_{\tau(t)} = \int_0^{\tau(t)} g(Z_u)^2 du = t. \tag{6.5.10}$$

Differentiating this latter relation we get

$$1 = g(Z_{\tau(t)})^2 \tau'(t) = g(B(t))^2 \tau'(t). \tag{6.5.11}$$

Hence

$$\tau'(t) = \frac{1}{g(B(t))^2}, \tag{6.5.12}$$

and hence

$$\tau(t) = \int_0^t \frac{1}{g(B(u))^2} du, \tag{6.5.13}$$

τ being the inverse of the time change, we see that the time change $[Z]_t$ is really the inverse of the function τ, which is given purely in terms of the Brownian motion B.
□

It remains to generalize the first part of our construction. Thus consider the general form of the SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{6.5.14}$$

with $\sigma > 0$. We know already that the equilibrium potential will be of the form (6.5.1) with s being an integral of

$$s'(x) = e^{-\int_0^x 2b(z)/\sigma^2(z)dz}. \tag{6.5.15}$$

Again $Z_t = s(X_t)$ then has the same harmonic function as Brownian motion, and the same calculation as in (6.5.3) shows that Z_t is a solution of (6.5.4), this time with

$$g(x) = s'(s^{-1}(x))\sigma(s^{-1}(x)). \tag{6.5.16}$$

We summarize these results in the following theorem.

Theorem 6.16. *Assume that $\sigma(x) > 0$, as long as $\int_0^x b(z)\sigma^{-2}(z) < \infty$ exists for $x \in I$, then the SDE has a unique weak solution given by*

$$X_t = s^{-1}\left(B(\gamma_t)\right), \tag{6.5.17}$$

where γ_t is the continuous inverse of the function

$$A_t \equiv \int_0^t \frac{1}{g(B(u))^2} du, \tag{6.5.18}$$

where g is given by (6.5.16), and B is Brownian motion.

The strong point of this result is that very little regularity is required for the drift or difussivity. For example, this theorem allows to make sense of the *Brownian motion in a Brownian potential*: Let $W(t)$ be a realization of a Brownian motion, and consider the formal expression

$$dX_t = W'(X_t)dt + dB_t. \tag{6.5.19}$$

Since W is not differentiable, this expression is formal. However, the corresponding potential, W, is well defined, and so is the scale function $s(x) = \exp(W(x))$. Thus we can interpret the process obtained from Eqs. (6.5.17)-(6.5.18) as the solution of (6.5.19).

6.6 Brownian local time and speed measures

The aim of this section is to give an alternative representation of the time change formula that will give rise to the possibility to construct even larger classes of one dimensional diffusion processes. At the same time we will deepen the discussion of *local time* that was initiated in Theorem 4.9. The following discussion draws on lecture notes by Steve Lalley.

Let us first observe what would be the natural notion of a occupation measure. Let $A \in \mathscr{B}(\mathbb{R})$ be a Borel subset of the real line. Then we can introduce

$$\Gamma_t(A) \equiv \int_0^t \mathbb{1}_A(B_s)ds \tag{6.6.1}$$

as a random measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. The main result we will use and need, is the following theorem.

Theorem 6.17. *With probability one, for each $t < \infty$, the occupation measure Γ_t is absolutely continuous with respect to Lebesgue measure, and it density, l_t^x, is jointly continuous in t and x.*

l_t^x is called the *local time* of Brownian motion at x. We have already seen that the local time at zero can be represented as a stochastic integral via an extension of Itô's formula (Tanaka's formula). This gives the representation

$$l_t^a \equiv |B_t - a| - |B_0 - a| - \int_0^t \operatorname{sign}(B_s - a)dB_s. \tag{6.6.2}$$

We will first show that

Theorem 6.18. *There exists a version of the process* $\{l_t^a, a \in \mathbb{R}, t \in \mathbb{R}_+\}$ *that is jointly continuous in* t *and* a.

Proof. We will deal with a fixed time horizon $T < \infty$. Define

$$\xi_1(a,t) \equiv \int_0^t \operatorname{sign}(B_s - a) dB_s \tag{6.6.3}$$

and

$$\xi_2(a,t) \equiv |B_t - a| - |B_0 - a|. \tag{6.6.4}$$

Obviously ξ_2 is jointly continuous, since B is continuous. Thus we need to prove that ξ_1 has a jointly continuous version. The tool to prove this is a lemma due to Kolmogorov.

Lemma 6.19. *Let* ξ *be a stochastic process indexed by* $\mathbb{R}^n$ *with values in a complete metric space with metric* ρ. *If there exist positive constants,* $\alpha, \beta, \varepsilon$, *such that*

$$\mathbb{E}\left(\rho(\xi_x, \xi_y)\right)^\alpha \leq \beta |x-y|^{n+\varepsilon}, \tag{6.6.5}$$

for all $x, y \in \mathbb{R}^n$, *then there exists a continuous version of* X.

