Markov Processes

Andreas Eberle

July 7, 2020

Contents

Introduction

Let $I = \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ (discrete time) or $I = \mathbb{R}_+ = [0, \infty)$ (continuous time), and let $(\Omega, \mathfrak{A}, P)$ be a probability space. If $(S, \mathcal{B})$ is a measurable space then a **stochastic process with state space S** is a collection $(X_t)_{t \in I}$ of random variables

$$X_t : \Omega \to S.$$

More generally, we will consider processes with finite life-time. Here we add an extra point Δ to the state space and we endow $S_\Delta = S\dot{\cup}\{\Delta\}$ with the σ-algebra $\mathcal{B}_\Delta = \{B, B \cup \{\Delta\} : B \in \mathcal{B}\}$. A **stochastic process with state space S and life time** ζ is then defined as a process

$$X_t : \Omega \to S_\Delta \quad \text{such that} \quad X_t(\omega) = \Delta \quad \text{if and only if} \quad t \geq \zeta(\omega).$$

Here $\zeta : \Omega \to [0, \infty]$ is a random variable.

We will usually assume that the state space S is a **Polish space**, i.e., there exists a metric $d : S \times S \to \mathbb{R}_+$ such that (S, d) is complete and separable. Note that for example open sets in $\mathbb{R}^n$ are Polish spaces, although they are not complete w.r.t. the Euclidean metric. Indeed, most state spaces encountered in applications are Polish. Moreover, on Polish spaces regular versions of conditional probability distributions exist. This will be crucial for much of the theory developed below. If S is Polish then we will always endow it with its Borel σ-algebra $\mathcal{B} = \mathcal{B}(S)$.

A **filtration** on $(\Omega, \mathfrak{A}, P)$ is an increasing collection $(\mathcal{F}_t)_{t \in I}$ of σ-algebras $\mathcal{F}_t \subset \mathfrak{A}$. A stochastic process $(X_t)_{t \in I}$ is **adapted** w.r.t. a filtration $(\mathcal{F}_t)_{t \in I}$ iff X_t is $\mathcal{F}_t$-measurable for any $t \in I$. In particular, any process $X = (X_t)_{t \in I}$ is adapted to the filtrations $(\mathcal{F}_t^X)$ and $(\mathcal{F}_t^{X,P})$ where

$$\mathcal{F}_t^X \;=\; \sigma(X_s : s \in I, s \leq t), \quad t \in I,$$

is the **filtration generated by X**, and $\mathcal{F}_t^{X,P}$ denotes the **completion** of the σ-algebra $\mathcal{F}_t$ w.r.t. the probability measure P:

$$\mathcal{F}_t^{X,P} = \{A \in \mathfrak{A} : \exists\, \widetilde{A} \in \mathcal{F}_t^X \text{ with } P[\widetilde{A} \Delta A] = 0\}.$$

Finally, a stochastic process $(X_t)_{t \in I}$ on $(\Omega, \mathfrak{A}, P)$ with state space $(S, \mathcal{B})$ is called an $(\mathcal{F}_t)$ **Markov process** iff (X_t) is adapted w.r.t. the filtration $(\mathcal{F}_t)_{t \in I}$, and

$$P[X_t \in B | \mathcal{F}_s] = P[X_t \in B | X_s] \quad P\text{-a.s. for any } B \in \mathcal{B} \text{ and } s, t \in I \text{ with } s \leq t. \tag{0.1}$$

Any $(\mathcal{F}_t)$ Markov process is also a Markov process w.r.t. the filtration $(\mathcal{F}_t^X)$ generated by the process. Hence an $(\mathcal{F}_t^X)$ Markov process will be called simply a **Markov process**. We will see other equivalent forms of the Markov property below. For the moment we just note that (0.1) implies

$$P[X_t \in B | \mathcal{F}_s] = p_{s,t}(X_s, B) \quad P\text{-a.s. for any } B \in \mathcal{B} \text{ and } s \leq t, \text{ and} \tag{0.2}$$

$$E[f(X_t) | \mathcal{F}_s] = (p_{s,t} f)(X_s) \quad P\text{-a.s. for any measurable function } f : S \to \mathbb{R}_+ \text{ and } s \leq t, \tag{0.3}$$

where $p_{s,t}(x, dy)$ is a regular version of the conditional probability distribution of X_t given X_s, and

$$(p_{s,t} f)(x) = \int_S p_{s,t}(x, dy) f(y).$$

Furthermore, by the tower property of conditional expectations, the kernels $p_{s,t}$ $(s,t \in I$ with $s \le t)$ satisfy the consistency condition

$$p_{s,u}(X_s, B) = \int_S p_{s,t}(X_s, dy)\, p_{t,u}(y, B) \tag{0.4}$$

P-almost surely for any $B \in \mathcal{B}$ and $s \le t \le u$, i.e., for any measurable function $f : S \to \mathbb{R}_+$,

$$p_{s,u} f = p_{s,t} p_{t,u} f \quad P \circ X_s^{-1}\text{-almost surely for any } 0 \le s \le t \le u. \tag{0.5}$$

Exercise. Show that the consistency conditions (0.4) and (0.5) follow from the defining property (0.2) of the kernels $p_{s,t}$.

1. Transition functions and Markov processes

From now on we assume that S is a Polish space and $\mathcal{B}$ is the Borel σ-algebra on S. We denote the collection of all non-negative respectively bounded measurable functions $f : S \to \mathbb{R}$ by $\mathcal{F}_+(S), \mathcal{F}_b(S)$ respectively. The space of all probability measures resp. finite signed measures are denoted by $\mathcal{P}(S)$ and $\mathcal{M}(S)$. For $\mu \in \mathcal{M}(S)$ and $f \in \mathcal{F}_b(S)$, as well as for $\mu \in \mathcal{P}(S)$ and $f \in \mathcal{F}_+(S)$, we set

$$\mu(f) = \int f\, d\mu.$$

The following definition is natural by the considerations above:

Definition 0.1 (Sub-probability kernel, transition function). 1) A **(sub) probability kernel** p on $(S, \mathcal{B})$ is a map $(x, B) \mapsto p(x, B)$ from $S \times \mathcal{B}$ to $[0, 1]$ such that

(i) for any $x \in S$, $p(x, \cdot)$ is a positive measure on $(S, \mathcal{B})$ with total mass $p(x, S) = 1$ $(p(x, S) \le 1$ respectively), and

(ii) for any $B \in \mathcal{B}$, $p(\cdot, B)$ is a measurable function on $(S, \mathcal{B})$.

2) A **transition function** is a collection $p_{s,t}$ $(s,t \in I$ with $s \le t)$ of sub-probability kernels on $(S, \mathcal{B})$ satisfying

$$p_{t,t}(x, \cdot) = \delta_x \quad \text{for any } x \in S \text{ and } t \in I, \text{ and} \tag{0.6}$$

$$p_{s,t} p_{t,u} = p_{s,u} \quad \text{for any } s \le t \le u, \tag{0.7}$$

where the composition of two sub-probability kernels p and q on $(S, \mathcal{B})$ is the sub-probability kernel pq defined by

$$(pq)(x, B) = \int p(x, dy)\, q(y, B) \quad \text{for any } x \in S, B \in \mathcal{B}.$$

The equations in (0.7) are called the **Chapman-Kolmogorov equations**. They correspond to the consistency conditions in (0.4). Note, however, that we are now assuming that the consistency conditions hold everywhere. This will allow us to relate a family of Markov processes with arbitrary starting points and starting times to a transition function.

The reason for considering sub-probability instead of probability kernels is that mass may be lost during the evolution if the process has a finite life-time. We extend transition functions of sub-probability kernels on $(S, \mathcal{B})$ to transition functions of probability kernels on $(S_\Delta, \mathcal{B}_\Delta)$ by setting $p_{s,t}(\Delta, \cdot) := \delta_\Delta$, and

$$p_{s,t}(x, \{\Delta\}) := 1 - p_{s,t}(x, S) \qquad \text{for any } x \in S \text{ and } s \le t. \tag{0.8}$$

Example (Discrete and absolutely continuous transition kernels). A sub-probability kernel on a countable set S takes the form $p(x,\{y\}) = p(x,y)$ where $p : S\times S \to [0,1]$ is a non-negative function satisfying $\sum_{y\in S} p(x,y) \le 1$. More generally, let λ be a non-negative measure on a general Polish state space (e.g. the counting measure on a discrete space or Lebesgue measure on $\mathbb{R}^n$). If $p : S \times S \to \mathbb{R}_+$ is a measurable function satisfying

$$\int p(x,y)\lambda(dy) \le 1 \qquad \text{for any } x \in S,$$

then p is the density of a sub-probability kernel given by

$$p(x,B) = \int_B p(x,y)\lambda(dy).$$

The collection of corresponding densities $p_{s,t}(x,y)$ for the kernels of a transition function w.r.t. a fixed measure λ is called a **transition density**. Note, however, that many interesting Markov processes on general state spaces do not possess a transition density w.r.t. a natural reference measure. A simple example is the Random Walk Metropolis algorithm on $\mathbb{R}^d$. This Markov chain moves in each time step with a positive probability according to an absolutely continuous transition density, whereas with the opposite probability, it stays at its current position, see Section 3.2 below.

Definition 0.2 (Markov process with transition function $\mathbf{p_{s,t}}$). Let $p_{s,t}$ ($s,t \in I$ with $s \le t$) be a transition function on $(S,\mathcal{B})$, and let $(\mathcal{F}_t)_{t\in I}$ be a filtration on a probability space $(\Omega,\mathfrak{A},P)$.

1) A stochastic process $(X_t)_{t\in I}$ on $(\Omega,\mathfrak{A},P)$ with values in S_Δ is called an $(\mathcal{F}_t)$ **Markov process with transition function** $(\mathbf{p_{s,t}})$ iff it is $(\mathcal{F}_t)$ adapted, and

   (MP) $P[X_t \in B|\mathcal{F}_s] = p_{s,t}(X_s,B)$ P-a.s. for any $s \le t$ and $B \in \mathcal{B}$.

2) It is called **time-homogeneous** iff the transition function is time-homogeneous, i.e., iff there exist sub-probability kernels p_t ($t \in I$) such that

$$p_{s,t} = p_{t-s} \qquad \text{for any } s \le t.$$

Notice that time-homogeneity does not mean that the law of X_t is independent of t; it is only a property of the transition function. For the transition kernels $(p_t)_{t\in I}$ of a time-homogeneous Markov process, the Chapman-Kolmogorov equations take the simple form

$$p_{s+t} \;=\; p_s p_t \qquad \text{for any } s,t \in I. \tag{0.9}$$

A time-inhomogeneous Markov process (X_t) with state space S can be identified with the time-homogeneous Markov process (t,X_t) on the enlarged state space $\mathbb{R}_+ \times S$:

Exercise (Reduction to time-homogeneous case). Let $((X_t)_{t\in I},P)$ be a Markov process with transition function $(p_{s,t})$. Show that for any $t_0 \in I$ the time-space process $\hat{X}_t = (t_0+t, X_{t_0+t})$ is a time-homogeneous Markov process with state space $\mathbb{R}_+ \times S$ and transition function

$$\hat{p}_t\,((s,x),\cdot) = \delta_{s+t} \otimes p_{s,s+t}(x,\cdot).$$

Kolmogorov's Theorem states that for any transition function and any given initial distribution there is a unique canonical Markov process on the product space

$$\Omega_{\text{can}} = S_\Delta^I = \{\omega : I \to S_\Delta\}.$$

Indeed, let $X_t : \Omega_{\text{can}} \to S_\Delta, X_t(\omega) = \omega(t)$, denote the evaluation at time t, and endow Ω_{can} with the product σ-algebra

$$\mathfrak{A}_{\text{can}} = \bigotimes_{t \in I} \mathcal{B}_\Delta = \sigma(X_t : t \in I).$$

Theorem 0.3 (Kolmogorov's Theorem). Let $p_{s,t}$ $(s, t \in I$ with $s \leq t)$ be a transition function on $(S, \mathcal{B})$. Then for any probability measure ν on $(S, \mathcal{B})$, there exists a unique probability measure P_ν on $(\Omega_{\text{can}}, \mathfrak{A}_{\text{can}})$ such that $((X_t)_{t \in I}, P_\nu)$ is a Markov process with transition function $(p_{s,t})$ and initial distribution $P_\nu \circ X_0^{-1} = \nu$.

Since the Markov property (MP) is equivalent to the fact that the finite-dimensional marginal laws of the process are given by

$$\left(X_{t_0}, X_{t_1}, \ldots, X_{t_n}\right) \sim \mu(dx_0) p_{0,t_1}(x_0, dx_1) p_{t_1,t_2}(x_1, dx_2) \cdots p_{t_{n-1},t_n}(x_{n-1}, dx_n)$$

for any $0 = t_0 \leq t_1 \leq \cdots \leq t_n$, the proof of Theorem 0.3 is a consequence of Kolmogorov's extension theorem (which follows from Carathéodory's extension theorem). Thus Theorem 0.3 is a purely measure-theoretic statement. Its main disadvantage is that the space S^I is too large and the product σ-algebra is too small when $I = \mathbb{R}_+$. Indeed, in this case important events such as the event that the process $(X_t)_{t \geq 0}$ has continuous trajectories are not measurable w.r.t. $\mathfrak{A}_{\text{can}}$. Therefore, in continuous time we will usually replace Ω_{can} by the space $\mathcal{D}(\mathbb{R}_+, S_\Delta)$ of all right-continuous functions $\omega : \mathbb{R}_+ \to S_\Delta$ with left limits $\omega(t-)$ for any $t > 0$. To realize a Markov process with a given transition function on $\Omega = \mathcal{D}(\mathbb{R}_+, S_\Delta)$ requires modest additional regularity conditions, cf. e.g. Rogers & Williams I [49].

For $x \in S$ and $t_0 \in I$, we denote by $P_x^{(t_0)}$ the canonical measure corresponding to the initial distribution δ_x and the time-shifted transition kernels

$$p_{s,t}^{(t_0)} = p_{t_0+s, t_0+t}.$$

In particular, in the time-homogeneous case, $p_{s,t}^{(t_0)} = p_{s,t}$ for any t_0, whence $P_x^{(t_0)}$ coincides with the law P_{δ_x} of the original Markov process starting at x.

Theorem 0.4 (Markov properties). Let $(X_t)_{t \in I}$ be a stochastic process with state space $(S_\Delta, \mathcal{B}_\Delta)$ defined on a probability space $(\Omega, \mathfrak{A}, P)$. Then the following statements are equivalent:

(i) (X_t, P) is a Markov process with initial distribution ν and transition function $(p_{s,t})$.

(ii) For any $n \in \mathbb{Z}_+$ and $0 = t_0 \leq t_1 \leq \ldots \leq t_n$,

$$(X_{t_0}, X_{t_1}, \ldots, X_{t_n}) \sim \nu \otimes p_{t_0,t_1} \otimes p_{t_1,t_2} \otimes \cdots \otimes p_{t_{n-1},t_n} \qquad \text{w.r.t. } P.$$

(iii) $(X_t)_{t \in I} \sim P_\nu$.

(iv) For any $s \in I$, $P_{X_s}^{(s)}$ is a version of the conditional distribution of $(X_t)_{t \geq s}$ given $\mathcal{F}_s^X$, i.e.,

$$E[F((X_t)_{t \geq s}) | \mathcal{F}_s^X] = E_{X_s}^{(s)}[F] \quad P\text{-a.s.}$$

for any $\mathfrak{A}_{\text{can}}$-measurable function $F : \Omega_{\text{can}} \to \mathbb{R}_+$.

2. Some classes of Markov processes

Markov chains

Markov processes $(X_t)_{t\in\mathbb{Z}_+}$ in discrete time are called **Markov chains**. The transition function of a Markov chain is completely determined by its one-step transition kernels $\pi_n = p_{n-1,n}$ $(n \in \mathbb{N})$. Indeed, by the Chapman-Kolmogorov equation,

$$p_{s,t} = \pi_{s+1}\pi_{s+2}\cdots\pi_t \qquad \text{for any } s,t \in \mathbb{Z}_+ \text{ with } s \leq t.$$

In particular, in the time-homogeneous case, the transition function takes the form

$$p_t = \pi^t \qquad \text{for any } t \in \mathbb{Z}_+,$$

where $\pi = p_{n-1,n}$ is the one-step transition kernel that does not depend on n.

The canonical law P_μ of a Markov chain with initial distribution μ on

$$\Omega_{\text{can}} = S^{\{0,1,2,\dots\}} = \{(\omega_n)_{n\in\mathbb{Z}_+} : \omega_n \in S\}$$

is the infinite product $P_\mu = \mu \otimes \pi_1 \otimes \pi_2 \otimes \pi_3 \otimes \dots$. In particular,

$$(X_0, \dots, X_n) \sim \mu \otimes \pi_1 \otimes \pi_2 \otimes \dots \otimes \pi_n \qquad \text{for any } n \geq 0.$$

For Markov chains, the Markov property (iv) in Theorem 0.4 says that for any $n \in \mathbb{Z}_+$, $P^{(n)}_{X_n}$ is a version of the conditional distribution of $(X_n, X_{n+1}, \dots)$ given $(X_0, X_1, \dots, X_n)$, i.e.,

$$E[F(X_n, X_{n+1}, \dots)|X_0, \dots, X_n] \;=\; E^{(n)}_{X_n}[F] \quad P\text{-a.s.}$$

for any $\mathfrak{A}_{\text{can}}$-measurable function $F : \Omega_{\text{can}} \to \mathbb{R}_+$.

One way in which Markov chains frequently arise in applications is as **random dynamical systems**: A stochastic process on a probability space $(\Omega, \mathfrak{A}, P)$ defined recursively by

$$X_{n+1} = \Phi_{n+1}(X_n, W_{n+1}) \qquad \text{for } n \in \mathbb{Z}_+ \tag{0.10}$$

is a Markov chain if $X_0 : \Omega \to S$ and $W_1, W_2, \dots : \Omega \to T$ are independent random variables taking values in measurable spaces $(S, \mathcal{B})$ and $(T, \mathcal{C})$, and $\Phi_1, \Phi_2, \dots$ are measurable functions from $S \times T$ to S. The one-step transition kernels are

$$\pi_n(x, B) = P[\Phi_n(x, W_n) \in B],$$

and the transition function is given by

$$p_{s,t}(x, B) = P[X_t(s, x) \in B],$$

where $X_t(s, x)$ for $t \geq s$ denotes the solution of the recurrence relation (0.10) with initial value $X_s(s, x) = x$ at time s. The Markov chain is time-homogeneous if the random variables W_n are identically distributed, and the functions Φ_n coincide for all $n \in \mathbb{N}$.

On a Polish state space S, every Markov chain can be represented as a random dynamical system in the form

$$X_{n+1} = \Phi_{n+1}(X_n, W_{n+1})$$

with independent random variables $X_0, W_1, W_2, W_3, \dots$ and measurable functions $\Phi_1, \Phi_2, \Phi_3, \dots$, see e.g. Kallenberg [27]. Often such representations arise naturally:

Example. 1) **Random Walk on $\mathbb{R}^{\mathbf{d}}$.** A d-dimensional Random Walk is defined by a recurrence relation $X_{n+1} = X_n + W_{n+1}$ with i.i.d. random variables $W_1, W_2, W_3, \ldots : \Omega \to \mathbb{R}^d$ and a independent initial value $X_0 : \Omega \to \mathbb{R}^d$.

2) **Reflected Random Walk on $\mathbf{S} \subset \mathbb{R}^{\mathbf{d}}$.** There are several possibilities for defining a reflected random walk on a measurable subset $S \subset \mathbb{R}^d$. The easiest is to set

$$X_{n+1} = X_n + W_{n+1} 1_{\{X_n + W_{n+1} \in S\}}$$

with i.i.d. random variables $W_i : \Omega \to \mathbb{R}^d$. One application where reflected random walks are of interest is the simulation of **hard-core models**. Suppose there are d particles of diameter r in a box $B \subset \mathbb{R}^3$. The configuration space of the system is given by

$$S = \left\{(x_1, \ldots, x_d) \in \mathbb{R}^{3d} : x_i \in B \text{ and } |x_i - x_j| > r \ \forall i \neq j\right\}.$$

If the law of the increments is invariant under reflection, i.e., $-W_n \sim W_n$, then the uniform distribution on S is a stationary distribution of the reflected random walk on S defined above.

3) **State Space Models with additive noise.** Several important models of Markov chains in $\mathbb{R}^d$ are defined by recurrence relations of the form

$$X_{n+1} = \Phi(X_n) + W_{n+1}$$

with i.i.d. random variables W_i ($i \in \mathbb{N}$). Besides random walks these include e.g. **linear state space models** where

$$X_{n+1} = AX_n + W_{n+1} \qquad \text{for some matrix } A \in \mathbb{R}^{d \times d},$$

and stochastic volatility models defined e.g. by

$$\begin{aligned} X_{n+1} &= X_n + e^{V_n/2} W_{n+1}, \\ V_{n+1} &= m + \alpha(V_n - m) + \sigma Z_{n+1} \end{aligned}$$

with constants $\alpha, \sigma \in \mathbb{R}_+, m \in \mathbb{R}$, and i.i.d. random variables W_i and Z_i. In the latter class of models X_n stands for the logarithmic price of an asset and V_n for the logarithmic volatility.

Markov chains in continuous time

If $(Y_n)_{n \in \mathbb{Z}_+}$ is a time-homogeneous Markov chain on a probability space $(\Omega, \mathfrak{A}, P)$, and $(N_t)_{t \geq 0}$ is a Poisson process with intensity $\lambda > 0$ on $(\Omega, \mathfrak{A}, P)$ that is independent of $(Y_n)_{n \in \mathbb{Z}_+}$ then the process

$$X_t = Y_{N_t}, \quad t \in [0, \infty),$$

is a time-homogeneous Markov process in continuous time, see e.g. [18]. Conditioning on the value of N_t shows that the transition function is given by

$$p_t(x, B) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \pi^k(x, B) = e^{\lambda t(\pi - I)}(x, B).$$

The construction can be generalized to time-inhomogeneous jump processes with finite jump intensities, but in this case the processes (Y_n) and (N_t) determining the positions and the jump times are not necessarily Markov processes on their own, and they are not necessarily independent of each other, see Section 5.1 below.

Diffusion processes

A **Brownian motion** $((B_t)_{t\geq 0}, P)$ taking values in $\mathbb{R}^n$ is a time-homogeneous Markov process with continuous sample paths $t \mapsto B_t(\omega)$ and transition density

$$p_t(x,y) = (2\pi t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2t}\right)$$

with respect to the n-dimensional Lebesgue measure λ^n. In general, Markov processes with continuous sample paths are called **diffusion processes**. It can be shown that a solution to an Itô stochastic differential equation of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \; X_0 = x_0, \tag{0.11}$$

is a diffusion process if, for example, the coefficients are Lipschitz continuous functions $b : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$, and $(B_t)_{t\geq 0}$ is a Brownian motion in $\mathbb{R}^d$. In this case, the transition function is usually not known explicitly.

Example (Ornstein-Uhlenbeck process). The Ornstein-Uhlenbeck process is the solution to the SDE (0.11) with $b(t,x) = -\gamma x$ and $\sigma(t,x) = I_d$ where $\gamma > 0$ is a fixed parameter. The solution to this linear SDE is a Gaussian process that can be computed explicitly by variation of constants. It is given by

$$X_t \;=\; e^{-\gamma t} X_0 \;+\; \int_0^t e^{\gamma(t-s)}\, dB_s$$

where the integral is an Itô integral. Observe that the influence of the initial condition decays exponentially.

Markov processes with finite life-time

Given an arbitrary Markov process and a possibly time and state dependent rate function, we can define another Markov process that follows the same dynamics until it eventually dies with the given rate.

We first consider the discrete time case. Let $(X_n)_{n\in\mathbb{Z}_+}$ be the original Markov chain with state space S and transition probabilities π_n, and suppose that the death rates are given by measurable functions $w_n : S \times S \to [0,\infty]$, i.e., the survival probability is $e^{-w_n(x,y)}$ if the Markov chain is jumping from x to y in the n-th step. Let E_n $(n \in \mathbb{N})$ be independent exponential random variables with parameter 1 that are also independent of the Markov chain (X_n). Then we can define a Markov chain with state space $S\dot{\cup}\{\Delta\}$ recursively by $X_0^w = X_0$,

$$X_n^w \;=\; \begin{cases} X_n & \text{if } X_{n-1}^w \neq \Delta \text{ and } E_n > w_n(X_{n-1}, X_n), \\ \Delta & \text{otherwise.} \end{cases}$$

Example (Absorption and killing on the boundary). If D is a measurable subset of S, and we set

$$w_n(x,y) = \begin{cases} 0 & \text{for } y \in D, \\ \infty & \text{for } y \in S\backslash D, \end{cases}$$

then the Markov chain (X_n^w) is *killed* when exiting the domain D for the first time. Note that (X_n^w) differs slightly from the Markov chain with *absorption at the boundary* that is defined as

$$X_n^D \;=\; X_{n\wedge T},$$

where $T = \min\{n \geq 0 : X_n \notin D\}$ is the first exit time from D.

Lemma 0.5. *The process (X_n^w) is a Markov chain on S_Δ w.r.t. the filtration $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n, E_1, \ldots, E_n)$. The transition probabilities are given by*

$$\begin{aligned} \pi_n^w(x,dy) \;&=\; e^{-w_n(x,y)}\pi_n(x,dy) + \left(1 - \int e^{-w_n(x,z)}\pi_n(x,dz)\right)\delta_\Delta(dy) \quad \textit{for } x \in S, \\ \pi_n^w(\Delta,\cdot) \;&=\; \delta_\Delta\,. \end{aligned}$$

Proof. For any Borel subset B of S,

$$
\begin{aligned}
P\left[X_{n+1}^w \in B|\mathcal{F}_n\right] &= E\left[P\left[X_n^w \in S, X_{n+1} \in B, E_{n+1} > w_n(X_n, X_{n+1})|\sigma(X_{0:\infty}, E_{1:n})\right]|\mathcal{F}_n\right] \\
&= E\left[1_S(X_n^w)1_B(X_{n+1})e^{-w_{n+1}(X_n,X_{n+1})}|\mathcal{F}_n\right] \\
&= 1_S(X_n^w)\int_B e^{-w_{n+1}(X_n,y)}\pi_{n+1}(X_n, dy) \;=\; \pi_{n+1}(X_n^w, B).
\end{aligned}
$$

Here we have used the properties of conditional expectations and the Markov property for (X_n). The assertion follows since the σ-algebra on $S \cup \{\Delta\}$ is generated by the sets in $\mathcal{B}$, and $\mathcal{B}$ is stable under intersections. ■

In continuous time, we can not use a step-by-step construction of a Markov process with a given death rate. However, the memoryless property of the exponential distribution ensures that the following alternative construction leads to an equivalent result. Let $(X_t)_{t\in\mathbb{R}_+}$ be a time-homogeneous Markov process on $(\Omega, \mathfrak{A}, P)$ with state space $(S, \mathcal{B})$, and suppose that the death rate is given by a measurable function $V : \mathbb{R}_+ \times S \to [0, \infty]$. Then the accumulated death rate up to time t is given by setting

$$
A_t \;=\; \int_0^t V(s, X_s)\, ds.
$$

We define the corresponding process $(X_t^V)_{t\in\mathbb{R}_+}$ with death rate V by

$$
X_t^V \;=\; \begin{cases} X_t & \text{if } A_t < E, \\ \Delta & \text{if } A_t \geq E, \end{cases} \qquad \zeta \;=\; \inf\{t \geq 0 : X_t^V = \Delta\} = \inf\{t \geq 0 : A_t \geq E\},
$$

where E is an exponential random variable with parameter 1 that is independent of the process (X_t).

Lemma 0.6. *The process (X_t^V) is a Markov process on S_Δ w.r.t. the filtration $\mathcal{F}_t = \sigma(D, \{s < \zeta\} : D \in \mathcal{F}_t^X, s \in [0, t])$. The transition probabilities are given by*

$$
\begin{aligned}
&p_{s,t}^V(x, B) = E_x\left[\exp\left(-\int_0^{t-s} V(s+u, X_u)\, du\right); X_{t-s} \in B\right] \qquad \text{for } x \in S \text{ and } B \in \mathcal{B}, \\
&p_{s,t}^V(\Delta, \cdot) = \delta_\Delta.
\end{aligned}
$$

Proof. Fix $B \in \mathcal{B}$ and $0 \leq s \leq t$. We are going to show that

$$
P[X_t^V|\mathcal{F}_s] \;=\; p_{s,t}^V(X_s, B) \qquad P\text{-a.s} \tag{0.12}
$$

To this end we observe that the σ-algebra $\mathcal{F}_s$ is generated by the collection of events of the form $D \cap \{r < \zeta\}$ with $D \in \mathcal{F}_s$ and $r \in [0, s]$. Furthermore, for events of this form, a computation based on the Markov property for the original process (X_t) and on the properties of conditional expectations yields

$$
\begin{aligned}
&P\left[\{X_t^V \in B\} \cap D \cap \{r < \zeta\}\right] \;=\; P\left[\{X_t \in B\} \cap D \cap \{t < \zeta\}\right] \\
&\quad= \;E\left[P\left[t < \zeta|\mathcal{F}_\infty^X\right]; \{X_t \in B\} \cap D\right] \\
&\quad= \;E\left[\exp(-A_t); \{X_t \in B\} \cap D\right] \\
&\quad= \;E\left[\exp(-A_s)1_D\, E\left[\exp\left(-\int_s^t V(s+u, X_u)\, du\right)1_B(X_t)\middle|\mathcal{F}_s^X\right]\right] \\
&\quad= \;E\left[p_{s,t}^V(X_s, B); D \cap \{s < \zeta\}\right] \;=\; E\left[p_{s,t}^V(X_s, B); D \cap \{r < \zeta\}\right].
\end{aligned}
$$

Here we have used in the first and last step that $\{t < \zeta\} \subseteq \{s < \zeta\} \subseteq \{r < \zeta\}$ and $p_{s,t}(X_s, B) = 0$ on $s \geq \zeta$. The assertion follows since the collection of events generating $\mathcal{F}_s$ considered above is stable under intersections. ■

Exercise. Complete the proof by verifying that $(p_{s,t})$ is indeed a transition function.

3. Generators and Martingales

Since the transition function of a Markov process is usually not known explicitly, one is looking for other natural ways to describe the evolution. An obvious idea is to consider the rate of change of the transition probabilities or expectations at a given time t.

In discrete time this is straightforward: For $f \in \mathcal{F}_b(S)$ and $t \geq 0$,

$$E[f(X_{t+1}) - f(X_t)|\mathcal{F}_t] = (\mathcal{L}_t f)(X_t) \quad P\text{-a.s.} \tag{0.13}$$

where $\mathcal{L}_t : \mathcal{F}_b(S) \to \mathcal{F}_b(S)$ is the linear operator defined by

$$(\mathcal{L}_t f)(x) = (\pi_{t+1} f)(x) - f(x) = \int \pi_{t+1}(x, dy)(f(y) - f(x)).$$

$\mathcal{L}_t$ is called the **generator at time t** - in the time homogeneous case it does not depend on t.

Example. 1) **Simple random walk on $\mathbb{Z}$.** Here $\pi(x, \cdot) = \frac{1}{2}\delta_{x+1} + \frac{1}{2}\delta_{x-1}$. Hence the generator is given by the second difference (discrete Laplacian):

$$(\mathcal{L}f)(x) = \frac{1}{2}(f(x+1) + f(x-1)) - f(x) = \frac{1}{2}[(f(x+1) - f(x)) - (f(x) - f(x-1))].$$

2) **Random walk on $\mathbb{R}^{\mathbf{d}}$.** A random walk on $\mathbb{R}^d$ with increment distribution μ can be represented as

$$X_n = x + \sum_{k=1}^{n} W_k \quad (n \in \mathbb{Z}_+)$$

with independent random variables $W_k \sim \mu$. The generator is given by

$$(\mathcal{L}f)(x) = \int f(x+w)\mu(dw) - f(x) = \int (f(x+w) - f(x))\mu(dw).$$

3) **Markov chain with finite life-time.** Suppose that $\mathcal{L}$ is the generator of a time-homogeneous Markov chain with state space S and transition kernel π. Then the generator of the corresponding Markov chain on $S\dot{\cup}\{\Delta\}$ with death rate $w_n(x, y)$ is given by

$$(\mathcal{L}_n^w f)(x) \;=\; \int \left(e^{-w_{n+1}(x,y)} f(y) - f(x)\right) \pi(x, dy)\Big) \qquad \text{for } x \in S.$$

In particular, if $w_{n+1}(x, y) = v_n(x)$ for measurable functions $v_0, v_1, \ldots$ (i.e., the death rate in the next step only depends on the current position), then

$$\mathcal{L}_n^w f \;=\; e^{-v_n}\mathcal{L}f \;+\; (e^{-v_n} - 1)\, f.$$

for any bounded measurable function $f : S \to \mathbb{R}$, and for any $x \in S$.

In continuous time, the situation is more involved. Here we have to consider the instantaneous rate of change, i.e., the derivative of the transition function. We would like to define

$$(\mathcal{L}_t f)(x) = \lim_{h \downarrow 0} \frac{(p_{t,t+h} f)(x) - f(x)}{h} = \lim_{h \downarrow 0} \frac{1}{h} E[f(X_{t+h}) - f(X_t)|X_t = x]. \tag{0.14}$$

By an informal calculation based on the Chapman-Kolmogorov equation, we could then hope that the transition function satisfies the differential equations

$$\text{(FE)} \qquad \frac{d}{dt} p_{s,t} f = \frac{d}{dh}\left(p_{s,t} p_{t,t+h} f\right)|_{h=0} = p_{s,t}\mathcal{L}_t f, \quad \text{and} \tag{0.15}$$

$$\text{(BE)} \qquad -\frac{d}{ds} p_{s,t} f = -\frac{d}{dh}\left(p_{s,s+h} p_{s+h,t} f\right)|_{h=0} + p_{s,s}\mathcal{L}_s p_{s,t} f = \mathcal{L}_s p_{s,t} f. \tag{0.16}$$

These equations are called **Kolmogorov's forward and backward equation** respectively, since they describe the forward and backward in time evolution of the transition probabilities.

However, making these informal computations rigorous is not a triviality in general. The problem is that the right-sided derivative in (0.14) may not exist for all bounded functions f. Moreover, different notions of convergence on function spaces lead to different definitions of $\mathcal{L}_t$ (or at least of its domain). Indeed, we will see that in many cases, the generator of a Markov process in continuous time is an unbounded linear operator - for instance, generators of diffusion processes are (generalized) second order differential operators. One way to circumvent these difficulties partially is the martingale problem of Stroock and Varadhan which sets up a connection to the generator only on a fixed class of nice functions.

We first consider again the discrete time case. Here, the generator can be used immediately to identify martingales associated to a Markov chain. Indeed if (X_n, P) is an $(\mathcal{F}_n)$ Markov chain with transition kernels π_n then by (0.13), for any $f \in \mathcal{F}_b(S)$, the process $M^{[f]}$ defined by

$$M_n^{[f]} \;=\; f(X_n) - \sum_{k=0}^{n-1} (\mathcal{L}_k f)(X_k), \quad n \in \mathbb{Z}_+, \tag{0.17}$$

is an $(\mathcal{F}_n)$ martingale. We even have:

Theorem 0.7 (Martingale problem characterization of Markov chains). Let $X_n : \Omega \to S$ be an $(\mathcal{F}_n)$ adapted stochastic process defined on a probability space $(\Omega, \mathfrak{A}, P)$. Then (X_n, P) is an $(\mathcal{F}_n)$ Markov chain with transition kernels π_n if and only if the process $M^{[f]}$ defined by (0.17) is an $(\mathcal{F}_n)$ martingale for every function $f \in \mathcal{F}_b(S)$.

The proof is a direct consequence of the fact that (X_n, P) is an $(\mathcal{F}_n)$ Markov chain with transition kernels π_n if and only if (0.13) holds for any $f \in \mathcal{F}_b(S)$.

The martingale problem provides a Doob decomposition for arbitrary bounded functions of a Markov chain into a martingale and a predictable process:

$$f(X_n) \;=\; M_n^{[f]} \;+\; A_n^{[f]}, \quad \text{where} \quad A_n^{[f]} \;:=\; \sum_{k=0}^{n-1} (\mathcal{L}_k f)(X_k) \quad \text{is } \mathcal{F}_{n-1}\text{-measurable.}$$

This decomposition can also be extended to time-dependent functions. Indeed, if (X_n, P) is a Markov chain with state space S and transition kernels π_n, then the time-space process $\hat{X}_n := (n, X_n)$ is a time-homogeneous Markov chain with state space $\mathbb{Z}_+ \times S$. Let

$$\begin{aligned}(\hat{\mathcal{L}} f)(n,x) &= \int \pi_{n+1}(x, dy)(f(n+1,y) - f(n,x)) \\ &= (\mathcal{L}_n\, f(n+1,\cdot))(x) \;+\; f(n+1,x) - f(n,x)\end{aligned}$$

denote the corresponding **time-space generator**.

Corollary 0.8 (Time-dependent martingale problem for Markov chains). Let $X_n : \Omega \to S$ be an $(\mathcal{F}_n)$ adapted stochastic process defined on a probability space $(\Omega, \mathfrak{A}, P)$. Then (X_n, P) is an $(\mathcal{F}_n)$ Markov chain with transition kernels $\pi_1, \pi_2, \ldots$ if and only if the processes

$$M_n^{[f]} \;:=\; f(n, X_n) - \sum_{k=0}^{n-1} (\hat{\mathcal{L}} f)(k, X_k) \quad (n \in \mathbb{Z}_+)$$

are $(\mathcal{F}_n)$ martingales for all functions $f \in \mathcal{F}_b(\mathbb{Z}_+ \times S)$.