We apply this theorem to the process $\xi_1(\cdot, t)$, $t \in [0, T]$. We may also consider the process on a bounded interval. Then it will be enough to show that

$$\begin{aligned} \mathbb{E}\,|\xi_1(x,t) - \xi_1(y,t)|^p &\leq C|x-y|^{2+\delta} \\ \mathbb{E}\left|\xi_1(x,t) - \xi_1(x,t')\right|^p &\leq C|t-t'|^{2+\delta} \end{aligned} \tag{6.6.6}$$

Now

$$\begin{aligned} |\xi_1(x,t) - \xi_1(y,t)| &= \left|\int_0^t (\operatorname{sign}(B_s - a) - \operatorname{sign}(B_s - b)) dB_s\right| \\ &\leq 2\int_0^t \mathbb{1}_{(x,y)}(B_s) dB_s. \end{aligned} \tag{6.6.7}$$

Hence, using the Burkholder inequalities,

$$\begin{aligned} \mathbb{E}\,|\xi_1(x,t) - \xi_1(y,t)|^{2m} &\leq 2^{2m}\mathbb{E}\left|\int_0^t \mathbb{1}_{(x,y)}(B_s) dB_s\right|^{2m} \\ &\leq C_m 2^{2m}\mathbb{E}\left|\int_0^t \mathbb{1}_{(x,y)}(B_s) ds\right|^{2m}. \end{aligned} \tag{6.6.8}$$

In the final expression we can now estimate

$$\mathbb{E}\left|\int_0^t \mathbb{1}_{(x,y)}(B_s)ds\right|^{2m} \tag{6.6.9}$$
$$= m!\int_{0\leq t_1\leq t_{m-1}\leq\cdots\leq t_m\leq t} \mathbb{P}\left[B_{t_1}\in(x,y),\ldots,B_{t_m}\in(x,y)\right]$$
$$\leq m!\int_{0\leq t_1\leq t_{m-1}\leq\cdots\leq t_m\leq t} \mathbb{E}\Big[P_{B_0}(B_{t_1}\in(x,y)]P_{B_{t_1}}(B_{t_2}\in(x,y))\ldots$$
$$\ldots P_{B_{t_{m-1}}}(B_{t_m}\in(x,y))\Big]$$
$$\leq m!(y-x)^m\sqrt{t}^m.$$

The corresponding estimate for different times is similar. □

With this absolutely continuous local time process we can of course write the time change function (6.5.18) in the form

$$A_t = \int_0^t \frac{1}{g(B_u)^2}du = \int dx \frac{1}{g(x)^2}\int_0^t \delta_{B_u}(x)du = \int m(dz)l_t^z, \tag{6.6.10}$$

where $m(dz) = g^{-2}(z)dz$ and l_t^z is the density of the Brownian local time process. The measure $m(z)$ is called he *speed measure*(essentially it tells us how the local time of Brownian motion is transformed to the time of the new process in the point z). . This formulation gives rise to an even wider class of one-dimensional diffusions that can be constructed as time changes of Brownian motion through more general speed measures. Note that these processes are all Markovian, which is not a trivial fact (see [15], Chapter III.21 and [14], Chapter V.47).

6.7 The one-dimensional trap model and a singular diffusion

In the following we show that these processes are not totally hypothetical, but that they can arise from more or less reasonable discrete models.

In the following we will give the construction of a random motion in a random environment that was studied by Fontes, Isopi, and Newman some years ago [7].

We begin be prescribing a random environment on $\mathbb{Z}$ as a family of iid random variables, τ_i, $i\in\mathbb{Z}$, whose distribution will be be assumed to satisfy

$$\lim_{t\uparrow\infty} t^{\alpha}\mathbb{P}[\tau_1 > t] = 1 \tag{6.7.1}$$

for $\alpha < 1$. Note that this implies in particular that $\mathbb{E}\tau_1 = +\infty$. Our next ingredient will be a continuous time, unbiased simple random walk, Z_k, $k\in\mathbb{N}$, on $\mathbb{Z}$.