Proof. By definition, the process (X_n, P) is a Markov chain with transition kernels π_n if and only if the time-space process $((n, X_n), P)$ is a time-homogeneous Markov chain with transition kernel $\hat{\pi}((n,x),\cdot) = \delta_{n+1} \otimes \pi_{n+1}(x,\cdot)$. The assertion now follows from Theorem 0.7. ■

We now return to general Markov processes. Let $\mathcal{A}$ be a linear space of bounded measurable functions on $(S, \mathcal{B})$, and let $\mathcal{L}_t : \mathcal{A} \to \mathcal{F}(S)$, $t \in I$, be a collection of linear operators with domain $\mathcal{A}$ taking values in the space $\mathcal{F}(S)$ of measurable (not necessarily bounded) functions on $(S, \mathcal{B})$.

Definition 0.9 (Martingale problem). A stochastic process $((X_t)_{t\in I}, P)$ that is adapted to a filtration $(\mathcal{F}_t)$ is said to be a **solution of the martingale problem for** $((\mathcal{L}_\mathbf{t})_{\mathbf{t}\in\mathbf{I}}, \mathcal{A})$ iff the real valued processes

$$M_t^f = f(X_t) - \sum_{s=0}^{t-1} (\mathcal{L}_s f)(X_s) \quad \text{if } I = \mathbb{Z}_+, \text{ resp.}$$

$$M_t^f = f(X_t) - \int_0^t (\mathcal{L}_s f)(X_s) \quad \text{if } I = \mathbb{R}_+,$$

are $(\mathcal{F}_t)$ martingales for all functions $f \in \mathcal{A}$. Here it is implicitly assumed that the integral exists almost surely, and defines an integrable random variable.

We have remarked above that in the discrete time case, a process $((X_t), P)$ is a solution to the martingale problem w.r.t. the operators $\mathcal{L}_s = \pi_s - I$ with domain $\mathcal{A} = \mathcal{F}_b(S)$ if and only if it is a Markov chain with one-step transition kernels π_s. Again, in continuous time the situation is much more tricky since the solution to the martingale problem may not be unique, and not all solutions are Markov processes. Indeed, the price to pay in the martingale formulation is that it is usually not easy to establish uniqueness. Nevertheless, if uniqueness holds, and even in cases where uniqueness does not hold, the martingale problem turns out to be a powerful tool for deriving properties of a Markov process in an elegant and general way. This together with stability under weak convergence turns the martingale problem into a fundamental concept in a modern approach to Markov processes.

Example. 1) **Markov chains.** As remarked above, a Markov chain solves the martingale problem for the operators $(\mathcal{L}_t, \mathcal{F}_b(S))$ where $(\mathcal{L}_t f)(x) = \int (f(y) - f(x))\pi_t(x, dy)$.

2) **Continuous time Markov chains.** A continuous time process $X_t = Y_{N_t}$ constructed from a time-homogeneous Markov chain $(Y_n)_{n\in\mathbb{Z}_+}$ with transition kernel π and an independent Poisson process $(N_t)_{t\geq 0}$ solves the martingale problem for the operator $(\mathcal{L}, \mathcal{F}_b(S))$ defined by

$$(\mathcal{L}f)(x) = \int (f(y) - f(x))q(x, dy)$$

where $q(x, dy) = \lambda\pi(x, dy)$ are the jump rates of the process $(X_t)_{t\geq 0}$. More generally, we will construct in Section 5.1 Markov jump processes with general finite time-dependent jump intensities $q_t(x, dy)$.

3) **Diffusion processes.** By Itô's formula, a Brownian motion in $\mathbb{R}^n$ solves the martingale problem for

$$\mathcal{L}f = \frac{1}{2}\Delta f \quad \text{with domain } \mathcal{A} = C_b^2(\mathbb{R}^n).$$

More generally, an Itô diffusion solving the stochastic differential equation (0.11) solves the martingale problem for

$$\mathcal{L}_t f = b(t,x)\cdot\nabla f + \frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,x)\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad \mathcal{A} = C_0^\infty(\mathbb{R}^n),$$

where $a(t,x) = \sigma(t,x)\sigma(t,x)^T$. This is again a consequence of Itô's formula, cf. Stochastic Analysis, e.g. [16, 17].

4. Strong Markov property

For time homogeneous Markov processes in discrete time, the Markov property is equivalent to the following strong Markov property:

Exercise (Strong Markov property for time homogeneous Markov chains). Let (X_n, P_x) be a time homogeneous $(\mathcal{F}_n)$ Markov chain on the state space $(S, \mathcal{B})$ with transition kernel $\pi(x, dy)$. Show that for every $(\mathcal{F}_n^X)$ stopping time $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$, and every $\mathfrak{A}_{\mathrm{can}}$-measurable function $F : \Omega_{\mathrm{can}} \to \mathbb{R}_+$,

$$E\left[F(X_T, X_{T+1}, \dots)|\mathcal{F}_T^X\right] = E_{X_T}[F] \quad P\text{-a.s. on } \{T < \infty\}.$$

For Markov processes in continuous time, corresponding strong Markov properties hold under an additional regularity condition on the transition function.

Exercise (Strong Markov property in continuous time). Suppose that (X_t, P_x) is a time homogeneous $(\mathcal{F}_t)$ Markov process in continuous time with state space $\mathbb{R}^d$ and transition semigroup (p_t).

a) Let T be an $(\mathcal{F}_t)$ stopping time taking only the discrete values $t_i = ih$, $i \in \mathbb{Z}_+$, for some fixed $h \in (0, \infty)$. Prove that for every initial value $x \in \mathbb{R}^d$ and every non-negative measurable function $F : (\mathbb{R}^d)^{[0,\infty)} \to \mathbb{R}$,

$$E_x\left[F(X_{T+\bullet})|\mathcal{F}_T\right] \;=\; E_{X_T}[F(X)] \qquad P_x\text{-almost surely.} \tag{0.18}$$

b) The transition semigroup (p_t) is called *Feller* iff for every $t \geq 0$ and every bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto (p_t f)(x)$ is continuous. Prove that if $t \mapsto X_t(\omega)$ is right continuous for all ω and (p_t) is a Feller semigroup, then the strong Markov property (0.18) holds for every $(\mathcal{F}_t)$ stopping time $T : \Omega \to [0, \infty)$.
Hint: Show first that for any $t \geq 0$ and $f \in C_b(\mathbb{R}^d)$,

$$E_x\left[f(X_{T+t})|\mathcal{F}_T\right] \;=\; E_{X_T}[f(X_t)] \qquad P_x\text{-almost surely.}$$

5. Stability and asymptotic stationarity

A question of fundamental importance in the theory of Markov processes are the long-time stability properties of the process and its transition function. In the time-homogeneous case that we will mostly consider here, many Markov processes approach an equilibrium distribution μ in the long-time limit, i.e.,

$$\mathrm{Law}(X_t) \to \mu \quad \text{ as } t \to \infty \tag{0.19}$$

w.r.t. an appropriate notion of convergence of probability measures. The limit is then necessarily a **stationary distribution** for the transition kernels, i.e.,

$$\mu(B) = (\mu p_t)(B) = \int \mu(dx) p_t(x, B) \quad \text{ for any } t \in I \text{ and } B \in \mathcal{B}.$$

More generally, the laws of the trajectories $X_{t:\infty} = (X_s)_{s \geq t}$ from time t onwards converge to the law P_μ of the Markov process with initial distribution μ, and ergodic averages approach expectations w.r.t. P_μ, i.e.,

$$\frac{1}{t}\sum_{n=0}^{t-1} F(X_n, X_{n+1}, \dots) \to \int_{S^{\mathbb{Z}_+}} F\, dP_\mu, \tag{0.20}$$

$$\frac{1}{t}\int_0^t F(X_{s:\infty})ds \to \int_{\mathcal{D}(\mathbb{R}_+, S)} F\, dP_\mu \quad \text{respectively} \tag{0.21}$$

w.r.t. appropriate notions of convergence.

Statements as in (0.20) and (0.21) are called **ergodic theorems**. They provide far-reaching generalizations of the classical law of large numbers. We will spend a substantial amount of time on proving convergence statements as in (0.19), (0.20) and (0.21) w.r.t. different notions of convergence, and on quantifying the approximation errors asymptotically and non-asymptotically w.r.t. different metrics. This includes studying the existence and uniqueness of stationary distributions. In particular, we will see in Section 5.4 that for Markov processes on infinite dimensional spaces (e.g. interacting particle systems with an infinite number of particles), the non-uniqueness of stationary distributions is often related to a **phase transition**. On spaces with high finite dimension the phase transition will sometimes correspond to a slowdown of the equilibration/mixing properties of the process as the dimension (or some other system parameter) tends to infinity.

In the first part of these notes, we study ergodic properties. We start in Chapter 1 by applying martingale theory to Markov processes in discrete and continuous time. Chapter 2 focuses on ergodic theorems and bounds for ergodic averages as in (0.20) and (0.21). A key idea in the theory of Markov processes is to relate long-time properties of the process to short-time properties described in terms of its generator. Two important approaches for doing this are the coupling/transportation approach considered in Chapter 3, and the L^2/Dirichlet form approach considered in Chapter 9.

In the second part, we outline several approaches for constructing Markov processes in continuous time. Chapter 5 contains direct probabilistic constructions for jump processes with finite jump intensity and for interacting particle systems. In the latter case, there may be infinitely jumps in a finite time interval. Other jump processes with infinite jump intensities (e.g. general Lévy processes as well as jump diffusions) are constructed and analysed in the stochastic analysis course. Chapter 4 discusses the characterization of Markov processes in terms of their generator in depth, and it outlines the construction of Feller processes starting from the generator. Finally, Chapter 6 introduces a general and powerful approach for constructing Markov processes as limits of solutions to approximating martingale problems.

Part I.

Ergodicity

1. Lyapunov functions and stochastic stability

In applications it is often not possible to identify relevant martingales explicitly. Instead one is frequently using supermartingales (or, equivalently, submartingales) to derive upper or lower bounds on expectation values one is interested in. It is then convenient to drop the integrability assumption in the martingale definition:

Definition 1.1 (Non-negative supermartingale). Let $I = \mathbb{Z}_+$ or $I = \mathbb{R}_+$. A real-valued stochastic process $((M_t)_{t\in I}, P)$ is called a **non-negative supermartingale** w.r.t. a filtration $(\mathcal{F}_t)$ if and only if for any $s, t \in I$ with $s \le t$,

(i) $M_t \ge 0 \quad P$-almost surely,

(ii) M_t is $\mathcal{F}_t$-measurable, and

(iii) $E[M_t|\mathcal{F}_s] \le M_s \quad P$-almost surely.

The optional stopping theorem and the supermartingale convergence theorem have versions for non-negative supermartingales. Indeed by Fatou's lemma,

$$E[M_T; T < \infty] \ \le\ \liminf_{n\to\infty} E[M_{T\wedge n}] \ \le\ E[M_0]$$

holds for an **arbitrary** $(\mathcal{F}_t)$ stopping time $T : \Omega \to I \cup \{\infty\}$. Similarly, the limit $M_\infty = \lim_{t\to\infty} M_t$ exists almost surely in $[0, \infty)$.

Lyapunov functions are functions of Markov processes that are supermartingales outside of a bounded subset of the state space. They can be applied effectively in order to prove asymptotic properties of Markov processes such as recurrence, existence of stationary distributions, and different forms of ergodicity.

1.1. Potential theory for Markov chains

Let S be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$. We consider a canonical time-homogeneous Markov chain (X_n, P_x) with state space $(S, \mathcal{B})$ and one-step transition kernel π. The corresponding generator is given by

$$(\mathcal{L}f)(x) = (\pi f)(x) - f(x) = E_x[f(X_1) - f(X_0)]$$

By Theorem 0.7,

$$M_n^{[f]} = f(X_n) - \sum_{i<n} (\mathcal{L}f)(X_i)$$

is a martingale w.r.t. $(\mathcal{F}_n^X)$ and P_x for any $x \in S$ and $f \in \mathcal{F}_b(S)$. Similarly, one easily verifies that if the inequality $\mathcal{L}f \le -c$ holds for non-negative functions $f, c \in \mathcal{F}_+(S)$, then the process

$$M_n^{[f,c]} = f(X_n) + \sum_{i<n} c(X_i)$$

is a non-negative supermartingale w.r.t. $(\mathcal{F}_n^X)$ and P_x for any $x \in S$. By applying optional stopping to these processes, we will derive upper bounds for various expectations of the Markov chain.

Let $D \in \mathcal{B}$ be a measurable subset of S. We define the **exterior boundary** of D w.r.t. the Markov chain as

$$\partial D = \bigcup_{x \in D} \operatorname{supp} \pi(x, \cdot) \setminus D$$

where the support $\operatorname{supp}(\mu)$ of a measure μ on $(S, \mathcal{B})$ is defined as the smallest closed set A such that μ vanishes on A^c. Thus, open sets contained in the complement of $D \cup \partial D$ can not be reached by the Markov chain in a single transition step from D.

Example. (i) For the simple random walk on $\mathbb{Z}^d$, the exterior boundary of a subset $D \subset \mathbb{Z}^d$ is given by

$$\partial D = \{x \in \mathbb{Z}^d \setminus D : |x - y| = 1 \text{ for some } y \in D\}.$$

(ii) For the ball walk on $\mathbb{R}^d$ with transition kernel

$$\pi(x, \cdot) = \operatorname{Unif}(B(x, r)),$$

the exterior boundary of a Borel set $D \in \mathcal{B}$ is the r-neighbourhood

$$\partial D = \{x \in \mathbb{R}^d \setminus D : \operatorname{dist}(x, D) \le r\}.$$

Let

$$T = \min\{n \ge 0 : X_n \in D^c\}$$

denote the first exit time from D. Then for any $x \in D$,

$$X_T \in \partial D \quad P_x\text{-a.s. on } \{T < \infty\}.$$

Our aim is to compute or bound expectations of the form

$$u(x) \; = \; E_x\left[\exp\left(-\sum_{i=0}^{T-1} w(X_i)\right) f(X_T)\,;\, T < \infty\right] + E_x\left[\sum_{n=0}^{T-1} \exp\left(-\sum_{i=0}^{n-1} w(X_i)\right) c(X_n)\right] \tag{1.1}$$

for given non-negative measurable functions $f : \partial D \to \mathbb{R}_+$, and $c, w : D \to \mathbb{R}_+$. The general expression (1.1) combines a number of important probabilities and expectations related to the Markov chain:

Example. (i) **Exit probability from** D**.** $w \equiv 0, c \equiv 0, f \equiv 1$:

$$u(x) = P_x[T < \infty].$$

(ii) **Law of the exit point** X_T**.** $w \equiv 0, c \equiv 0, f = 1_B$ for some measurable subset $B \subset \partial D$:

$$u(x) = P_x[X_T \in B; T < \infty].$$

For instance, if ∂D is the disjoint union of sets A and B and $f = 1_B$ then $u(x) = P_x[T_B < T_A]$.

(iii) **Mean exit time from** D**.** $w \equiv 0, f \equiv 0, c \equiv 1$:

$$u(x) = E_x[T].$$

(iv) **Average occupation time of** B **before exiting** D**.** $w \equiv 0, f \equiv 0, c = 1_B$: $u(x) = G_D(x, B)$, where

$$G_D(x, B) = E_x\left[\sum_{n=0}^{T-1} 1_B(X_n)\right] = \sum_{n=0}^{\infty} P_x[X_n \in B, n < T].$$

G_D is called the **potential kernel** or **Green kernel** of the domain D.

(v) **Laplace transform of mean exit time.** $c \equiv 0, f \equiv 1, w \equiv \lambda$ for some constant $\lambda \geq 0$:

$$u(x) = E_x[\exp(-\lambda T)].$$

(vi) **Laplace transform of occupation time.** $c \equiv 0, f \equiv 1, w = \lambda 1_B$ for some $\lambda > 0, B \subset D$:

$$u(x) = E_x\left[\exp\left(-\lambda \sum_{n=0}^{T-1} 1_B(X_n)\right)\right].$$

The next fundamental theorem shows that supersolutions to an associated boundary value problem provide upper bounds for expectations of the form (1.1). This observation is crucial for studying stability properties of Markov chains.

Theorem 1.2 (Maximum principle). Suppose $v \in \mathcal{F}_+(S)$ is a non-negative function satisfying

$$\begin{aligned} \mathcal{L}v &\leq (e^w - 1)v - e^w c && \text{on } D, \\ v &\geq f && \text{on } \partial D. \end{aligned} \tag{1.2}$$

Then $u \leq v$.

The proof will be given below. It is an application of the optional stopping theorem for non-negative supermartingales. The expectation $u(x)$ can be identified precisely as the minimal non-negative solution of the corresponding boundary value problem:

Theorem 1.3 (Dirichlet problem, Poisson equation, Feynman-Kac formula). The function u is the *minimal non-negative solution* of the boundary value problem

$$\begin{aligned} \mathcal{L}v &= (e^w - 1)v - e^w c && \text{on } D, \\ v &= f && \text{on } \partial D. \end{aligned} \tag{1.3}$$

If $c \equiv 0$, f is bounded, and $T < \infty$ P_x-almost surely for any $x \in S$, then u is the **unique bounded solution** of (1.3).

We first prove both theorems in the case $w \equiv 0$. The extension to the general case will be discussed afterwards. The proof for $w \equiv 0$ is based on the following simple observation:

Lemma 1.4 (Locally superharmonic functions and supermartingales). *Let $A \in \mathcal{B}$ and suppose that $V \in \mathcal{F}_+(S)$ is a non-negative function satisfying*

$$\mathcal{L}V \leq -c \quad \text{on } S \setminus A$$

for a non-negative function $c \in \mathcal{F}_+(S \setminus A)$. Then the process

$$M_n = V(X_{n \wedge T_A}) + \sum_{i < n \wedge T_A} c(X_i) \tag{1.4}$$

is a non-negative supermartingale.

The elementary proof of the lemma is left as an exercise.

Proof (of Theorem 1.2 for $w \equiv 0$). Let $v \in \mathcal{F}_+(S)$ such that $\mathcal{L}v \leq -c$ on D. Then by the lemma,

$$M_n = V(X_{n\wedge T}) + \sum_{i<n\wedge T} c(X_i)$$

is a non-negative supermartingale. In particular, (M_n) converges almost surely to a limit $M_\infty \geq 0$, and thus M_T is defined and non-negative even on $\{T = \infty\}$. If $v \geq f$ on ∂D then

$$M_T \geq f(X_T)1_{\{T<\infty\}} + \sum_{i=0}^{T-1} c(X_i). \tag{1.5}$$

Therefore, by optional stopping combined with Fatou's lemma,

$$u(x) \leq E_x[M_T] \leq E_x[M_0] = v(x). \tag{1.6}$$

■

Proof (of Theorem 1.3 for $w \equiv 0$). By Theorem 1.2, all non-negative solutions v of (1.3) dominate u from above. This proves minimality. Moreover, if $c \equiv 0$, f is bounded, and $T < \infty$ P_x-a.s. for any x, then (M_n) is a bounded martingale, and hence all inequalities in (1.5) and (1.6) are equalities. Thus if a non-negative solution of (1.3) exists then it coincides with u, i.e., uniqueness holds.

It remains to verify that u satisfies (1.19). This can be done by conditioning on the first step of the Markov chain: For $x \in D$, we have $T \geq 1$ P_x-almost surely. In particular, if $T < \infty$ then X_T coincides with the exit point of the shifted Markov chain $(X_{n+1})_{n\geq 0}$, and $T-1$ is the exit time of (X_{n+1}). Therefore, the Markov property implies that P_x-almost surely,

$$\begin{aligned} & E_x\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T} c(X_n) \,\middle|\, X_1 \right] \\ = \;& c(x) + E_x\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T-1} c(X_{n+1}) \,\middle|\, X_1 \right] \\ = \;& c(x) + E_{X_1}\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T} c(X_n) \right] \\ = \;& c(x) + u(X_1), \end{aligned}$$

and hence

$$u(x) = E_x\left[c(x) + u(X_1)\right] = c(x) + (pu)(x),$$

i.e., $\mathcal{L}u(x) = -c(x)$. Moreover, for $x \in \partial D$, we have $T = 0$ P_x-almost surely and hence

$$u(x) = E_x[f(X_0)] = f(x).$$

■

We now extend the results to the case $w \not\equiv 0$. This can be done by representing the expectation in (1.3) as a corresponding expectation with $w \equiv 0$ for a Markov chain with finite life-time:

Proof (Reduction of general case to $w \equiv 0$). We consider the Markov chain (X_n^w) with death rate w defined on the extended state space $S\dot{\cup}\{\Delta\}$ by $X_0^w = X_0$,

$$X_{n+1}^w = \begin{cases} X_{n+1} & \text{if } X_n^w \neq \Delta \text{ and } E_{n+1} \geq w(X_n), \\ \Delta & \text{otherwise}, \end{cases}$$

with independent Exp(1) distributed random variables E_i ($i \in \mathbb{N}$) that are independent of (X_n) as well. Setting $f(\Delta) = c(\Delta) = 0$ one easily verifies that

$$u(x) = E_x[f(X_T^w); T < \infty] + E_x[\sum_{n=0}^{T-1} c(X_n^w)].$$

By applying Theorems 1.2 and 1.3 with $w \equiv 0$ to the Markov chain (X_n^w), we see that u is the minimal non-negative solution of

$$\mathcal{L}^w u = -c \quad \text{on } D, \quad u = f \quad \text{on } \partial D, \tag{1.7}$$

and any non-negative supersolution v of (1.7) dominates u from above. Moreover, the boundary value problem (1.7) is equivalent to (1.3) since

$$\mathcal{L}^w u = e^{-w}\pi u - u = e^{-w}\mathcal{L}u + (e^{-w} - 1)u = -c \quad \text{if and only if} \quad \mathcal{L}u = (e^w - 1)u - e^w c.$$

This proves Theorem 1.2 and the main part of Theorem 1.3 in the case $w \not\equiv 0$. The proof of the last assertion of Theorem 1.3 is left as an exercise. ■

Example (Random walks with bounded steps). We consider a random walk on $\mathbb{R}$ with transition step $x \mapsto x + W$ where the increment $W : \Omega \to \mathbb{R}$ is a bounded random variable, i.e., $|W| \le r$ for some constant $r \in (0, \infty)$. Our goal is to derive tail estimates for passage times.

$$T_a = \min\{n \ge 0 : X_n \ge a\}.$$

Note that T_a is the first exit time from the domain $D = (-\infty, a)$. Since the increments are bounded by r, $\partial D \subset [a, a + r]$. Moreover, the moment generating function $Z(\lambda) = E[\exp(\lambda W)]$,
$\lambda \in \mathbb{R}$, is bounded by $e^{\lambda r}$, and for $\lambda \le 0$, the function $v(x) = e^{\lambda x}$ satisfies

$$\begin{aligned} (\mathcal{L}v)(x) &= E_x\left[e^{\lambda(x+W)}\right] - e^{\lambda x} = (Z(\lambda) - 1)\, v(x) \quad \text{for } x \in D, \\ v(x) &\ge e^{\lambda(a+r)} \qquad \text{for } x \in \partial D. \end{aligned}$$

By applying Theorem 1.2 with the constant functions w and f satisfying $e^{w(x)} \equiv Z(\lambda)$ and $f(x) \equiv e^{\lambda(a+r)}$ we conclude that for any $x \in \mathbb{R}$,

$$E_x\left[Z(\lambda)^{-T_a} e^{\lambda(a+r)}; T_a < \infty\right] \le e^{\lambda x} \tag{1.8}$$

Indeed, it can be verified that the second part of the theorem applies even if w is a negative constant. We now distinguish cases:

(i) $E[W] > 0$: In this case, by the law of large numbers, $X_n \to \infty$ P_x-a.s., and hence $P_x[T_a < \infty] = 1$ for any $x \in \mathbb{R}$. Moreover, for $\lambda < 0$ with $|\lambda|$ sufficiently small,

$$Z(\lambda) = E[e^{\lambda W}] = 1 + \lambda E[W] + O(\lambda^2) < 1.$$

Therefore, (1.8) yields the exponential moment bound

$$E_x\left[\left(\tfrac{1}{Z(\lambda)}\right)^{T_a}\right] \le e^{-\lambda(a+r-x)} \tag{1.9}$$

for any $x \in \mathbb{R}$ and $\lambda < 0$ as above. In particular, by Markov's inequality, the passage time T_a has exponential tails:

$$P_x[T_a \ge n] \le Z(\lambda)^n E_x[Z(\lambda)^{-T_a}] \le Z(\lambda)^n e^{-\lambda(a+r-x)}.$$

(ii) $E[W] = 0$: In this case, we may have $Z(\lambda) \ge 1$ for any $\lambda \in \mathbb{R}$, and thus we can not apply the argument above. Indeed, it is well known that for instance for the simple random walk on $\mathbb{Z}$ even the first moment $E_x[T_a]$ is infinite, cf. [18]. However, we may apply a similar approach as above to the exit time $T_{\mathbb{R}\setminus(-a,a)}$ from a finite interval. We assume that W has a symmetric distribution, i.e., $W \sim -W$. By choosing $u(x) = cos(\lambda x)$ for some $\lambda > 0$ with $\lambda(a + r) < \pi/2$, we obtain

$$\begin{aligned} (\mathcal{L}u)(x) &= E[cos(\lambda x + \lambda W)] - cos(\lambda x) \\ &= cos(\lambda x)E[cos(\lambda W)] + sin(\lambda x)E[sin(\lambda W)] - cos(\lambda x) \\ &= (C(\lambda) - 1)\, cos(\lambda x) \end{aligned}$$

where $C(\lambda) := E[cos(\lambda W)]$, and $cos(\lambda x) \ge cos\,(\lambda(a + r)) > 0$ for $x \in \partial(-a, a)$. Here we have used that $\partial(-a, a) \subset [-a - r, a + r]$ and $\lambda(a + r) < \pi/2$. If W does not vanish almost surely then $C(\lambda) < 1$ for sufficiently small λ. Hence we obtain similarly as above the exponential tail estimate

$$P_x\left[T_{(-a,a)^c} \ge n\right] \le C(\lambda)^n E\left[C(\lambda)^{-T_{(-a,a)^c}}\right] \le C(\lambda)^n \frac{cos(\lambda x)}{cos(\lambda(a + r))} \qquad \text{for } |x| < a.$$

1.2. Lyapunov functions and recurrence

The results in the last section already indicated that superharmonic functions can be used to control stability properties of Markov chains, i.e., they can serve as stochastic Lyapunov functions. This idea will be developed systematically in this and the next section. As before we consider a time-homogeneous Markov chain (X_n, P_x) with generator $\mathcal{L} = \pi - I$ on a Polish state space S endowed with the Borel σ-algebra $\mathcal{B}$.

Recurrence of sets

The first return time to a set A is given by

$$T_A^+ = \inf\{n \geq 1 : X_n \in A\}.$$

Notice that

$$T_A = T_A^+ \cdot 1_{\{X_0 \notin A\}},$$

i.e., the first hitting time and the first return time coincide if and only if the chain is not started in A.

Definition 1.5 (Harris recurrence and positive recurrence). A set $A \in \mathcal{B}$ is called **Harris recurrent** iff

$$P_x[T_A^+ < \infty] = 1 \quad \text{for any } x \in A.$$

It is called **positive recurrent** iff

$$E_x[T_A^+] < \infty \quad \text{for any } x \in A.$$

The name "Harris recurrence" is used to be able to differentiate between several possible notions of recurrence that are all equivalent on a discrete state space but not necessarily on a general state space, cf. Meyn and Tweedie [40]. Harris recurrence is the most widely used notion of recurrence on general state spaces. By the strong Markov property, the following alternative characterisations holds:

Exercise. Prove that a set $A \in \mathcal{B}$ is Harris recurrent if and only if

$$P_x[X_n \in A \text{ infinitely often}] = 1 \quad \text{for any } x \in A.$$

We will now show that the existence of superharmonic functions with certain properties provides sufficient conditions for non-recurrence, Harris recurrence and positive recurrence, respectively. Below, we will see that for irreducible Markov chains on countable spaces, these conditions are essentially sharp. The conditions are:

(LT) There exists a function $V \in \mathcal{F}_+(S)$ and $y \in S$ such that

$$\mathcal{L}V \leq 0 \text{ on } A^c \qquad \text{and} \qquad V(y) < \inf_A V.$$

(LR) There exists a function $V \in \mathcal{F}_+(S)$ such that

$$\mathcal{L}V \leq 0 \text{ on } A^c \text{ and } T_{\{V>c\}} < \infty \quad P_x\text{-a.s. for any } x \in S \qquad \text{and} \qquad c \geq 0.$$

(LP) There exists a function $V \in \mathcal{F}_+(S)$ such that

$$\mathcal{L}V \leq -1 \text{ on } A^c \qquad \text{and} \qquad \pi V < \infty \text{ on } A.$$

Theorem 1.6. (Foster-Lyapunov conditions for non-recurrence, Harris recurrence and positive recurrence)

(i) If (LT) holds then
$$P_y[T_A < \infty] \;\leq\; V(y)/\inf_A V \;<\; 1.$$

(ii) If (LR) holds then
$$P_x[T_A < \infty] \;=\; 1 \qquad \text{for any } x \in S.$$
In particular, the set A is Harris recurrent.

(iii) If (LP) holds then
$$\begin{aligned} E_x[T_A] \;&\leq\; V(x) \;<\; \infty \qquad \text{for any } x \in A^c, \text{ and} \\ E_x[T_A^+] \;&\leq\; (\pi V)(x) \;<\; \infty \qquad \text{for any } x \in A. \end{aligned}$$
In particular, the set A is positive recurrent.

Proof. (i) If $\mathcal{L}V \leq 0$ on A^c then by Lemma 1.4, the process $M_n = V(X_{n\wedge T_A})$ is a non-negative supermartingale w.r.t. P_x for any x. Hence by optional stopping and Fatou's lemma,
$$V(y) = E_y[M_0] \geq E_y[M_{T_A}; T_A < \infty] \geq P_y[T_A < \infty] \cdot \inf_A V.$$
Assuming (LT), we obtain $P_y[T_A < \infty] < 1$.

(ii) Now assume that (LR) holds. Then by applying optional stopping to (M_n), we obtain
$$V(x) = E_x[M_0] \geq E_x[M_{T_{\{V>c\}}}] = E_x[V(X_{T_A \wedge T_{\{V>c\}}})] \geq cP_x[T_A = \infty]$$
for any $c > 0$ and $x \in S$. Here we have used that $T_{\{V>c\}} < \infty$ P_x-almost surely and hence $V(X_{T_A\wedge T_{\{V>c\}}}) \geq c$ P_x-almost surely on $\{T_A = \infty\}$. By letting c tend to infinity, we conclude that $P_x[T_A = \infty] = 0$ for any x.

(iii) Finally, suppose that $\mathcal{L}V \leq -1$ on A^c. Then by Lemma 1.4,
$$M_n = V(X_{n\wedge T_A}) + n \wedge T_A$$
is a non-negative supermartingale w.r.t. P_x for any x. In particular, (M_n) converges P_x-almost surely to a finite limit, and hence $P_x[T_A < \infty] = 1$. Thus by optional stopping and since $V \geq 0$,
$$E_x[T_A] \leq E_x[M_{T_A}] \leq E_x[M_0] = V(x) \qquad \text{for any } x \in S. \tag{1.10}$$
Moreover, we can also estimate the first return time by conditioning on the first step. Indeed, for $x \in A$ we obtain by (1.10):
$$E_x[T_A^+] = E_x\left[E_x[T_A^+|X_1]\right] = E_x\left[E_{X_1}[T_A]\right] \leq E_x[V(X_1)] = (\pi V)(x)$$
Thus A is positive recurrent if (LP) holds. ∎

Example (State space model on $\mathbb{R}^d$). We consider a simple state space model with one-step transition

$$x \mapsto x + hb(x) + \sqrt{h}W$$

where h is a positive constant, $b : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable vector field and $W : \Omega \to \mathbb{R}^d$ is a square-integrable random vector with $E[W] = 0$ and $\mathrm{Cov}(W^i, W^j) = \delta_{ij}$. The corresponding Markov chain arises as the Euler discretization of the stochastic differential equation $dX_t = b(X_t)\,dt + dB_t$ driven by a d-dimensional Brownian motion (B_t). As a Lyapunov function we try

$$V(x) = |x|^2/\varepsilon$$

where ε is a positive constant. A simple calculation shows that

$$\begin{aligned}\varepsilon(\mathcal{L}V)(x) &= E\left[|x + hb(x) + \sqrt{h}W|^2\right] - |x|^2 = |x + hb(x)|^2 + h\,E[|W|^2] - |x|^2 \\ &= (2x \cdot b(x) + h|b(x)|^2 + d)\,h.\end{aligned}$$

Therefore, the condition $\mathcal{L}V(x) \le -1$ is satisfied if and only if

$$2x \cdot b(x) + h|b(x)|^2 + d \le -\varepsilon/h.$$

By choosing ε small enough we see that positive recurrence holds for the ball $B(0, r)$ with r sufficiently large provided

$$\limsup_{|x|\to\infty} \left(2x \cdot b(x) + h|b(x)|^2\right) < -d. \tag{1.11}$$

This condition is satisfied in particular if outside of a ball, the radial component $b_r(x) = \frac{x}{|x|} \cdot b(x)$ of the drift satisfies $(1 - \delta)b_r(x) \le -\frac{d}{2|x|}$ for some $\delta > 0$, and $h|b(x)|^2 \le -\delta|x|b_r(x)$.

Exercise. Derive a sufficient condition similar to (1.11) for positive recurrence of state space models with transition step

$$x \mapsto x + hb(x) + \sqrt{h}\sigma(x)W$$

where b and W are chosen as in the example above, and σ is a measurable function from $\mathbb{R}^d$ to $\mathbb{R}^{d\times d}$.

Example (Recurrence and transience for the simple random walk on $\mathbb{Z}^d$). The simple random walk is the Markov chain on $\mathbb{Z}^d$ with transition probabilities $\pi(x, y) = \frac{1}{2d}$ if $|x - y| = 1$, and $\pi(x, y) = 0$ otherwise. The generator is given by

$$(\mathcal{L}f)(x) = \frac{1}{2d}(\Delta_{\mathbb{Z}^d} f)(x) = \frac{1}{2d}\sum_{i=1}^{d}\left[(f(x + e_i) - f(x)) - (f(x) - f(x - e_i))\right].$$

In order to find suitable Lyapunov functions, we approximate the discrete Laplacian on $\mathbb{Z}^d$ by the Laplacian on $\mathbb{R}^d$. By Taylor's theorem, for $f \in C^4(\mathbb{R}^d)$,

$$\begin{aligned}f(x + e_i) - f(x) &= \partial_i f(x) + \frac{1}{2}\partial_{ii}^2 f(x) + \frac{1}{6}\partial_{iii}^3 f(x) + \frac{1}{24}\partial_{iiii}^4 f(\xi),\\ f(x - e_i) - f(x) &= -\partial_i f(x) + \frac{1}{2}\partial_{ii}^2 f(x) - \frac{1}{6}\partial_{iii}^3 f(x) + \frac{1}{24}\partial_{iiii}^4 f(\eta),\end{aligned}$$

where ξ and η are intermediate points on the line segments between x and $x + e_i$, x and $x - e_i$, respectively. Adding these $2d$ equations, we see that

$$\Delta_{\mathbb{Z}^d} f(x) = \Delta f(x) + R(x), \qquad \text{where} \qquad |R(x)| \le \frac{d}{12}\sup_{B(x,1)} \|\partial^4 f\|. \tag{1.12}$$

This suggests to choose Lyapunov functions that are close to harmonic functions on $\mathbb{R}^d$ outside a ball. However, since there is a perturbation involved, we will not be able to use exactly harmonic functions, but we will have to choose functions that are strictly superharmonic instead. We try

$$V(x) = |x|^p \qquad \text{for some } p \in \mathbb{R}.$$

By the expression for the Laplacian in polar coordinates,

$$\Delta V(x) = \left(\frac{d^2}{dr^2} + \frac{d-1}{r}\frac{d}{dr}\right) r^p = p \cdot (p-1+d-1)\, r^{p-2}$$

where $r = |x|$. In particular, V is superharmonic on the complement of a ball if and only if $p \in [0, 2-d]$ or $p \in [2-d, 0]$, respectively. The perturbation term can be controlled by noting that there exists a finite constant C such that outside a ball,

$$\|\partial^4 V(x)\| \le C \cdot |x|^{p-4}.$$

This bound shows that the approximation of the discrete Laplacian by the Laplacian on $\mathbb{R}^d$ improves if $|x|$ is large. Indeed, by (1.12), we obtain

$$\mathcal{L}V(x) = \frac{1}{2d}\Delta_{\mathbb{Z}^d} V(x) \le \frac{p}{2d}(p+d-2) r^{p-2} + \frac{C}{2d} r^{p-4}.$$

Thus V is superharmonic for $\mathcal{L}$ outside a ball provided $p \in (0, 2-d)$ or $p \in (2-d, 0)$, respectively. We now distinguish cases:

$d > 2$: In this case we can choose $p < 0$ such that $\mathcal{L}V \le 0$ outside some ball $B(0, r_0)$. Since r^p is decreasing, we have

$$V(x) < \inf_{B(0,r_0)} V \qquad \text{for any } x \text{ with } |x| > r_0,$$

and hence by Theorem 1.6,

$$P_x[T_{B(0,r_0)} < \infty] < 1 \qquad \text{whenever } |x| > r_0.$$

Theorem 1.8 below shows that this implies that every finite set is transient, i.e., it is almost surely visited only finitely many times by the random walk with an arbitrary starting point.

$d < 2$: In this case we can choose $p \in (0, 2-d)$ to obtain $\mathcal{L}V \le 0$ outside some ball $B(0, r_0)$. Now $V(x) \to \infty$ as $|x| \to \infty$. Since $\limsup |X_n| = \infty$ almost surely, we see that

$$T_{\{V>c\}} < \infty \qquad P_x\text{-almost surely for any } x \in \mathbb{Z}^d \text{ and } c \in \mathbb{R}_+.$$

Therefore, by Theorem 1.6, the ball $B(0, r_0)$ is (Harris) recurrent. By irreducibility this implies that any state $x \in \mathbb{Z}^d$ is recurrent, cf. Theorem 1.8 below.