(Note that a continuous time random walk can be described as follows: Let Y_k, $k\in\mathbb{N}$, be a discrete time simple random walk on $\mathbb{Z}$ (i.e. $Y_k = \sum_{i=1}^k u_i$, where u_i are iid with $\mathbb{P}[u_i = \pm 1] = \frac{1}{2}$). Let $C(k) = \sum_{i=0}^{k-1} e_i$, where e_i, $i\in\mathbb{N}$, are iid exponential random variables with rate 1. Then

$$Z_t = Y_{C^{-1}(t)} \tag{6.7.2}$$

where C^{-1} is the inverse of C,

$$f^{-1}(t) = \inf\{k : f(k) > t\}. \tag{6.7.3}$$

We will now construct a continuous time process a time change of the simple radom walk Z as follows. Define the so-called *clock process*

$$S(u) \equiv \int_0^u \tau_{Z_r} dr \tag{6.7.4}$$

The process X is then defined, for a given realization of the random variables τ_i, as

$$X_t = Z_{S^{-1}(u)} \tag{6.7.5}$$

We now want to re-write the clock process in terms of a speed measure. For this we define the local time process of Z as

$$L(j,t) = \int_0^t \mathbb{1}_{Z_u = j} du. \tag{6.7.6}$$

Then we can re-write the clock process as

$$S(u) = \sum_{j \in \mathbb{Z}} \tau_j L(j,u). \tag{6.7.7}$$

One sees easily that there is a complete analogy between the construction of a diffusion from Brownian motion.

We now consider a rescaling of space and time to obtain a continuous process limit. Clearly we have from (a complete analog of) Donsker's invariance principle that

$$\lim_{\varepsilon \downarrow 0} \varepsilon Z_{t/\varepsilon^2} = B_t. \tag{6.7.8}$$

Now assume that for some β,

$$\varepsilon^2 S^{-1}(\varepsilon^{-\beta} t) \equiv S_\varepsilon^{-1}(t) \to \Sigma^{-1}(t), \tag{6.7.9}$$

then

$$X_t^\varepsilon \equiv \varepsilon X_{\varepsilon^{-\beta} t} = \varepsilon Z_{\varepsilon^{-2} S_\varepsilon^{-1}(t)} \equiv Z^\varepsilon_{S_\varepsilon^{-1}(t)} \tag{6.7.10}$$

and we may expect that

$$Z^\varepsilon_{S_\varepsilon^{-1}(t)} \to B_{\Sigma^{-1}(t)}. \tag{6.7.11}$$

The question is thus to see whether and to what the process $\varepsilon^2 S^{-1}(\varepsilon^{-\beta} t)$, respectively its inverse,

$$S_\varepsilon(u) \equiv \varepsilon^\beta S(u/\varepsilon^2), \tag{6.7.12}$$

converges. Now

$$\varepsilon^{\beta} S(u/\varepsilon^2) = \sum_{i\in\mathbb{Z}} \varepsilon^{\beta} \tau_i L(i, u/\varepsilon^2) \equiv \sum_{i\in\mathbb{Z}} \varepsilon^{\beta} \tau_i L_{\varepsilon}(\varepsilon i, u), \tag{6.7.13}$$

where by definition $L_\varepsilon(\varepsilon i, u) = L(i, u/\varepsilon^2)$.We may expect this to converge to the local time process of Brownian motion. On the other hand, we can think of the sum as an integral over the random measure

$$m_\varepsilon(dx) \equiv \sum_{i\in\mathbb{Z}} \delta_{i\varepsilon}(dx)\tau_i\varepsilon^{\beta}, \tag{6.7.14}$$

i.e.

$$S_\varepsilon(t) = \int m_\varepsilon(dx) L_\varepsilon(x,t), \tag{6.7.15}$$

It is a curious fact that in distribution,

$$\int m_\varepsilon(dx) L_\varepsilon(x,t) = \int m_\varepsilon(dx) l_t^x \tag{6.7.16}$$

This is due to the fact that the local time density of Brownian motion on an integer point i before visiting one of its neighbors is an exponential random variable with mean one. To see this, observe that the continuous time simple random walk on $\mathbb{Z}$ can be coupled to a Brownian motion: consider the measure $m_0(dx) \equiv \sum_{i\in\mathbb{Z}} \delta_i(dx)$ as a speed measure and let $\widetilde{Y}_t$ be the time change of a Brownian motion B_t with this speed measure. This is a Markov process that spends all of its time on the integers and jumps with infinite speed between them. It is clear that this process visits the sites $i \pm 1$ with equal probability after i. Moreover, form the fact that the process observed on the sites i is Markov, the waiting time at i before the process reaches $i \pm 1$ is exponential. Its mean value is given by $\mathbb{E}_0 \ell^0_{\tau_1 \wedge \tau_{-1}}$. Using Tanaka's formula for the local time of Brownian motion, we get

$$\mathbb{E}_0 \ell^0_{\tau_1 \wedge \tau_{-1}} = \mathbb{E}_0 \left| B_{\tau_1 \wedge \tau_{-1}} \right| = 1. \tag{6.7.17}$$

Thus we see that we have indeed a relisation of the process Y as timechange of Brownian motion.

Thus we have actually immediately an expression of our (rescaled) process X^ε immediately as a time change of Brownian motion with speed measure m_ε.

Thus the key question is whether $m_\varepsilon(dx)$ converges. this will be the case, due to (6.7.1), if $\beta = 1/\alpha$. This follows from a more general result about the convergence of so-called extremal processes.

(for a proof see, e.g., [13]):

Theorem 6.20. *Assume that X_i are iid random variables that satisfy*

$$\lim_{n\uparrow\infty} \varepsilon^{-1} \mathbb{P}[X_i > u_\varepsilon(c)] = \nu(c). \tag{6.7.18}$$

where ν is an increasing (respectively decreasing) function. Then, the point process

$$\sum_{i\in\mathbb{Z}} \delta_{(i\varepsilon, u_\varepsilon^{-1}(X_i))} \tag{6.7.19}$$

converges in distribution to the Poisson point process, $\mathscr{R}$ *on* $\mathbb{R}_+ \times \mathbb{R}$ *with intensity measure* $dt \times d\nu(x)$ *(respectively* $-d\nu$ *if* ν *is decreasing).*

Using the property (6.7.1), we see that in our case, with $u_\varepsilon(c) \equiv \varepsilon^{-1/\alpha}c$, we have that

$$\varepsilon^{-1}\mathbb{P}[\tau_1 > \varepsilon^{-1/\alpha}c] = c^{-\alpha}[\varepsilon^{-1/\alpha}c]^{\alpha}\mathbb{P}[\tau_1 > \varepsilon^{-1/\alpha}c] \to c^{-\alpha}.$$

Thus the theorem yields

Corollary 6.21. *The point process*

$$R_\varepsilon \equiv \sum_{i\in\mathbb{Z}} \delta_{(i\varepsilon,\varepsilon^{1/\alpha}\tau_i)} \to \mathscr{R} \tag{6.7.20}$$

converges to the Poisson point process on $\mathbb{R} \times \mathbb{R}_+$ *with intensity measure* $dt \times \alpha c^{-1-\alpha}dc$.

One can show that this implies that, if $\alpha < 1$, the measures

$$m_\varepsilon(dx) = \int R_\varepsilon(dx,dt)t \tag{6.7.21}$$

converge to the measure

$$m(dx) \equiv \int R(dx,dt)t. \tag{6.7.22}$$

Note of course that we are speaking of random measures here. So what converges is the distribution of these random measures. In proper terms, we would have to equip the space of measures with a topology (e.g. the vague topology) and speak of *weak convergence* of the family of random measures with respect to this underlying topology.

One can easily check that the measure $m(dx)$ is singular (in fact it is a pure point measure) with respect to Lebesgue's measure. Nonetheless one can use it to construct a singular diffusion as a time change of Brownian motion form it, that will be the natural candidate for the limit process in our model.

It is known that if a sequence of (point) measures, ν_n, converges to a point measure ν (in a suitable topology that I will not discuss here), then the corresponding time-changed processes converge to the process with time change obtained from the speed measure ν.

Can we apply this fact in our case, when the measures μ_ε converge only in weakly? The answer is yes, in general due to Skorohod's theorem, that states that weak convergence of a family of random variables, X_n, is equivalent to the existence of a another family, $\overline{X}_n$, such that for each n, X_n and $\overline{X}_n$ have the same distribution, while $\overline{X}_n$ converges almost surely.

A coupling

. It is an amusing observation that in the case of our random measures m_ε, this construction can be made in a very explicit way. It will also exhibit a deep relation between these measures and Lévy processes.

Let us first briefly recall what an α-stable Lévy subordinator, U, is. There are in fact at least two ways to describe it: one is to say that U is a non-decreasing stationary process with independent increments whose Laplace transform is given by

$$\mathbb{E}e^{-\lambda U(x)} = \exp\left[x\alpha\int_0^\infty (e^{-\lambda w}-1)w^{-1-\alpha}dw\right]. \tag{6.7.23}$$

Another way to characterize it is to say that it is the distribution function on $\mathbb{R}$ associated with the measure $m(dx)$, normalized s.t. $U(0)=0$ (see my lecture notes on ageing [2]).

Now introduce the scaling function G such that

$$\mathbb{P}[U(1)\leq G(a)] = \mathbb{P}[\tau_0\leq a]. \tag{6.7.24}$$

Then define

$$\tau_i^\varepsilon \equiv G^{-1}\left(\varepsilon^{-1/\alpha}\left(U(\varepsilon(i+1)-U(\varepsilon i)\right)\right). \tag{6.7.25}$$

Lemma 6.22. *The family of random variables τ_i^ε, $i\in\mathbb{Z}$ is iid and τ_i^ε has the same distribution as τ_1.*

Proof. The proof of this lemma follows form the fact that the subordinator U is α-stable, i.e. that $\varepsilon^{-1/\alpha}U(\varepsilon)$ has the same law as $U(1)$. □