$d = 2$: This is the critical case and therefore more delicate. The Lyapunov functions considered above can not be used. Since a rotationally symmetric harmonic function for the Laplacian on $\mathbb{R}^2$ is $\log |x|$, it is natural to try choosing $V(x) = (\log |x|)^\alpha$ for some $\alpha \in \mathbb{R}_+$. Indeed, one can show by choosing α appropriately that the Lyapunov condition for recurrence is satisfied in this case as well:

Exercise (Recurrence of the two-dimensional simple random walk). Show by choosing an appropriate Lyapunov function that the simple random walk on $\mathbb{Z}^2$ is recurrent.

Exercise (Recurrence and transience of Brownian motion). Let $\big((B_t)_{t\in[0,\infty)}, P_x\big)$ be an $\mathbb{R}^d$-valued Brownian motion starting at x, and let $T_a = \inf\{t \ge 0 : |B_t| = a\}$.

a) Compute $P_x[T_a < T_b]$ for $a < |x| < b$.

b) Show that for $d \le 2$, Brownian motion is recurrent in the sense that $P_x[T_a < \infty] = 1$ for any $a < |x|$.

c) Show that for $d \ge 3$, Brownian motion is transient in the sense that $P_x[T_a < \infty] \to 0$ as $|x| \to \infty$.

Global recurrence

For irreducible Markov chains on countable state spaces, recurrence respectively transience of an arbitrary finite set already implies that recurrence resp. transience holds for every finite set. This allows to show that the Lyapunov conditions for recurrence and transience are both necessary and sufficient. On general state spaces this is not necessarily true, and proving corresponding statements under appropriate conditions is much more delicate. We recall the results on countable state spaces, and we state a result on general state spaces without proof. For a thorough treatment of recurrence properties for Markov chains on general state spaces we refer to the monograph “Markov chains and stochastic stability” by Meyn and Tweedie [40].

a) Countable state space

Suppose that $\pi(x,y) = \pi(x,\{y\})$ are the transition probabilities of a homogeneous Markov chain (X_n, P_x) taking values in a countable set S, and let T_y and T_y^+ denote the first hitting resp. return time to a set $\{y\}$ consisting of a single state $y \in S$.

Definition 1.7 (Irreducibility on countable state spaces). The transition matrix π and the Markov chain (X_n, P_x) are called **irreducible** if and only if

(i) $\forall x, y \in S : \exists n \in \mathbb{Z}_+ : \pi^n(x,y) > 0,$ or, equivalently, if and only if

(ii) $\forall x, y \in S : P_x[T_y < \infty] > 0.$

If the transition matrix is irreducible then recurrence and positive recurrence of different states are equivalent to each other, since between two visits to a recurrent state the Markov chain will visit any other state with positive probability:

Theorem 1.8 (Recurrence and positive recurrence of irreducible Markov chains). Suppose that S is countable and the transition matrix p is irreducible.

1) The following statements are all equivalent:
   (i) There exists a finite recurrent set $A \subseteq S$.
   (ii) For any $x \in S$, the set $\{x\}$ is recurrent.
   (iii) For any $x, y \in S$,
   $$P_x[X_n = y \text{ infinitely often }] = 1.$$
2) The following statements are all equivalent:
   (i) There exists a finite positive recurrent set $A \subset S$.
   (ii) For any $x \in S$, the set $\{x\}$ is positive recurrent.
   (iii) For any $x, y \in S$,
   $$E_x[T_y] < \infty.$$

The proof is left as an exercise, see also the lecture notes on "Stochastic Processes" [18]. The Markov chain is called **(globally) recurrent** iff the equivalent conditions in 1) hold, and transient iff these conditions do not hold. Similarly, it is called **(globally) positive recurrent** iff the conditions in 2) are satisfied. By the example above, for $d \leq 2$ the simple random walk on $\mathbb{Z}^d$ is globally recurrent, however it is not positive recurrent. For $d \geq 3$ it is transient.

As a consequence of Theorem 1.8, we obtain Lyapunov conditions for transience, recurrence and positive recurrence on a countable state space that are both necessary and sufficient:

Corollary 1.9 (Foster-Lyapunov conditions for transience and recurrence on a countable state space). Suppose that S is countable and the transition matrix p is irreducible. Then:

1) The Markov chain is transient if and only if there exists a finite set $A \subseteq S$ and a function $V \in \mathcal{F}_+(S)$ such that (LT) holds.

2) The Markov chain is recurrent if and only if there exists a finite set $A \subseteq S$ and a function $V \in \mathcal{F}_+(S)$ such that

$$(LR') \quad \mathcal{L}V \le 0 \text{ on } A^c, \text{ and } \{V \le c\} \text{ is finite for any } c \in \mathbb{R}_+.$$

3) The Markov chain is positive recurrent if and only if there exists a finite set $A \subseteq S$ and a function $V \in \mathcal{F}_+(S)$ such that (LP) holds.

Proof. Sufficiency of the Lyapunov conditions follows directly by Theorems 1.6 and 1.8: If (LT) holds then by Theorem 1.6, there exists $y \in S$ such that $P_y[T_A < \infty]$, and hence the Markov chain is transient by Theorem 1.8. Similarly, if (LP) holds then A is positive recurrent by Theorem 1.6, and hence global positive recurrence holds by Theorem 1.8. Finally, if (LR') holds and the state space is not finite, then for every $c \in \mathbb{R}_+$, the set $\{V > c\}$ is not empty. Therefore, (LR) holds by irreducibility, and the recurrence follows again from Theorems 1.6 and 1.8. If S is finite then every irreducible chain is globally recurrent.

We now prove that the Lyapunov conditions are also necessary.

1) If the Markov chain is transient then we can find a state $x \in S$ and a finite set $A \subseteq S$ such that the function $V(x) = P_x[T_A < \infty]$ satisfies

$$V(x) < 1 = \inf_A V.$$

By Theorem 1.3, V is harmonic on A^c and thus (LT) is satisfied.

2) Now suppose that the Markov chain is recurrent. If S is finite then (LR') holds with $A = S$ for an arbitrary function $V \in \mathcal{F}_+(S)$. If S is not finite then we choose a finite set $A \subseteq S$ and an arbitrary decreasing sequence of sets $D_n \subseteq S$ such that $A \subseteq S \setminus D_1$, the set $S \setminus D_n$ is finite for all n, and $\bigcap D_n = \emptyset$, and we set

$$V_n(x) = P_x[T_{D_n} < T_A].$$

Then $V_n \equiv 1$ on D_n, and as $n \to \infty$,

$$V_n(x) \searrow P_x[T_A = \infty] = 0 \qquad \text{for any } x \in S.$$

Since S is countable, we can apply a diagonal argument to extract a subsequence such that

$$V(x) := \sum_{k=1}^{\infty} V_{n_k}(x) < \infty \qquad \text{for any } x \in S.$$

By Theorem 1.3, the functions V_n are harmonic on $S \setminus (A \cup D_n)$. Moreover, for $x \in D_n$,

$$\mathcal{L}V_n(x) \;=\; E_x[V_n(X_1)] - V(x) \;\le\; 1 - 1 \;=\; 0.$$

Hence the functions V_n and V are superharmonic on $S \setminus A$. Moreover, $V \ge k$ on D_{n_k}. Thus the sub-level sets of V are finite, and (LR') is satisfied.

3) Finally if the chain is positive recurrent then for an arbitrary finite set $A \subseteq S$, the function $V(x) = E_x[T_A]$ is finite and satisfies $\mathcal{L}V = -1$ on A^c. Since

$$(\pi V)(x) = E_x\left[E_{X_1}[T_A]\right] = E_x\left[E_x[T_A^+ | X_1]\right] = E_x[T_A^+] < \infty$$

for any x, condition (LP) is satisfied. ■

b) Extension to locally compact state spaces

Extensions of Corollary 1.9 to general state spaces are not trivial. Suppose for example that S is a separable metric space that is **locally compact**, i.e., every point $x \in S$ has a compact neighbourhood. Let π be a transition kernel on $(S, \mathcal{B})$, and let λ be a positive measure on $(S, \mathcal{B})$ with full support, i.e., $\lambda(B) > 0$ for any non-empty open set $B \subset S$. For instance, $S = \mathbb{R}^d$ and λ is the Lebesgue measure.

Definition 1.10 (λ-irreducibility). The transition kernel π is called **λ-irreducible** if and only if for any $x \in S$ and for any Borel set $A \in \mathcal{B}$ with $\lambda(A) > 0$, there exists $n \in \mathbb{Z}_+$ such that $\pi^n(x, A) > 0$.

One of the difficulties on general state spaces is that there are different concepts of irreducibility. In general, λ-irreducibility is a strictly stronger condition than **topological irreducibility** which means that every non-empty open set $B \subset S$ is accessible from any state $x \in S$. The following equivalences are proven in Chapter 9 of Meyn and Tweedie [40]:

Theorem 1.11 (Necessary and sufficient conditions for Harris recurrence). Suppose that S is separable and locally compact with Borel σ-algebra $\mathcal{B}$, and π is a λ-irreducible transition kernel on $(S, \mathcal{B})$. Moreover, suppose that for every function $f \in C_b(S)$, πf is continuous (Feller property). Then the following statements are all equivalent:

(i) There exists a compact set $K \subseteq S$ and a function $V \in \mathcal{F}_+(S)$ such that

$$(LR'') \qquad \mathcal{L}V \le 0 \text{ on } K^c, \text{ and } \overline{\{V \le c\}} \text{ is compact for any } c \in \mathbb{R}_+.$$

(ii) $P_x[X_n \text{ visits every compact set only finitely many times}] = 0 \quad$ for any $x \in S$.

(iii) Every non-empty open ball $B \subset S$ is Harris recurrent.

(iv) For any $x \in S$ and any set $A \in \mathcal{B}$ with $\lambda(A) > 0$,

$$P_x[X_n \in A \text{ infinitely often }] = 1.$$

The result summarizes Theorems 9.4.1 ("(i)⇒(ii)"), 9.2.2 ("(ii)⇒(iii)"), 9.1.4 ("(iii)⇒(iv)") and 9.4.2 ("(iv)⇒(i)") in [40]. For a part of the implications, the assumptions can be relaxed. For example, the Feller property is only required for the implications "(i)⇒(iv)" and "(ii)⇒(iii)", and in the second case it can be replaced by a weaker condition. The idea of the proof is to show at first that for every compact set $K \subset S$, there exist a probability mass function (a_n) on $\mathbb{Z}_+$, a probability measure ν on $(S, \mathcal{B})$, and a constant $\varepsilon > 0$ such that the minorization condition

$$\sum_{n=0}^{\infty} a_n \pi^n(x, \cdot) \ \ge \ \varepsilon \nu \tag{1.13}$$

holds for any $x \in K$. In the theory of Markov chains on general state spaces, a set K with this property is called **petite**. Given a petite set K and a Lyapunov condition on K^c one can then find a strictly increasing sequence of regeneration times T_n ($n \in \mathbb{N}$) such that the law of X_{T_n} dominates the measure $\varepsilon\nu$ from above. By the strong Markov property, the Markov chain makes a "fresh start" with probability ε at each of the regeneration times, and during each excursion between two fresh starts it visits a given set A satisfying $\lambda(A) > 0$ with a fixed strictly positive probability.

1.3. Invariant probability measures

A central topic in Markov chain theory is the existence, uniqueness and convergence of Markov chains to stationary distributions. To this end we will consider different topologies and metrics on the space $\mathcal{P}(S)$ of probability measures on a Polish space S endowed with its Borel σ-algebra $\mathcal{B}$. In this section, we study weak convergence of probability measures, and applications to existence of invariant probability measures. Convergence in Wasserstein and total variation metrics will be considered in Chapter 3. A useful additional reference for this section is the classical monograph "Convergence of probability measures" by Billingsley [3].

Recall that $\mathcal{P}(S)$ is a convex subset of the vector space

$$\mathcal{M}(S) = \{\alpha\mu_+ - \beta\mu_- : \mu_+, \mu_- \in \mathcal{P}(S), \alpha, \beta \geq 0\}$$

consisting of all finite signed measures on $(S, \mathcal{B})$. By $\mathcal{M}_+(S)$ we denote the set of all (not necessarily finite) non-negative measures on $(S, \mathcal{B})$. For a measure μ and a measurable function f we set

$$\mu(f) = \int f\, d\mu \quad \text{whenever the integral exists.}$$

Definition 1.12 (Invariant measure, stationary distribution). A measure $\mu \in \mathcal{M}_+(S)$ is called **invariant** w.r.t. a transition kernel π on $(S, \mathcal{B})$ iff $\mu\pi = \mu$, i.e., iff

$$\int \mu(dx)\, \pi(x, B) = \mu(B) \quad \text{for any } B \in \mathcal{B}.$$

An invariant probability measure is also called a **stationary (initial) distribution** or an **equilibrium distribution** of π.

Exercise. Show that the set of invariant probability measures for a given transition kernel π is a convex subset of $\mathcal{P}(S)$.

Below, we are going to prove the existence of an invariant probability measure μ for a given transition kernel π as a subsequential limit of Césaro averages of the form $\mu_n = \frac{1}{n}\sum_{i<n} \nu\pi^i$, where ν is an arbitrary initial distribution. We first need some preparations on weak convergence and the existence of subsequential limits for sequences of probability measures.

Weak convergence of probability measures

Recall that a sequence $(\mu_k)_{k\in\mathbb{N}}$ of probability measures on $(S, \mathcal{B})$ is said to **converge weakly** to a measure $\mu \in \mathcal{P}(S)$ if and only if

(i) $\mu_k(f) \to \mu(f)$ for any $f \in C_b(S)$.

The **Portemanteau Theorem** states that weak convergence is equivalent to each of the following properties:

(ii) $\mu_k(f) \to \mu(f)$ for any uniformly continuous $f \in C(S)$.

(iii) $\limsup \mu_k(A) \leq \mu(A)$ for any closed set $A \subset S$.

(iv) $\liminf \mu_k(O) \geq \mu(O)$ for any open set $O \subset S$.

(v) $\limsup \mu_k(f) \leq \mu(f)$ for any upper semicontinuous function $f : S \to \mathbb{R}$ that is bounded from above.

(vi) $\liminf \mu_k(f) \geq \mu(f)$ for any lower semicontinuous function $f : S \to \mathbb{R}$ that is bounded from below.

(vii) $\mu_k(f) \to \mu(f)$ for any function $f \in \mathcal{F}_b(S)$ that is continuous at μ-almost every $x \in S$.

For the proof see e.g. the monographs by Billingsley [3] or Stroock [52, Theorem 3.1.5]. The following observation is crucial for studying weak convergence on Polish spaces:

Remark (Polish spaces as measurable subset of $[0,1]^{\mathbb{N}}$). Suppose that (S, ϱ) is a separable metric space, and $\{x_n : n \in \mathbb{N}\}$ is a countable dense subset. Then the map $h : S \to [0,1]^{\mathbb{N}}$,

$$h(x) \;=\; (\varrho(x, x_n) \wedge 1)_{n\in\mathbb{N}} \tag{1.14}$$

is a homeomorphism from S to $h(S)$ provided $[0,1]^{\mathbb{N}}$ is endowed with the product topology (i.e., the topology corresponding to pointwise convergence). In general, it can be shown that $h(S)$ is a measurable subset of the compact space $[0,1]^{\mathbb{N}}$ (endowed with the product σ-algebra that is generated by the product topology). If S is compact then $h(S)$ is compact as well. In general, we can identify

$$S \;\cong\; h(S) \;\subseteq\; \hat{S} \;\subseteq\; [0,1]^{\mathbb{N}}$$

where $\hat{S} := \overline{h(S)}$ is compact since it is a closed subset of the compact space $[0,1]^{\mathbb{N}}$. Thus $\hat{S}$ can be viewed as a compactification of S.

On compact spaces, any sequence of probability measures has a weakly convergent subsequence.

Theorem 1.13. If S is compact then $\mathcal{P}(S)$ is compact w.r.t. weak convergence.

Proof (Sketch). Suppose that S is compact. Then it can be shown based on the remark above that $C(S)$ is separable w.r.t. uniform convergence. Thus there exists a sequence $g_n \in C(S)$ $(n \in \mathbb{N})$ such that $\|g_n\|_{\sup} \leq 1$ for any n, and the linear span of the functions g_n is dense in $C(S)$. Now consider an arbitrary sequence $(\mu_k)_{k\in\mathbb{N}}$ in $\mathcal{P}(S)$. We will show that (μ_k) has a convergent subsequence. Note first that $(\mu_k(g_n))_{k\in\mathbb{N}}$ is a bounded sequence of real numbers for any n. By a diagonal argument, we can extract a subsequence $(\mu_{k_l})_{l\in\mathbb{N}}$ of $(\mu_k)_{k\in\mathbb{N}}$ such that $\mu_{k_l}(g_n)$ converges as $l \to \infty$ for every $n \in \mathbb{N}$. Since the span of the functions g_n is dense in $C(S)$, this implies that

$$\Lambda(f) \;:=\; \lim_{l\to\infty} \mu_{k_l}(f) \tag{1.15}$$

exists for any $f \in C(S)$. It is easy to verify that Λ is a *positive* (i.e., $\Lambda(f) \geq 0$ whenever $f \geq 0$) linear functional on $C(S)$ with $\Lambda(1) = 1$. Moreover, if $(f_n)_{n\in\mathbb{N}}$ is a decreasing sequence in $C(S)$ such that $f_n \searrow 0$ pointwise, then $f_n \to 0$ uniformly by compactness of S, and hence $\Lambda(f_n) \to 0$. Therefore, there exists a probability measure μ on S such that

$$\Lambda(f) = \mu(f) \quad \text{for any } f \in C(S).$$

By (1.15), the sequence (μ_{k_l}) converges weakly to μ. ■

Remark (A metric for weak convergence). Choosing the functions g_n as in the proof above, we see that a sequence $(\mu_k)_{k\in\mathbb{N}}$ of probability measures in $\mathcal{P}(S)$ converges weakly to μ if and only if $\mu_k(g_n) \to \mu(g_n)$ for all $n \in \mathbb{N}$. Thus weak convergence in $\mathcal{P}(S)$ is equivalent to convergence w.r.t. the metric

$$d(\mu, \nu) \;=\; \sum_{n=1}^{\infty} 2^{-n} |\mu(g_n) - \nu(g_n)|.$$

Prokhorov's theorem

We now consider the case where S is a Polish space that is not necessarily compact. By identifying S with the image $h(S)$ under the map h defined by (1.14), we can still view S as a measurable subset of the compact space $\hat{S} \subseteq [0,1]^{\mathbb{N}}$. Hence $\mathcal{P}(S)$ can be viewed as a subset of the compact space $\mathcal{P}(\hat{S})$:

$$\mathcal{P}(S) \;=\; \{\mu \in \mathcal{P}(\hat{S}) : \mu(\hat{S} \setminus S) = 0\} \;\subseteq\; \mathcal{P}(\hat{S}).$$

If μ_k ($k \in \mathbb{N}$) and μ are probability measures on S (that trivially extend to $\hat{S}$) then by the Portemanteau theorem, and since uniformly continuous functions on S extend uniquely to continuous functions on the closure $\hat{S}$,

$$\begin{aligned} & \mu_k \to \mu \text{ weakly in } \mathcal{P}(S) \\ \Leftrightarrow\quad & \mu_k(f) \to \mu(f) \text{ for any uniformly continuous } f \in C_b(S) \\ \Leftrightarrow\quad & \mu_k(f) \to \mu(f) \text{ for any } f \in C(\hat{S}) \\ \Leftrightarrow\quad & \mu_k \to \mu \text{ weakly in } \mathcal{P}(\hat{S}). \end{aligned} \tag{1.16}$$

Thus $\mathcal{P}(S)$ inherits the weak topology from $\mathcal{P}(\hat{S})$. The problem is, however, that since S is not necessarily a closed subset of $\hat{S}$, it can happen that a sequence (μ_k) in $\mathcal{P}(S)$ converges to a probability measure μ on $\hat{S}$ with $\mu(S) < 1$. To exclude this possibility, the following tightness condition is required:

Definition 1.14 (Tightness of collections of probability measures). Let $\mathcal{R} \subseteq \mathcal{P}(S)$ be a set consisting of probability measures on S. Then $\mathcal{R}$ is called **tight** iff for any $\varepsilon > 0$, there exists a compact set $K \subseteq S$ such that

$$\sup_{\mu \in \mathbb{R}} \mu(S \setminus K) \;<\; \varepsilon.$$

Thus tightness means that the measures in the set $\mathbb{R}$ are concentrated uniformly on a compact set up to an arbitrary small positive amount of mass. On a Polish space, a set $\mathcal{R} = \{\mu\}$ consisting of a single probability measure is always tight, cf. e.g. Billingsley [3] or Ethier and Kurtz [19, Ch. 3, Lemma 2.1].