Using these random variables we can construct measures

$$\bar{m}_\varepsilon \equiv \sum_{i\in\mathbb{Z}} \varepsilon^{1/\alpha}\tau_i^\varepsilon \delta_{\varepsilon i}, \tag{6.7.26}$$

which now converge almost surely to m, where of course the the distribution function of this m is used as U in the construction of the τ_i^ε. In the case when G is the identity, this is quite straightforward, whereas in the general case some care is required to show this fact (see [7]). The key is that by the assumptions on the law of τ_i, G is close to linear at infinity and so $\varepsilon^{1/\alpha}G^{-1}(\varepsilon^{-1/\alpha}x)\to x$, as $\varepsilon\downarrow\infty$.

The existence of a non-trivial scaling limit for this model have far-reaching consequences for its long time asymptotics. In particular, it implies so-called *aging behavior*. This notion refers to the long-time behavior of of certain *correlation functions* of the process, e.g.

$$R(t_w,t)\equiv \mathbb{P}[X_{t+t_w}=X_{t_w}]. \tag{6.7.27}$$

One says that a process shows aging, if

$$\lim_{t_w\uparrow\infty} R(t_w,\theta t) = f(\theta), \tag{6.7.28}$$

for some non-trivial function f. Now in our case we have that

$$R(t_w, \theta t) = \mathbb{P}[X_{t_w(1+\theta} = X_{t_w}] = \mathbb{P}[X_{1+\theta}^{t_w^{-1/\alpha}} = X_1^{t_w^{-1/\alpha}}]. \tag{6.7.29}$$

Again we may expect this to converge to

$$\mathbb{P}\left[B_{\Sigma(1+\theta)} = B_{\Sigma(1)}\right], \tag{6.7.30}$$

which would be our desired *ageing function* expressed in term of the limiting process.

Chapter 7
Appendix: 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.

7.1 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 $\mathscr{M}_1(J)$ 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 7.1. *[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 7.2. *[Riesz representation theorem] Let ϕ be a positive linear functional $\phi : C(J) \to \mathbb{R}$ with $\Phi(1) = 1$. Then there exists a unique inner regular probability measure, $\mu \in \mathscr{M}_1(J)$, such that*

$$\phi(f) = \mu(f) = \int_J f d\mu. \tag{7.1.1}$$

Recall (see [1] 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

$$B_{f_1,\dots,f_n,\varepsilon}(\phi_0) \equiv \{\phi \in C(J)^* : \forall_{1\leq i\leq n} |\phi(f_i) - \phi^0(f_i)| < \varepsilon\} \tag{7.1.2}$$

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 7.3. 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 7.4. *A net ϕ_α in $C(J)^*$ converges in the weak-* topology to some element, ϕ, 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 ε, there exists α_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 α_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 7.5. *The unit ball*

$$\{\phi \in C(J)^* : \|\phi\| \leq 1\} \tag{7.1.3}$$

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 7.6. *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.

Corollary 7.7. *The set of probability measures on a compact metrisable space is compact in the weak-$*$ topology.*

Proof. By Theorem 1.2.6 in [1], any probability measure on a compact metrisable space is inner regular, hence the restriction to inner regular measures in Corollary 7.6 can be dropped in this case.

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

Theorem 7.8. *Let J be a compact metrisable space. Then $C(J)$ is separable, and $\mathscr{M}_1(J)$ equipped with the weak-$*$ topology is compact metrisable.*

Proof. We may take J to be metric with metric ρ. 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 $\Phi : \mathscr{M}_1(J) \to V \equiv \times_{n\in\mathbb{N}}[-\|f_n\|_\infty, \|f_n\|_\infty]$, 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 μ_α converges to μ (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, $\mathscr{M}_1(J)$ is homeomorphic to a compact subset of this space. Thus it is compact and metrisable.

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

$$\hat{\rho}(\mu, \nu) \equiv \sum_{n=1}^{\infty} 2^{-n} \left(1 - e^{-|\mu(f_n) - \nu(f_n)|}\right). \tag{7.1.4}$$

7.2 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

$$W \equiv C([0,\infty),\mathbb{R}). \tag{7.2.1}$$

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

Lemma 7.9. *The space W equipped with the topology of uniform convergence on compact sets is a Polish space. The σ-algebra, $\mathscr{A}$, of cylinders generated by the projections $\pi_t : W \to \mathbb{R}$, $\pi_t(w) = w(t)$, is the Borel-σ-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 π_t are continuous, and hence $\mathscr{A} \subset \mathscr{B}(W)$. 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 ρ_n and hence ρ are $\mathscr{A}$-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 $F \in \mathscr{A}$, and thus $\mathscr{A} = \mathscr{B}(W)$.

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 7.10. *A topological space is Polish, if and only if it is homeomorphic to a G_δ 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_δ subset of the compact metrisable space $J \equiv [0,1]^{\mathbb{N}}$. Let ρ 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 α 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

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

Therefore, for all k,

$$\limsup \hat{\rho}(x(n),x) \leq 2\hat{\rho}(x_k,x). \tag{7.2.2}$$

Now take a sequence of x_k that converges to x. Then (7.2.2) 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 α^{-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 α 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 7.11. Let S be a Lousin space. Let $C_b(S)$ be the space of bounded, continuous functions on S, and let $\mathscr{M}_1(S)$ be the space of probability measures on S. Then a net, $\mu_\alpha \in \mathscr{M}_1(S)$ converges *weakly* to $\mu \in \mathscr{M}_1(S)$, if and only if, for all $f \in C_b(S)$,

$$\mu_a(f) \to \mu(f). \tag{7.2.3}$$

Weak convergence is related to convergence in probability.

Lemma 7.12. *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 μ_n, μ 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

$$\begin{aligned} |\mu_n(f) - \mu(f)| &= |\mathbb{E}(f(X_n) - f(X))| \\ &\leq \left|\mathbb{E}\left[(f(X_n) - f(X))\mathbb{1}_{\rho(X_n - X) \leq \varepsilon}\right]\right| \\ &\quad + \left|\mathbb{E}\left[(f(X_n) - f(X))\mathbb{1}_{\rho(X_n - X) > \varepsilon}\right]\right| \\ &\leq \delta + C\mathbb{P}(\rho(X_n - X) > \varepsilon) \end{aligned} \tag{7.2.4}$$

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 7.13. *Let μ_α be a net of elements of $\mathscr{M}_1(S)$ 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 $B \in \mathscr{B}(S)$ 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 $\mathscr{M}_1(S)$ is metrisable if S is Lousin. This is very convenient, and in particular will allow us to never use nets anymore!

Theorem 7.14. *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]) μ on B on J such that $\hat{\mu}(J\backslash B) = 0$. The map $\mu \to \hat{\mu}$ is a homeomorphisms from $\mathscr{M}_1(S)$ to the set $\{\nu \in \mathscr{M}_1(J) : \nu(B) = 1\}$ in the weak topologies. Therefore, the weak topology on $\mathscr{M}_1(S)$ is metrisable.*

Proof. We must show that, if μ_α is a net in $\mathscr{M}_1(S)$ and $\mu \in \mathscr{M}_1(S)$, 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 $f_B = f 1\!\!1_B$. 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 7.13,

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

Hence again by Proposition 7.13, (i) holds.

Now that we have shown that the space $\mathscr{M}_1(S)$ is homeomorphic to a subspace of the compact metrisable space $\mathscr{M}_1(J)$ (because of Theorem 7.2.1), $\mathscr{M}_1(S)$ 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

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

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 7.15. 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 7.16. 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, μ_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, $\mu \in \mathscr{M}_1(S)$, 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 7.17. A subset, $H \subset \mathscr{M}_1(S)$ is called *tight*, if and only if there exists, for any $\varepsilon > 0$, a compact set $K_\varepsilon \subset S$, such that, for all $\mu \in H$,

$$\mu(K_\varepsilon) > 1 - \varepsilon. \tag{7.2.5}$$

Theorem 7.18 (Prohorov). *If S is a Lousin space, then a subset $H \subset \mathscr{M}_1(S)$ is* conditionally compact, *if it is tight.*

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

Moreover, since the spaces $\mathscr{M}_1(S)$ 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 ϕ be a homeomorphiosm $\phi : \Sigma \to B \subset J$, for some Borel set B. We know that $\mathscr{M}_1(J)$ 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, $\hat{\mu}_n \in \mathscr{M}_1(J)$ has limit points in $\mathscr{M}_1(J)$. Now let $H = \{\mu_n, n \in \mathbb{N}\} \subset \mathscr{M}_1(S)$ 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 μ_n. For this we need to show that $\hat{\mu}(B) = 1$. Now let K_ε be the compact set in S such that $\mu_n(K_\varepsilon) > 1 - \varepsilon$. Then, by Proposition 7.13,

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

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 (7.2.1).

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

Theorem 7.19. *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*

$$\Delta(\delta,N.w) \equiv \sup\{|w(t)-w(s)| : t,s \in [0,N], |t-s| < \delta\}. \tag{7.2.6}$$

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

The Arzelà–Ascoli theorem allows us to describe conditionally compact sets in W.

This allows us to formulate the following tightness-criterion.