A set $\mathcal{R} \subseteq \mathcal{P}(S)$ is called **relatively compact** iff every sequence in $\mathcal{R}$ has a subsequence that converges weakly to a limit in $\mathcal{P}(S)$.

Theorem 1.15 (Prokhorov). Suppose that S is Polish, and let $\mathcal{R} \subseteq \mathcal{P}(S)$. Then

$$\mathcal{R} \text{ is relatively compact} \quad\Leftrightarrow\quad \mathcal{R} \text{ is tight}\,.$$

In particular, every tight sequence in $\mathcal{P}(S)$ has a weakly convergent subsequence.

We only prove the implication "$\Leftarrow$" that will be the more important one for our purposes. This implication holds in arbitrary separable metric spaces. For the proof of the converse implication cf. e.g. Billingsley [3].

Proof (of "$\Leftarrow$"). Let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. We have to show that (μ_k) has a weakly convergent subsequence in $\mathcal{P}(S)$. By Theorem 1.13, $\mathcal{P}(\hat{S})$ is compact. Thus there is a subsequence (μ_{k_l}) that converges weakly in $\mathcal{P}(\hat{S})$ to a probability measure μ on $\hat{S}$. We claim that by tightness, $\mu(S) = 1$ and $\mu_{k_l} \to \mu$ weakly in $\mathcal{P}(S)$. Let $\varepsilon > 0$ be given. Then there exists a compact subset K of S such that $\mu_{k_l}(K) \geq 1 - \varepsilon$ for any

l. Since K is compact, it is also a compact and (hence) closed subset of $\hat{S}$. Therefore, by the Portemanteau Theorem,

$$\mu(K) \;\geq\; \limsup_{l\to\infty} \mu_{k_l}(K) \;\geq\; 1-\varepsilon, \qquad \text{and thus}$$

$$\mu(\hat{S}\setminus S) \;\leq\; \mu(\hat{S}\setminus K) \;\leq\; \varepsilon.$$

Letting ε tend to 0, we see that $\mu(\hat{S}\setminus S)=0$. Hence $\mu\in\mathcal{P}(S)$. Moreover, by (1.16), $\mu_{k_l}\to\mu$ weakly in $\mathcal{P}(S)$. ■

Existence of invariant probability measures

We now apply Prokhorov's Theorem to derive sufficient conditions for the existence of an invariant probability measure for a given transition kernel $\pi(x,dy)$ on $(S,\mathcal{B})$.

Definition 1.16 (Feller property). The transition kernel π is called **Feller** iff for any $f\in C_b(S)$, πf is a continuous function.

A kernel π is Feller if and only if $x\mapsto\pi(x,\cdot)$ is a continuous map from S to $\mathcal{P}(S)$ w.r.t. the weak topology on $\mathcal{P}(S)$. Indeed, by definition, π is Feller if and only if for any $f\in C_b(S)$,

$$x_n\to x \qquad\Rightarrow\qquad (\pi f)(x_n)\to(\pi f)(x)\,.$$

For an intuitive interpretation of the following theorem note that if (X_n,P_ν) is a Markov chain with transition kernel π and initial distribution ν then

$$(\nu\overline{p}_n)(K) \;=\; E_\nu\left[\frac{1}{n}\sum_{i=0}^{n-1}1_K(X_i)\right]$$

is the *average proportion of time spent by the chain in the set K during the first n steps*.

Theorem 1.17 (Krylov-Bogoliubov, Foguel). Suppose that π is a Feller transition kernel on a Polish space S, and let $\overline{p}_n := \frac{1}{n}\sum_{i=0}^{n-1}\pi^i$. Then there exists an invariant probability measure μ of π if one of the following conditions is satisfied for some probability measure ν on S:

(i) For every $\varepsilon>0$, there exists a compact set $K\subseteq S$ such that

$$\liminf_{n\to\infty}(\nu\overline{p}_n)(K)\;\geq\;1-\varepsilon, \qquad\qquad \text{or}$$

(ii) S is locally compact, and there exists a compact set $K\subseteq S$ such that

$$\limsup_{n\to\infty}(\nu\overline{p}_n)(K)\;>\;0.$$

Clearly, the second condition is weaker than the first one in several respects. However, it requires the additional assumption that S is **locally compact**, i.e., every point $x\in S$ has a compact neighbourhood. This assumption is satisfied for example for $S=\mathbb{R}^d$, but not for infinite dimensional Hilbert spaces. Polish spaces that are locally compact are also σ**-compact**, i.e., they are the union of countably many compact subsets. This will be crucial in the proof below.

Proof (of Theorem 1.17). (i) On a Polish space, every finite collection of probability measures is tight [3, 19]. Therefore, Condition (i) implies that the sequence $\nu_n := \nu\overline{p}_n$ is tight. Hence by Prokhorov's Theorem, there exists a subsequence (ν_{n_k}) and a probability measure μ on S such that $\nu_{n_k} \to \mu$ weakly. We claim that $\mu\pi = \mu$. Indeed, for $f \in C_b(S)$ we have $\pi f \in C_b(S)$ by the Feller property. Therefore, for any $f \in C_b(S)$,

$$(\mu\pi)(f) = \mu(\pi f) = \lim_{k\to\infty} \nu_{n_k}(\pi f) = \lim_{k\to\infty} (\nu_{n_k}\pi)(f) = \lim_{k\to\infty} \nu_{n_k}(f) = \mu(f),$$

where the second last equality holds since

$$\nu_{n_k}\pi \;=\; \frac{1}{n_k}\sum_{i=0}^{n_k-1} \nu\pi^{i+1} \;=\; \nu_{n_k} - \frac{1}{n_k}\nu + \frac{1}{n_k}\nu\pi^{n_k}.$$

(ii) Now suppose that Condition (ii) holds. We may assume that S is a Borel subset of a compact space $\hat{S}$. Since $\mathcal{P}(\hat{S})$ is compact and (ii) holds, there exist a constant $\varepsilon > 0$, a compact set $K \subseteq S$, a subsequence (ν_{n_k}) of (ν_n), and a probability measure $\hat{\mu}$ on $\hat{S}$ such that

$$\nu_{n_k}(K) \geq \varepsilon \quad \text{for any } k \in \mathbb{N}, \qquad \text{and} \qquad \nu_{n_k} \to \hat{\mu} \quad \text{weakly in } \hat{S}.$$

Note that weak convergence of the probability measures on $\hat{S}$ does not imply weak convergence of the restricted measures on S. However, (ν_{n_k}) converges *vaguely* to the restriction of the probability measure $\hat{\mu}$ to S, i.e., $\nu_{n_k}(f) \to \hat{\mu}(f)$ holds for any function $f \in C(S)$ with compact support $K \subseteq S$, because these functions can be extended trivially to continuous functions on $\hat{S}$. Therefore, we can conclude that for any non-negative compactly supported $f \in C(S)$,

$$\hat{\mu}(f) = \lim_{k\to\infty} \nu_{n_k}(f) = \lim_{k\to\infty} (\nu_{n_k}\pi)(f) = \lim_{k\to\infty} \nu_{n_k}(\pi f) \geq \hat{\mu}(\pi f). \tag{1.17}$$

Here we have used in the last step that since S is a locally compact Polish space, every continuous function $g : S \to [0,\infty)$ can be represented as the limit of an increasing sequence of compactly supported, non-negative functions $g_i \in C(S)$. Therefore,

$$\liminf_{k\to\infty} \nu_{n_k}(g) \geq \liminf_{k\to\infty} \nu_{n_k}(g_i) = \hat{\mu}(g_i) \qquad \text{for all } i \in \mathbb{N},$$

and thus

$$\liminf_{k\to\infty} \nu_{n_k}(g) \geq \hat{\mu}(g).$$

Since $\hat{\mu}(S) \geq \hat{\mu}(K) \geq \limsup \nu_{n_k}(K) \geq \varepsilon$, the normalized measure

$$\mu(B) \;:=\; \hat{\mu}(B \cap S)/\hat{\mu}(S) \;=\; \hat{\mu}(B|S), \qquad B \in \mathcal{B}(S),$$

exists, and by (1.17), $\mu \geq \mu\pi$. But μ and $\mu\pi$ are both probability measures, and thus $\mu = \mu\pi$. ∎

In practice, the assumptions in Theorem 1.17 can be verified via appropriate Lyapunov functions:

Corollary 1.18 (Lyapunov condition for the existence of an invariant probability measure). Suppose that π is a Feller transition kernel and S is locally compact. Then an invariant probability measure for π exists if the following Lyapunov condition is satisfied:

(LI) There exists a function $V \in \mathcal{F}_+(S)$, a compact set $K \subseteq S$, and constants $c, \varepsilon \in (0,\infty)$ such that

$$\mathcal{L}V \leq c1_K - \varepsilon.$$

Proof. By (LI),

$$c1_K \;\geq\; \varepsilon + \mathcal{L}V \;=\; \varepsilon + \pi V - V.$$

By integrating the inequality w.r.t. the probability measures $\overline{p}_n(x,\cdot)$, we obtain

$$\begin{aligned} c\,\overline{p}_n(\cdot,K) \;&=\; c\,\overline{p}_n 1_K \;\geq\; \varepsilon + \frac{1}{n}\sum_{i=0}^{n-1}(\pi^{i+1}V - \pi^i V) \\ &=\; \varepsilon + \frac{1}{n}\pi^n V - \frac{1}{n}V \;\geq\; \varepsilon - \frac{1}{n}V \end{aligned}$$

for any $n \in \mathbb{N}$. Therefore, for any $x \in S$,

$$\liminf_{n\to\infty} \overline{p}_n(x,K) \;\geq\; \varepsilon\,.$$

The assertion now follows by Theorem 1.17. ■

Example. 1) *Countable state space.* If S is countable and π is irreducible then an invariant probability measure exists if and only if the Markov chain is positive recurrent. On the other hand, by Corollary 1.9, positive recurrence is equivalent to (LI). Hence for irreducible Markov chains on countable state spaces, Condition (LI) is both necessary and sufficient for the existence of a stationary distribution.

2) $S = \mathbb{R}^d$. On $\mathbb{R}^d$, Condition (LI) is satisfied in particular if $\mathcal{L}V$ is continuous and

$$\limsup_{|x|\to\infty} \mathcal{L}V(x) < 0.$$

1.4. Lyapunov functions and stability in continuous time

In this section we explain how Lyapunov function methods similar to those considered above can be applied to Markov processes in continuous time. An excellent reference is the book by Khasminskii [29] that focuses on diffusion processes in $\mathbb{R}^d$. Most results in [29] easily carry over to more general Markov processes in continuous time.

We assume that we are given a right continuous stochastic process $((X_t)_{t\in\mathbb{R}_+}, P)$ with Polish state space S, constant initial value $X_0 = x_0 \in S$, and life time ζ. Let $\hat{\mathcal{A}}$ be a linear subspace of the space $C^{1,0}([0,\infty)\times S)$ consisting of continuous functions $(t,x) \mapsto f(t,x)$ that are continuously differentiable in the first variable. Moreover, let $\hat{\mathcal{L}} : \hat{\mathcal{A}} \to \mathcal{F}([0,\infty)\times S)$ be a linear operator of the form

$$(\hat{\mathcal{L}}f)(t,x) \;=\; \left(\frac{\partial f}{\partial t} + \mathcal{L}_t f\right)(t,x)$$

where $\mathcal{L}_t$ is a linear operator acting only on the x-variable. For $f \in \hat{\mathcal{A}}$ and $t < \zeta$ we consider

$$M_t^f \;=\; f(t,X_t) - \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{L}_s f\right)(s,X_s)\,ds$$

where it is implicitly assumed that the integral exists almost surely and defines a measurable function. We assume that the process (X_t) is adapted to a filtration $(\mathcal{F}_t)$, and that (X_t, P) solves the local martingale problem for $(\hat{\mathcal{L}}, \hat{\mathcal{A}})$ up to the life-time ζ in the following sense:

Assumption (A). There exists an increasing sequence $(B_k)_{k\in\mathbb{N}}$ of open sets in S such that the following conditions are satisfied:

(i) $S = \bigcup B_k$.

(ii) The exit times $T_k := \inf\{t \geq 0 : X_t \notin B_k\}$ satisfy $T_k < \zeta$ on $\{\zeta < \infty\}$ for any $k \in \mathbb{N}$, and $\zeta = \sup T_k$.

(iii) For any $k \in \mathbb{N}$ and $f \in \hat{\mathcal{A}}$, the stopped process $\left(M^f_{t\wedge T_k}\right)_{t\geq 0}$ is a square integrable $(\mathcal{F}_t)$ martingale.

Example. 1) Itô diffusions. A stochastic differential equation $dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dB_t$ on $\mathbb{R}^d$ with locally Lipschitz continuous coefficients has a unique strong solution on a maximal time interval $[0,\zeta)$. This solution is a Markov process that satisfies Assumption (A) with $B_k = B(0,k)$, $\hat{\mathcal{A}} = C^{1,2}([0,\infty)\times\mathbb{R}^d)$, and

$$(\mathcal{L}_t f)(t,x) \;=\; b(t,x)\cdot\nabla f(t,x) \;+\; \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(t,x)\frac{\partial^2 f}{\partial x_i \partial x_j}(t,x),$$

where $a(t,x) = \sigma(t,x)\sigma(t,x)^T$, see e.g. [16, 17].

2) Minimal jump processes. In Section 5.1, we will construct a minimal jump process for given jump intensities. This process satisfies the assumption if (B_k) is an increasing sequence of open sets exhausting the state space such that the jump intensities $\lambda_t(x)$ are uniformly bounded for $(t,x) \in \mathbb{R}_+ \times B_k$, and $\hat{\mathcal{A}}$ consists of all functions $f \in C^{1,0}([0,\infty)\times S)$ such that f and $\partial f/\partial t$ are bounded on $[0,t]\times B_k$ for any $t \geq 0$ and $k \in \mathbb{N}$.

Upper bounds for expectations

Let $D \subseteq S$ be an open subset. Since the process is right-continuous, the first exit time

$$T \;=\; \inf\{t \geq 0 : X_t \in S_\Delta \setminus D\}$$

is an $(\mathcal{F}_t)$ stopping time. Similarly to the discrete time case, we can use superharmonic functions to bound expected values of the form

$$u(x_0) \;=\; E\left[e^{-\int_0^T w(s,X_s)\,ds} f(T,X_T); T<\zeta\right] + E\left[\int_0^T e^{-\int_0^s w(r,X_r)\,dr} c(s,X_s)\,ds\right] \tag{1.18}$$

for given non-negative measurable functions $f : \mathbb{R}_+ \times (S\setminus D) \to \mathbb{R}_+$ and $c, w : \mathbb{R}_+ \times D \to \mathbb{R}_+$.

Theorem 1.19 (Upper bounds via time-dependent superharmonic functions). Suppose that Assumption (A) holds, and let v be a non-negative function in $\hat{\mathcal{A}}$ satisfying

$$\begin{aligned} \frac{\partial v}{\partial t} + \mathcal{L}_t v \;&\leq\; w\,v - c \qquad && \text{on } \mathbb{R}_+\times D, \\ v \;&\geq\; f && \text{on } \mathbb{R}_+ \times (S\setminus D). \end{aligned} \tag{1.19}$$

Then $u(x_0) \;\leq\; v(0,x_0)$.

Proof. We first assume $w \equiv 0$. Then by (1.19),

$$v(t,X_t) \;\leq\; M^v_t - \int_0^t c(s,X_s)\,ds \qquad \text{for } t<T.$$

The process $\left(M^v_{t\wedge T_k}\right)_{t\geq 0}$ is a martingale for any $k \in \mathbb{N}$. Since $v \geq 0$, optional stopping implies

$$E\left[v\left(t\wedge T\wedge T_k, X_{t\wedge T\wedge T_k}\right) + \int_0^{t\wedge T\wedge T_k} c(s,X_s)\,ds\right] \;\leq\; v(0,x_0) \qquad \text{for any } k\in\mathbb{N} \text{ and } t \in \mathbb{R}_+.$$

As $k \to \infty$, $T_k \to \zeta$ almost surely. Recalling that v and c are non-negative functions, and letting both k and t tend to infinity, we obtain

$$E\left[v\left(T, X_T\right); T < \zeta\right] + E\left[\int_0^T c(s, X_s)\, ds\right] \leq v(0, x_0).$$

This proves the assertion, since for $T < \zeta$, $X_T \in S \setminus D$, and thus $v(T, X_T) \geq f(T, X_T)$.

In the general case let $A_t = \int_0^t w(s, X_s)\, ds$. Notice that for every ω, the function $t \mapsto A_t(\omega)$ is continuous and non-decreasing, and hence of finite variation. Therefore, by Itô's product rule (which in this case is just the integration by parts identity from Stieltjes calculus, see e.g. [17]),

$$\begin{aligned} e^{-A_t} v(t, X_t) &= v(0, x_0) + \int_0^t e^{-A_s}\, dM_s^v + \int_0^t e^{-A_s} \left(\frac{\partial v}{\partial s} + \mathcal{L}_s v - wv\right)(s, X_s)\, ds \\ &\leq v(0, x_0) + \int_0^t e^{-A_s}\, dM_s^v - \int_0^t e^{-A_s}\, c(s, X_s)\, ds. \end{aligned}$$

Now the argument can be completed as in the case $w \equiv 0$ because the Itô integral $\int_0^{t \wedge T_k} e^{-A_s}\, dM_s$, $t \geq 0$, is a martingale for any $k \in \mathbb{N}$. ■

Non-explosion criteria

A first important application of Lyapunov functions in continuous time are conditions for non-explosiveness of a Markov process:

Theorem 1.20 (Khasminskii). Suppose that Assumption (A) is satisfied, and assume that there exists a non-negative function $V \in \hat{\mathcal{A}}$ such that for any $t \geq 0$,

(i) $\inf_{s \in [0,t]} \inf_{x \in B_k^c} V(s, x) \to \infty$ as $k \to \infty$, and

(ii) $\frac{\partial V}{\partial t} + \mathcal{L}_t V \leq 0$.