Theorem 7.20. *A subset, $H \subset \mathscr{M}_1(W)$, is conditionally compact (equiv. tight), if and only if:*

(i) $\lim_{c\uparrow\infty} \sup_{\mu\in H} \mu(|w(0)| > c) = 0$*;*
(ii)for all $N \in \mathbb{N}$ and all $\varepsilon > 0$, $\lim_{\delta\downarrow 0} \sup_{\mu\in H} \mu(\Delta(\delta,N,w) > \varepsilon) = 0$*, where Δ is defined in (7.2.6)*

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, Γ_ε, 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 ε, 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.

The continuity module $\Delta(\delta,N,w)$ looks difficult to use due to the appearance of the supremum over t,s. The following proposition gives moment condition that is easier to use.

Proposition 7.21. *Let μ be the law of a continuous stochastic processes X on Wiener space. Then conditions (i) and two of Theorem 7.20 can be replaced by the conditions*

(i)' $\sup_{X\in H}\mathbb{E}|X_0|^\nu<\infty$,
(ii') *For all* $N\in\mathbb{N}$, $\sup_{X\in H}\mathbb{E}|X_t-X_s|^\alpha\leq C_N|t-s|^{1+\beta}$, *for all* $0\leq s,t\leq N$,

for some $\alpha,\beta,\nu>0$ *and* $C_N<\infty$.

Proof. The assertion concerning condition (i) follows trivially. The intresting part is to show that (ii') implies (ii). We fix N and for simplicity set $N=1$. By Chebychev's inequality, we get from (ii) that

$$\mathbb{P}\left[|X_t-X_s|\geq\varepsilon\right]\leq C\varepsilon^{-\alpha}|t-s|^{1+\beta}. \tag{7.2.7}$$

Now take a dyadic sequence of time $t_k^n=k2^{-n}$ and set $\varepsilon_n=e^{-\gamma n}$ for some $\gamma>0$. Then

$$\mathbb{P}\left[|X_{t_k^n}-X_{t_{k-1}^n}|\geq\varepsilon_n\right]\leq C2^{-n(1+\beta-\alpha\gamma)}. \tag{7.2.8}$$

Thus a trivial estimate shows that for any n,

$$\mathbb{P}\left[\max_{k=1}^{2^n}\left|X_{t_k^n}-X_{t_{k-1}^n}\right|\geq\varepsilon_n\right]\leq C2^{-n(\beta-\alpha\gamma)}. \tag{7.2.9}$$

Note that this is exponentially small and hence summable provided that we choose $\gamma<\beta/\alpha$, which is always possible. Therefore, by the Borel-Cantelli lemma, with probability one, the events in (7.2.9) happen only for finitely many n. Let us call the last such value $n^*(\omega)$.

Now any point t can be approximated by a sequence of diadic point of order n, which we call $t_n=k_n(t)2^{-n}$, such that $|t_n-t_{-1}|=2^{-n}$ while $t_n\to t$. Similarly we call s_n a sequence of dyadic points converging to s. Since X is continuous, it follows that $X_{t_n}\to X_t$ and $X_{s_n}\to X_s$.

Given s and t, let n_0 be the smallest value such that $t_{n_0}=s_{n_0}$. Then by telescopic expansion,

$$X_t-X_s=\sum_{n=n_0+1}^{\infty}\left(X_{t_n}-X_{t_{n-1}}-X_{s_n}+X_{s_{n-1}}\right). \tag{7.2.10}$$

Thus

$$|X_t-X_s|\leq\sum_{n=n_0+1}^{\infty}\left(|X_{t_n}-X_{t_{n-1}}|+|X_{s_n}-X_{s_{n-1}}|\right). \tag{7.2.11}$$

For each t,s, such that $|t-s|\leq 2^{-n^*(\omega)}$, we have $n_0\geq n^*(\omega)$, and thus, on a set of probability one, for such, t,s,

$$|X_t-X_s|\leq 2\sum_{j=n+1}^{\infty}2^{-j\gamma}=\frac{2^{-n\gamma}}{1-2^{\gamma}}\leq\frac{2}{1-2^{-\gamma}}|t-s|^{\gamma}. \tag{7.2.12}$$

Let P denote the law of X. Therefore, we have shown that for all $P\in H$, there exists $\Omega_P\subset\Omega$ s such that $P(\Omega_P)=1$ and for all $\omega\in\Omega_P$, there exist $\delta(\omega)$, such that for all

$s,t \in [0,N]$ with $|s-t| \leq \delta(\omega)$, $|X_t(\omega) - X_s(\omega)| \leq C_N |s-t|^\gamma$, which will be smaller then ε for $|s-t|$ small enough. Now define $\Omega_{P,\delta} = \{\omega \in \Omega_P : \delta(\omega) \leq \delta\}$. Then, for $C_N \delta^\gamma < \varepsilon$,

$$P\left(\sup_{s,t\in[0,N]:|s-t|<\delta} |X_t - X_s| > \varepsilon\right) \leq P_X\left(\Omega_{P,\delta}\right). \tag{7.2.13}$$

But $\Omega_{P,\delta} \downarrow \emptyset$, as $\delta \downarrow 0$, uniformly in $P \in H$. But this implies that

$$\lim_{\delta\downarrow 0} \sup_{P\in H} P\left(\sup_{s,t\in[0,N]:|s-t|<\delta} |X_t - X_s| > \varepsilon\right) = 0, \tag{7.2.14}$$

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

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

Lemma 7.22. *Let μ_n, μ be probability measures in W. Then μ_n converges weakly to μ, if and only if*

(i) the finite dimensional distributions of μ_n converge to those of μ;
(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 μ. Assume that there is another subsequence, $m(k)$, such that $\mu_{m(k)}$ converges weakly to a measure ν. But then also the finite dimensional distibutions of $\mu_{n(k)}$, respectively, $\mu_{m(k)}$, converge to those of μ, repectively ν. But by (i), the finite dimensional marginals of μ_n converge, so that μ and ν 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 [1]) without using the Skorokhod embedding that was used in in the last section of [1]. Note that we already have: (i) convergence of the finite dimensional distributions (Exercise in [1]) 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 [1].

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 7.23. *Let S be a Lousin space and assume the μ_n, μ are probability measures on S. Assume that $\mu_n \to \mu$ weakly. Then there exists a probability space*

$(\Omega, \mathscr{F}, \mathbb{P})$ *and random variables* X_n *with law* μ_n*, and* X *with law* μ*, 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

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

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.

7.3 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.

7.3.1 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 [17]. Here we consider only one case.

Definition 7.24. Let Λ denote the set of all strictly increasing maps $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, such that λ is Lipshitz continuous and

$$\gamma(\lambda) \equiv \sup_{0\leq t<s} \left| \ln \frac{\lambda(s)-\lambda(t)}{s-t} \right| < \infty. \tag{7.3.1}$$

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

$$d(x,y,\lambda,u) \equiv \sup_{\geq 0} \rho\left(x(t\wedge u), y(\lambda(t)\wedge u)\right). \tag{7.3.2}$$

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

$$d(x,y) \equiv \inf_{\lambda\in\Lambda} \left(\gamma(\lambda) \vee \int_0^\infty e^{-u} d(x,y,u,\lambda) du \right). \tag{7.3.3}$$

To get the idea behind this definition, note that with λ the identity, this is just the metric on the space of continuous functions. The rôle of the λ 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 7.25. *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 λ_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

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

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

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

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

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

The next theorem completes our task.

Theorem 7.26. *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 7.8) 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 $n,m \geq n_k$, $d(x_n,x_m) \leq C^{-k}$. Then we can select sequences u_k, and λ_k, such that

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

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 closure of the space of step functions that are uniformly bounded and where the distance between steps is uniformly bounded from below is compact:

Lemma 7.27. *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

$$
w(x,\delta,T)\equiv\inf_{t_i}\max_i\sup_{s,t\in[t_{i-1},t_i)}\rho(x(s),x(t)), \tag{7.3.4}
$$

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 7.28. *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 $T < \infty$,

$$\lim_{\delta \downarrow 0} \sup_{x \in A} w(x, \delta, T) = 0. \tag{7.3.5}$$

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

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

Theorem 7.29. *Let* E *be complete and separable, and let* X_α *be a family of processes with càdlàg paths. Then the family of probability laws,* μ_α*, of* X_α*, 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*

$$\inf_\alpha \mu_\alpha \left(x(t) \in \Gamma_{\eta,t} \right) \geq 1 - \eta, \tag{7.3.6}$$

and

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

$$\sup_\alpha \mu_\alpha \left(w(x, \delta, T) \geq \eta \right) \leq \eta. \tag{7.3.7}$$

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.

References

1. Anton Bovier. Stochastic processes. Lecture notes, Bonn University, 2009.
2. Anton Bovier. Extremes, sums, Lévy processes, and ageing. Lecture notes, Bonn University, 2011.
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. L. R. G. Fontes, M. Isopi, and C. M. Newman. Random walks with strongly inhomogeneous rates and singular diffusions: convergence, localization and aging in one dimension. *Ann. Probab.*, 30(2):579–604, 2002.
8. 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, 1984. Translated from the Russian by Joseph Szücs.
9. Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda. *Dirichlet forms and symmetric Markov processes*, volume 19 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, extended edition, 2011.
10. David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1983.
11. Jean Jacod. *Calcul stochastique et problèmes de martingales*, volume 714 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.
12. Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Graduate Texts in Mathematics. Springer, New York, 1988.
13. M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York, 1983.
14. L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1987. Itô calculus.
15. 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.

16. Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.
17. W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002.

Index