Then $P[\zeta = \infty] = 1$.

Proof. Since $V \geq 0$, $V(t, X_t) = M_t^V + \int_0^t \left(\frac{\partial V}{\partial s} + \mathcal{L}_s V\right)(s, X_s)\, ds$, and Condition (ii) holds, optional stopping implies that for any $t \geq 0$ and $k \in \mathbb{N}$,

$$V(0, x_0) \geq E[V(t \wedge T_k, X_{t \wedge T_k})] \geq P[T_k \leq t] \cdot \inf_{s \in [0,t]} \inf_{x \in B_k^c} V(s, x).$$

Therefore, by (i),

$$P[T_k \leq t] \to 0 \quad \text{as } k \to \infty$$

for any $t \geq 0$, and hence $P[\zeta < \infty] = \lim_{t \to \infty} P[\zeta \leq t] = 0$. ■

In applications, the following simple criterion is used frequently:

Corollary 1.21. Suppose that Assumption (A) is satisfied, and there exist a non-negative function $U \in C(S)$ and a constant $\alpha \in \mathbb{R}_+$ such that $V(t, x) = \exp(-\alpha t) U(x)$ is contained in $\hat{A}$, and

(i) $\inf_{x \in B_k^c} U(x) \to \infty$ as $k \to \infty$, and

(ii) $\mathcal{L}_t U \leq \alpha U$.

Then $P[\zeta = \infty] = 1$.

Proof. Theorem 1.20 can be applied with the function V. ■

Hitting times and positive recurrence

From now on we assume that the process is non-explosive. We now apply Lyapunov functions to prove upper bounds for moments of hitting times. Let

$$T_A \;=\; \inf\{t \geq 0 : X_t \in A\},$$

where A is a closed subset of S. By right-continuity of (X_t), T_A is a stopping time.

Theorem 1.22 (Lyapunov bound for hitting times). Suppose that Assumption (A) holds, and the process (X_t, P) is non-explosive. Furthermore, assume that there exist a non-negative function $V \in \hat{\mathscr{A}}$ and a measurable function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that

(i) $\beta(t) := \int_0^t \alpha(s)\,ds \to \infty$ as $t \to \infty$, and

(ii) $\left(\frac{\partial V}{\partial t} + \mathcal{L}_t V\right)(t,x) \leq -\alpha(t)$ for any $t \geq 0$ and $x \in S \setminus A$,

Then $P[T_A < \infty] = 1$, and

$$E[\beta(T_A)] \;\leq\; V(0,x_0). \tag{1.20}$$

Proof. By applying Theorem 1.19 with $D = S \setminus A$, $c(s,x) = \alpha(s)$, $w \equiv 0$ and $f \equiv 0$, we obtain

$$E\left[\int_0^{T_A} \alpha(s)\,ds\right] \;\leq\; V(0,x_0).$$

Hence by (i), T_A is almost surely finite and (1.20) holds. ■

Example (Moments of hitting times). If $\alpha(t) = ct^{p-1}$ for constants $c,p \in (0,\infty)$ then $\beta(t) = ct^p/p$. In this case, the inequality (1.20) is a bound for the p-th moment of the hitting time:

$$E[T_A^p] \;\leq\; pV(0,x_0)/c.$$

Similarly, if $\alpha(t) = ce^{\varepsilon t}$ for constants $c,\varepsilon \in (0,\infty)$ then $\beta(t) = (e^{\varepsilon t} - 1)/\varepsilon$. In this case, (1.20) is an exponential moment bound:

$$E[\exp(\varepsilon T_A)] \;\leq\; 1 + \varepsilon V(0,x_0).$$

For Itô diffusions on $\mathbb{R}^d$, we can apply the theorem with the Lyapunov function $V(t,x) = |x|^2$. In this case, $(\mathcal{L}_t V)(t,x) = \operatorname{tr} a(t,x) + 2x \cdot b(t,x)$. If this expression is bounded from above by a negative constant for large $|x|$ then sufficiently large balls are positive recurrent. For example, the Itô diffusion solving the SDE

$$X_t \;=\; x_0 \;+\; B_t \;+\; \int_0^t b(X_s)\,ds, \qquad (B_t) \sim \mathrm{BM}(\mathbb{R}^d),$$

is positive recurrent if there exist constants $R,\varepsilon \in (0,\infty)$ such that for $|x| \geq R$,

$$b_r(x) \;=\; \frac{x}{|x|} \cdot b(x) \;\leq\; -\frac{d+\varepsilon}{2|x|}.$$

Occupation times and existence of stationary distributions

Similarly to the discrete time case, Lyapunov conditions can also be used in continuous time to show the existence of stationary distributions. Let

$$A_t(B) = \frac{1}{t}\int_0^t 1_B(X_s)\,ds$$

denote the relative amount of time spent by the process in the set B during the time interval $[0,t]$. The proofs of the following results are similar to the discrete time case. Carrying out the details is left as an exercise, see Section 1.5.

Lemma 1.23 (Lyapunov bound for occupation times). *Suppose that Assumption (A) holds, the process is almost surely non-explosive, and there exist constants $\varepsilon, c \in \mathbb{R}_+$ and a non-negative function $V \in \hat{\mathcal{A}}$ such that*

$$\frac{\partial V}{\partial t} + \mathcal{L}_t V \;\le\; -\varepsilon + c 1_B \qquad \text{on } \mathbb{R}_+ \times S.$$

Then

$$E[A_t(B)] \;\ge\; \frac{\varepsilon}{c} - \frac{V(0,x_0)}{ct}.$$

Now assume that (X_t, P) is a **time-homogeneous** Markov process with transition semigroup $(p_t)_{t\ge 0}$, and, correspondingly, $\mathcal{L}_t$ does not depend on t. Then by Fubini's Theorem,

$$E[A_t(B)] \;=\; \frac{1}{t}\int_0^t p_s(x_0,B)\,ds \;=:\; \overline{p}_t(x_0,B).$$

Theorem 1.24 (Krylov-Bogoliubov, Foguel). Suppose that p_t is Feller for every $t \ge 0$. Then there exists an invariant probability measure μ of $(p_t)_{t\ge 0}$ if one of the following conditions is satisfied for some probability measure ν on S:

(i) For every $\varepsilon > 0$, there exists a compact set $K \subseteq S$ such that

$$\liminf_{t\to\infty} (\nu\overline{p}_t)(K) \;\ge\; 1-\varepsilon, \qquad\qquad \text{or}$$

(ii) S is σ-compact, and there exists a compact set $K \subseteq S$ such that

$$\limsup_{t\to\infty} (\nu\overline{p}_t)(K) \;>\; 0.$$

As a consequence, we obtain:

Corollary 1.25 (Lyapunov condition for the existence of an invariant probability measure). Suppose that Assumption (A) holds, and the process is almost surely non-explosive. Moreover, assume that p_t is Feller for every $t \ge 0$, S is locally compact, and there exist constants $\varepsilon, c \in \mathbb{R}_+$, a compact set $K \subseteq S$, and a continuous function $U : S \to [0,\infty)$ such that the function $V(t,x) = U(x)$ is in $\hat{\mathcal{A}}$, and

$$\mathcal{L}U \;\le\; -\varepsilon + c 1_K.$$

Then there exists a stationary distribution μ of $(p_t)_{t\ge 0}$.

Proof. By Lemma 1.23, the assumptions imply $\liminf_{t\to\infty} \overline{p}_t(x_0,K) > 0$. ■

1.5. Stability of diffusion processes and Euler approximations

The goal of this problem section is to apply Lyapunov functions to study diffusion processes on $\mathbb{R}^d$ and Markov chains corresponding to Euler approximations of the respective stochastic differential equations.

Let $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be locally Lipschitz continuous functions. We consider a diffusion process (X_t, P_x) with possibly finite life-time ζ solving a stochastic differential equation

$$dX_t \;=\; \sigma(X_t)\,dB_t \;+\; b(X_t)\,dt \qquad \text{for } t<\zeta, \qquad X_0 = x, \tag{1.21}$$

where (B_t) is a d-dimensional Brownian motion. Let $a(x) \;=\; \sigma(x)\sigma(x)^T$. For the exercises below, it will only be important to know that (X_t, P_x) solves the local martingale problem for the generator

$$\mathcal{L}f \;=\; \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i\,\partial x_j} \;+\; \sum_{i=1}^{d} b_i(x)\frac{\partial f}{\partial x_i}, \tag{1.22}$$

in the sense that for any $x\in\mathbb{R}^2$ and $f\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^n)$,

$$M_t^f \;=\; f(t,X_t) \;-\; \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{L}f\right)(s,X_s)\,ds \tag{1.23}$$

is a local martingale w.r.t. P_x. More precisely, let $T_k = \inf\{t\geq 0 : |X_t|\geq k\}$. Then $\zeta = \sup T_k$, and for any $k\in\mathbb{N}$, the stopped process $(M_{t\wedge T_k})_{t\geq 0}$ is a martingale under P_x.

For a fixed time step $h>0$, the **Euler-Maruyama approximation** of the diffusion process above is the time-homogeneous Markov chain (X_n^h, P_x) with transition step

$$x \;\mapsto\; x \;+\; \sqrt{h}\,\sigma(x)\,Z \;+\; h\,b(x), \qquad Z\sim N(0,I_d).$$

We denote the corresponding transition kernel and generator by π_h and $\mathcal{L}_h$, respectively.

Problems

Exercise (Explosions). a) Prove that the diffusion process is non-explosive if

$$\limsup_{|x|\to\infty} \frac{\operatorname{tr} a(x)/2 + x\cdot b(x)}{|x|^2} < \infty$$

b) Implement the Euler scheme on a computer, e.g. for $d=1$. Do some experiments. Can you observe a different behavior in cases where the condition above is satisfied or violated, respectively?

c) As discrete time Markov chains, the Euler approximations always have infinite life-time. Which properties can you prove for the Euler approximations if the condition in a) is satisfied?

Exercise (Stationary distributions I). Suppose that $\zeta = \infty$ almost surely, and that there exist $V\in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^d)$, $\varepsilon, c\in\mathbb{R}_+$, and a ball $B\subset\mathbb{R}^d$ such that $V\geq 0$ and

$$\frac{\partial V}{\partial t} + \mathcal{L}V \;\leq\; -\varepsilon + c1_B \qquad \text{on } \mathbb{R}_+\times\mathbb{R}^d.$$

a) Prove that

$$E\left[\frac{1}{t}\int_0^t 1_B(X_s)\,ds\right] \;\geq\; \frac{\varepsilon}{c} - \frac{V(0,x_0)}{ct}.$$

b) Conclude, assuming that (X_t, P_x) is a time-homogeneous Markov process with Feller transition semigroup $(p_t)_{t\geq 0}$, that there exists an invariant probability measure μ. *Hint: Try to carry over the corresponding proof in discrete time to the continuous time case.*

Exercise (Stationary distributions II). Suppose that the conditions in the last exercise hold, and let μ be an invariant probability measure.

a) Show that $\int \mathcal{L}f \, d\mu = 0$ for any $f \in C_0^\infty(\mathbb{R}^d)$.

b) Use this to compute μ explicitly in the case $d = 1$. Here, assume that b and σ are twice continuously differentiable, and $\sigma(x) > 0$ for all x. You may also assume without proof that μ has a twice continuously differentiable density $\varrho(x)$ which is strictly positive.

Exercise (Stationary distributions III). a) Show that an invariant probability measure for the diffusion process on $\mathbb{R}^d$ exists if

$$\limsup_{|x|\to\infty} \left(\operatorname{tr} a(x)/2 \; + \; x \cdot b(x)\right) \; < 0. \tag{1.24}$$

b) Give conditions ensuring the existence of an invariant probability measure for the Euler approximations. Why is not enough to assume (1.24) ?

c) Study different cases experimentally using an implementation of the Euler scheme. Can you see the difference between cases where an invariant probability measure exists or does not exist ?

2. Ergodic averages

Suppose that (X_n, P_x) is a canonical time-homogeneous Markov chain with transition kernel π. Recall that the process (X_n, P_μ) with initial distribution μ is **stationary**, i.e.,

$$X_{n:\infty} \sim X_{0:\infty} \quad \text{for any } n \geq 0,$$

if and only if $\mu = \mu\pi$. A probability measure μ with this property is called a **stationary (initial) distribution** or an **invariant probability measure for the transition kernel** π. In this chapter we will prove law of large number type theorems for ergodic averages of the form

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \to \int f d\mu \quad \text{as } n \to \infty,$$

and, more generally,

$$\frac{1}{n}\sum_{i=0}^{n-1} F(X_i, X_{i+1}, \dots) \to \int F dP_\mu \quad \text{as } n \to \infty$$

where μ is a stationary distribution for the transition kernel. At first these limit theorems are derived almost surely or in L^p w.r.t. the law P_μ of the Markov chain in stationarity. Indeed, they turn out to be special cases of more general ergodic theorems for stationary (not necessarily Markovian) stochastic processes. After the derivation of the basic results we will consider extensions to continuous time. Moreover, we will study the fluctuations of ergodic averages around their limit. The validity of ergodic theorems for Markov chains that are not started in stationarity is considered in Section 3.2.

As usual, S will denote a Polish space endowed with its Borel σ-algebra $\mathcal{B}$.

2.1. Ergodic theorems

Supplementary references for this section are the probability theory textbooks by Breiman [5], Durrett [13] and Varadhan [53]. We first introduce the more general setup of ergodic theory that includes stationary Markov chains as a special case:

Let $(\Omega, \mathfrak{A}, P)$ be a probability space, and let

$$\Theta : \Omega \to \Omega$$

be a measure-preserving measurable map on $(\Omega, \mathfrak{A}, P)$, i.e.,

$$P \circ \Theta^{-1} = P.$$

The main example is the following: Let

$$\Omega = S^{\mathbb{Z}_+}, \quad X_n(\omega) = \omega_n, \quad \mathfrak{A} = \sigma(X_n : n \in \mathbb{Z}_+),$$

be the canonical model for a stochastic process with state space S. Then the shift transformation $\Theta = X_{1:\infty}$ given by

$$\Theta(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots) \quad \text{for any } \omega \in \Omega$$

is measure-preserving on $(\Omega, \mathfrak{A}, P)$ if and only if (X_n, P) is a stationary process.

Ergodicity

We denote by $\mathcal{J}$ the sub-σ-algebra of $\mathfrak{A}$ consisting of all Θ-invariant events, i.e.,

$$\mathcal{J} := \left\{A \in \mathfrak{A} : \Theta^{-1}(A) = A\right\}.$$

It is easy to verify that $\mathcal{J}$ is indeed a σ-algebra, and that a function $F : \Omega \to \mathbb{R}$ is $\mathcal{J}$-measurable if and only if

$$F = F \circ \Theta.$$

Definition 2.1 (Ergodic probability measure). The probability measure P on $(\Omega, \mathfrak{A})$ is called **ergodic** (w.r.t. Θ) if and only if any event $A \in \mathcal{J}$ has probability zero or one.

Exercise (Characterization of ergodicity). 1) Show that P is not ergodic if and only if there exists a non-trivial decomposition $\Omega = A \cup A^c$ of Ω into disjoint sets A and A^c with $P[A] > 0$ and $P[A^c] > 0$ such that

$$\Theta(A) \subseteq A \quad \text{and} \quad \Theta(A^c) \subseteq A^c.$$

2) Prove that P is ergodic if and only if any measurable function $F : \Omega \to \mathbb{R}$ satisfying $F = F \circ \Theta$ is P-almost surely constant.

Before considering general stationary Markov chains we look at two elementary examples:

Example (Deterministic rotations of the unit circle).
Let $\Omega = \mathbb{R}/\mathbb{Z}$ or, equivalently, $\Omega = [0,1]/\sim$ where "$\sim$" is the equivalence relation that identifies the boundary points 0 and 1. We endow Ω with the Borel σ-algebra $\mathfrak{A} = \mathcal{B}(\Omega)$ and the uniform distribution (Lebesgue measure) $P = \mathrm{Unif}(\Omega)$. Then for any fixed $a \in \mathbb{R}$, the rotation

$$\Theta(\omega) = \omega + a \quad (\text{modulo } 1)$$

is a measure preserving transformation of $(\Omega, \mathfrak{A}, P)$. Moreover, P is ergodic w.r.t. Θ if and only if a is irrational:

$\underline{a \in \mathbb{Q}}$: If $a = p/q$ with $p, q \in \mathbb{Z}$ relatively prime then

$$\Theta^n(\omega) \in \left\{\omega + \frac{k}{q} : k = 0, 1, \ldots, q-1\right\} \quad \text{for any } n \in \mathbb{Z}.$$

This shows that for instance the union

$$A = \bigcup_{n \in \mathbb{Z}} \Theta^n\left(\left[0, \frac{1}{2q}\right)\right)$$

is Θ-invariant with $P[A] \notin \{0,1\}$, i.e., P is **not ergodic**.

$\underline{a \notin \mathbb{Q}}$: Suppose a is irrational and F is a bounded measurable function on Ω with $F = F \circ \Theta$. Then F has a Fourier representation

$$F(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \omega} \quad \text{for } P\text{-almost every } \omega \in \Omega,$$

and Θ invariance of F implies

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n(\omega + a)} = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n \omega} \quad \text{for } P\text{-almost every } \omega \in \Omega,$$

i.e., $c_n e^{2\pi i n a} = c_n$ for any $n \in \mathbb{Z}$. Since a is irrational this implies that all Fourier coefficients c_n except c_0 vanish, i.e., F is P-almost surely a constant function. Thus P is **ergodic** in this case.

Example (IID Sequences). Let μ be a probability measure on $(S, \mathcal{B})$. The canonical process $X_n(\omega) = \omega_n$ is an i.i.d. sequence w.r.t. the product measure $P = \bigotimes_{n=0}^{\infty} \mu$ on $\Omega = S^{\mathbb{Z}_+}$. In particular, (X_n, P) is a stationary process, i.e., the shift $\Theta(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots)$ is measure-preserving. To see that P is ergodic w.r.t. Θ we consider an arbitrary event $A \in \mathcal{J}$. Then

$$A = \Theta^{-n}(A) = \{(X_n, X_{n+1}, \dots) \in A\} \qquad \text{for any } n \geq 0.$$

This shows that A is a tail event, and hence $P[A] \in \{0, 1\}$ by Kolmogorov's zero-one law.

Ergodicity of stationary Markov chains

Now suppose that (X_n, P_μ) is a general stationary Markov chain with initial distribution μ and transition kernel π satisfying $\mu = \mu\pi$. Note that by stationarity, the map $f \mapsto \pi f$ is a contraction on $\mathcal{L}^2(\mu)$. Indeed, by the Cauchy-Schwarz inequality,

$$\int (\pi f)^2 d\mu \leq \int \pi f^2 d\mu \leq \int f^2 d(\mu\pi) = \int f^2 d\mu \quad \forall f \in \mathcal{L}^2(\mu).$$

In particular,

$$\mathcal{L} f = \pi f - f$$

is an element in $\mathcal{L}^2(\mu)$ for any $f \in \mathcal{L}^2(\mu)$.

Theorem 2.2 (Characterizations of ergodicity for stationary Markov chains). The following statements are equivalent:

1) The measure P_μ is shift-ergodic.

2) Any function $h \in \mathcal{L}^2(\mu)$ satisfying $\mathcal{L}h = 0$ μ-almost surely is μ-almost surely constant.

3) Any Borel set $B \in \mathcal{B}$ satisfying $\pi 1_B = 1_B$ μ-almost surely has measure $\mu(B) \in \{0, 1\}$.

Proof.

1) $\Rightarrow$ 2). Suppose that P_μ is ergodic and let $h \in \mathcal{L}^2(\mu)$ with $\mathcal{L}h = 0$ μ-a.e. Then the process $M_n = h(X_n)$ is a square-integrable martingale w.r.t. P_μ. Moreover, the martingale is bounded in $L^2(P_\mu)$ since by stationarity,

$$E_\mu[h(X_n)^2] = \int h^2 d\mu \qquad \text{for any } n \in \mathbb{Z}_+.$$

Hence by the L^2 martingale convergence theorem, the limit $M_\infty = \lim_{n \to \infty} M_n$ exists in $L^2(P_\mu)$. We fix a version of M_∞ by defining

$$M_\infty(\omega) = \limsup_{n \to \infty} h(X_n(\omega)) \qquad \text{for every } \omega \in \Omega.$$

Note that M_∞ is a $\mathcal{J}$-measurable random variable, since

$$M_\infty \circ \Theta = \limsup_{n \to \infty} h(X_{n+1}) = \limsup_{n \to \infty} h(X_n) = M_\infty.$$

Therefore, by ergodicity of P_μ, M_∞ is P_μ-almost surely constant. Furthermore, by the martingale property,

$$h(X_0) = M_0 = E_\mu[M_\infty | \mathcal{F}_0^X] \quad P_\mu\text{-a.s.}$$

Hence $h(X_0)$ is P_μ-almost surely constant, and thus h is μ-almost surely constant.

2) $\Rightarrow$ 3). If B is a Borel set with $\pi 1_B = 1_B$ μ-almost surely then the function $h = 1_B$ satisfies $\mathcal{L}h = 0$ μ-almost surely. If 2) holds then h is μ-almost surely constant, i.e., $\mu(B)$ is equal to zero or one.

3) $\Rightarrow$ 1). For proving that 3) implies ergodicity of P_μ let $A \in \mathcal{J}$. Then $1_A = 1_A \circ \Theta$. We will show that this property implies that

$$h(x) := E_x[1_A]$$

satisfies $\pi h = h$, and h is μ-almost surely equal to an indicator function 1_B. Hence by 3), either $h = 0$ or $h = 1$ holds μ-almost surely, and thus $P_\mu[A] = \int h d\mu$ equals zero or one.
The fact that h is harmonic follows from the Markov property and the invariance of A: For any $x \in S$,

$$(\pi h)(x) = E_x\left[E_{X_1}[1_A]\right] = E_x[1_A \circ \Theta] = E_x[1_A] = h(x).$$

To see that h is μ-almost surely an indicator function observe that by the Markov property invariance of A and the martingale convergence theorem,

$$h(X_n) = E_{X_n}[1_A] = E_\mu[1_A \circ \Theta^n | \mathcal{F}_n^X] = E_\mu[1_A | \mathcal{F}_n^X] \to 1_A$$

P_μ-almost surely as $n \to \infty$. Hence

$$\mu \circ h^{-1} = P_\mu \circ (h(X_n))^{-1} \xrightarrow{w} P_\mu \circ 1_A^{-1}.$$

Since the left-hand side does not depend on n,

$$\mu \circ h^{-1} = P_\mu \circ 1_A^{-1},$$

and so h takes μ-almost surely values in $\{0, 1\}$. ■

The third condition in Theorem 2.2 says that every Borel set B satisfying $\pi(x, B) = 1$ for μ-a.e. $x \in B$ and $\pi(x, B) = 0$ for μ-a.e. $x \in B^c$ has μ-measure zero or one. In this sense, it is reminiscent of the definition of irreducibility. However, there is an important difference as the following example shows:

Example (Ergodicity and irreducibility). Consider the constant Markov chain on $S = \{0, 1\}$ with transition probabilities $\pi(0, 0) = \pi(1, 1) = 1$. Obviously, any probability measure on S is a stationary distribution for π. The matrix π is not irreducible, for instance $\pi 1_{\{1\}} = 1_{\{1\}}$. Nevertheless, condition 3) is satisfied and P_μ is ergodic if (and only if) μ is a Dirac measure.

One way to verify ergodicity in practice is the strong Feller property:

Definition 2.3 (Strong Feller property). A probability kernel π on $(S, \mathcal{B})$ is called **strong Feller** iff πf is continuous for any bounded measurable function $f : S \to \mathbb{R}$.

Corollary 2.4 (A sufficient condition for ergodicity). Suppose that the transition kernel π is strong Feller. Then P_μ is stationary and ergodic for any invariant probability measure μ of π that has connected support.

Proof. Consider $B \in \mathcal{B}$ such that $\pi 1_B = 1_B$ holds μ-almost surely. If π is strong Feller then $\pi 1_B$ is a continuous function. Therefore, and since the support of μ is connected, either $\pi 1_B \equiv 0$ or $\pi 1_B \equiv 1$ on $\mathrm{supp}(\mu)$. Hence

$$\mu(B) = \mu(1_B) = \mu(\pi 1_B) \in \{0, 1\}.$$ ■

Exercise (Invariant and almost invariant events). An event $A \in \mathfrak{A}$ is called **almost invariant** iff

$$P_\mu[A \,\Delta\, \Theta^{-1}(A)] = 0.$$

Prove that the following statements are equivalent for $A \in \mathfrak{A}$:

(i) A is almost invariant.

(ii) A is contained in the completion $\mathcal{J}^{P_\mu}$ of the σ-algebra $\mathcal{J}$ w.r.t. the measure P_μ.

(iii) There exist a set $B \in \mathcal{B}$ satisfying $\pi 1_B = 1_B$ μ-almost surely such that

$$P_\mu[A \,\Delta\, \{X_n \in B \text{ eventually}\}] = 0.$$

Birkhoff's ergodic theorem

We return to the general setup where Θ is a measure-preserving transformation on a probability space $(\Omega, \mathfrak{A}, P)$, and $\mathcal{J}$ denotes the σ-algebra of Θ-invariant events in $\mathfrak{A}$.

Theorem 2.5 (Birkhoff). Suppose that $P = P \circ \Theta^{-1}$ and let $p \in [1, \infty)$. Then as $n \to \infty$,

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F|\mathcal{J}] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P) \tag{2.1}$$

for any random variable $F \in L^p(\Omega, \mathfrak{A}, P)$. In particular, if P is ergodic then

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P). \tag{2.2}$$

Example (Law of large numbers for stationary processes). Suppose that (X_n, P) is a stationary stochastic process in the canonical model, i.e., $\Omega = S^{\mathbb{Z}_+}$ and $X_n(\omega) = \omega_n$. Then the shift $\Theta = X_{1:\infty}$ is measure-preserving. By applying Birkhoff's theorem to a function of the form $F(\omega) = f(\omega_0)$, we see that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) = \frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[f(X_0)|\mathcal{J}] \tag{2.3}$$

P-almost surely and in $L^p(\Omega, \mathfrak{A}, P)$ for any $f : S \to \mathbb{R}$ such that $f(X_0) \in \mathcal{L}^p$ and $p \in [1, \infty)$. If ergodicity holds then $E[f(X_0)|\mathcal{J}] = E[f(X_0)]$ P-almost surely, where (2.3) is a law of large numbers. In particular, we recover the classical law of large numbers for i.i.d. sequences. More generally, Birkhoff's ergodic can be applied to arbitrary $\mathcal{L}^p$ functions $F : S^{\mathbb{Z}_+} \to \mathbb{R}$. In this case,

$$\frac{1}{n}\sum_{i=0}^{n-1} F(X_i, X_{i+1}, \dots) = \frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F|\mathcal{J}] \tag{2.4}$$

P-almost surely and in L^p as $n \to \infty$. Even in the classical i.i.d. case where $E[F|\mathcal{J}] = E[F]$ almost surely, this result is an important extension of the law of large numbers.

Before proving Birkhoff's Theorem, we give a **functional analytic interpretation** for the L^p convergence.

Remark (Functional analytic interpretation). If Θ is measure preserving on $(\Omega, \mathfrak{A}, P)$ then the map U defined by

$$UF = F \circ \Theta$$

is a linear isometry on $\mathcal{L}^p(\Omega, \mathfrak{A}, P)$ for any $p \in [1, \infty]$. Indeed, if p is finite then

$$\int |UF|^p \, dP = \int |F \circ \Theta|^p \, dP = \int |F|^p \, dP \qquad \text{for any } F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P).$$

Similarly, it can be verified that U is isometric on $\mathcal{L}^\infty(\Omega, \mathfrak{A}, P)$. For $p = 2$, U induces a unitary transformation on the Hilbert space $L^2(\Omega, \mathfrak{A}, P)$, i.e.,

$$(UF, UG)_{L^2(P)} = \int (F \circ \Theta) \, (G \circ \Theta) dP = (F, G)_{L^2(P)} \qquad \text{for any } F, G \in \mathcal{L}^2(\Omega, \mathfrak{A}, P).$$

The L^p ergodic theorem states that for any $F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P)$,

$$\frac{1}{n} \sum_{i=0}^{n-1} U^i F \to \pi F \qquad \text{in } L^p(\Omega, \mathfrak{A}, P) \text{ as } n \to \infty, \text{ where } \pi F := E[F|\mathcal{J}]. \tag{2.5}$$

In the Hilbert space case $p = 2$, πF is the orthogonal projection of F onto the closed subspace

$$H_0 = L^2(\Omega, \mathcal{J}, P) = \left\{ F \in L^2(\Omega, \mathfrak{A}, P) : UF = F \right\} \tag{2.6}$$

of $L^2(\Omega, \mathfrak{A}, P)$. Note that H_0 is the kernel of the linear operator $U - I$. Since U is unitary, H_0 coincides with the orthogonal complement of the range of $U - I$, i.e.,

$$L^2(\Omega, \mathfrak{A}, P) = H_0 \oplus \overline{(U - I)(L^2)}. \tag{2.7}$$

Indeed, every function $F \in H_0$ is orthogonal to the range of $U - I$, since

$$(UG - G, F)_{L^2} = (UG, F)_{L^2} - (G, F)_{L^2} = (UG, F)_{L^2} - (UG, UF)_{L^2} = (UG, F - UF)_{L^2} = 0$$

for any $G \in L^2(\Omega, \mathfrak{A}, P)$. Conversely, every function $F \in \text{Range}(U - I)^\perp$ is contained in H_0 since

$$\|UF - F\|_{L^2}^2 = (UF, UF)_{L^2} - 2(F, UF)_{L^2} + (F, F)_{L^2} = 2(F, F - UF)_{L^2} = 0.$$

The L^2 convergence in (2.5) therefore reduces to a simple functional analytic statement that will be the starting point for the proof in the general case given below.

Exercise ($\mathbf{L^2}$ ergodic theorem). Prove that (2.5) holds for $p = 2$ and any $F \in \mathcal{L}^2(\Omega, \mathfrak{A}, P)$.

From now on we will use the notation

$$A_n F = \frac{1}{n} \sum_{i=0}^{n-1} F \circ \Theta^i = \frac{1}{n} \sum_{i=0}^{n-1} U^i F$$

for ergodic averages of $\mathcal{L}^p$ random variables. Note that A_n defines a linear operator. Moreover, A_n induces a contraction on $L^p(\Omega, \mathfrak{A}, P)$ for any $p \in [1, \infty]$ and $n \in \mathbb{N}$ since

$$\|A_n F\|_{L^p} \le \frac{1}{n} \sum_{i=0}^{n-1} \|U^i F\|_{L^p} = \|F\|_{L^p} \qquad \text{for any } F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P).$$

Proof (of Theorem 2.5). The proof of the ergodic theorem will be given in several steps. At first we will show in Step 1 below that for a broad class of functions the convergence in (2.1) follows in an elementary way. As in the remark above we denote by

$$H_0 = \{F \in L^2(\Omega, \mathfrak{A}, P) : UF = F\}$$

the kernel of the linear operator $U - I$ on the Hilbert space $L^2(\Omega, \mathfrak{A}, P)$. Moreover, let

$$H_1 = \{UG - G : G \in L^\infty(\Omega, \mathfrak{A}, P)\} = (U - I)(L^\infty),$$

and let $\pi F = E[F|\mathcal{J}]$.

Step 1: We show that for any $F \in H_0 + H_1$,

$$A_n F - \pi F \longrightarrow 0 \quad \text{in } L^\infty(\Omega, \mathfrak{A}, P). \tag{2.8}$$

Indeed, suppose that $F = F_0 + UG - G$ with $F_0 \in H_0$ and $G \in L^\infty$. By the remark above, πF is the orthogonal projection of F onto H_0 in the Hilbert space $L^2(\Omega, \mathfrak{A}, P)$, and $UG - G$ is orthogonal to H_0. Hence $\pi F = F_0$ and

$$\begin{aligned} A_n F - \pi F &= \frac{1}{n} \sum_{i=0}^{n-1} U^i F_0 - F_0 + \frac{1}{n} \sum_{i=0}^{n-1} U^i (UG - G) \\ &= \frac{1}{n}(U^n G - G). \end{aligned}$$

Since $G \in L^\infty(\Omega, \mathfrak{A}, P)$ and U is an L^∞-isometry, the right hand side converges to 0 in L^∞ as $n \to \infty$.

Step 2: **$\mathbf{L^2}$-convergence.** By Step 1,

$$A_n F \longrightarrow \pi F \quad \text{in } L^2(\Omega, \mathfrak{A}, P) \tag{2.9}$$

for any $F \in H_0 + H_1$. As the linear operators A_n and π are all contractions on $L^2(\Omega, \mathfrak{A}, P)$, the convergence extends to all random variables F in the L^2 closure of $H_0 + H_1$ by an $\varepsilon/3$ argument. Therefore, in order to extend (2.9) to all $F \in L^2$ it only remains to verify that $H_0 + H_1$ is dense in $L^2(\Omega, \mathfrak{A}, P)$. But indeed, since L^∞ is dense in L^2 and $U - I$ is a bounded linear operator on L^2, H_1 is dense in the L^2-range of $U - I$, and hence by (2.7),

$$L^2(\Omega, \mathfrak{A}, P) = H_0 + \overline{(U - I)(L^2)} = H_0 + \overline{H_1} \subseteq \overline{H_0 + H_1}.$$

Step 3: **$\mathbf{L^p}$-convergence.** For $F \in L^\infty(\Omega, \mathfrak{A}, P)$, the sequence $(A_n F)_{n \in \mathbb{N}}$ is bounded in L^∞. Hence for any $p \in [1, \infty)$,

$$A_n F \longrightarrow \pi F \quad \text{in } L^p(\Omega, \mathfrak{A}, P) \tag{2.10}$$

by (2.9) and the dominated convergence theorem. Since A_n and π are contractions on each L^p space, the convergence in (2.10) extends to all $F \in L^p(\Omega, \mathfrak{A}, P)$ by an $\varepsilon/3$ argument.

Step 4: **Almost sure convergence.** By Step 1,

$$A_n F \longrightarrow \pi F \quad P\text{-almost surely} \tag{2.11}$$

for any $F \in H_0 + H_1$. Furthermore, we have already shown that $H_0 + H_1$ is dense in $L^2(\Omega, \mathfrak{A}, P)$ and hence also in $L^1(\Omega, \mathfrak{A}, P)$. Now fix an arbitrary $F \in L^1(\Omega, \mathfrak{A}, P)$, and let $(F_k)_{k \in \mathbb{N}}$ be a sequence in $H_0 + H_1$ such that $F_k \to F$ in L^1. We want to show that $A_n F$ converges almost surely as $n \to \infty$, then the limit can be identified as πF by the L^1 convergence shown in Step 3. We already know that P-almost surely,

$$\limsup_{n \to \infty} A_n F_k = \liminf_{n \to \infty} A_n F_k \quad \text{for any } k \in \mathbb{N},$$

and therefore, for $k \in \mathbb{N}$ and $\varepsilon > 0$,

$$\begin{aligned} P[\limsup A_n F - \liminf A_n F \ge \varepsilon] &\le P[\sup_n |A_n F - A_n F_k| \ge \varepsilon/2] \\ &= P[\sup_n |A_n (F - F_k)| \ge \varepsilon/2]. \end{aligned} \tag{2.12}$$

Hence we are done if we can show for any $\varepsilon > 0$ that the right hand side in (2.12) converges to 0 as $k \to \infty$. Since $E[|F - F_k|] \to 0$, the proof is now completed by Lemma 2.6 below. ■

Lemma 2.6 (Maximal ergodic theorem). *Suppose that $P = P \circ \Theta^{-1}$. Then the following statements hold for any $F \in \mathcal{L}^1(\Omega, \mathfrak{A}, P)$:*

1) $E[F; \max_{1 \le i \le n} A_i F \ge 0] \ge 0$ *for any $n \in \mathbb{N}$,*

2) $P[\sup_{n \in \mathbb{N}} |A_n F| \ge c] \le \frac{1}{c} E[|F|]$ *for any $c \in (0, \infty)$.*

Note the similarity to the maximal inequality for martingales. The proof is not very intuitive but not difficult either:

Proof.

1) Let $M_n = \max_{1 \le i \le n} (F + F \circ \Theta + \cdots + F \circ \Theta^{i-1})$, and let $B = \{M_n \ge 0\} = \{\max_{1 \le i \le n} A_i F \ge 0\}$. Then $M_n = F + M_{n-1}^+ \circ \Theta$, and hence

$$F = M_n^+ - M_{n-1}^+ \circ \Theta \ge M_n^+ - M_n^+ \circ \Theta \quad \text{on } B.$$

Taking expectations we obtain

$$\begin{aligned} E[F; B] &\ge E[M_n^+; B] - E[M_n^+ \circ \Theta; \Theta^{-1}(\Theta(B))] \\ &\ge E[M_n^+] - E[(M_n^+ 1_{\Theta(B)}) \circ \Theta] \\ &= E[M_n^+] - E[M_n^+; \Theta(B)] \ge 0 \end{aligned}$$

since $B \subset \Theta^{-1}(\Theta(B))$.

2) We may assume that F is non-negative - otherwise we can apply the corresponding estimate for $|F|$. For $F \ge 0$ and $c \in (0, \infty)$,

$$E\left[F - c; \max_{1 \le i \le n} A_i F \ge c\right] \ge 0$$

by 1). Therefore,

$$c \cdot P\left[\max_{i \le n} A_i F \ge c\right] \le E\left[F; \max_{i \le n} A_i F \ge c\right] \le E[F]$$

for any $n \in \mathbb{N}$. As $n \to \infty$ we can conclude that

$$c \cdot P\left[\sup_{i \in \mathbb{N}} A_i F \ge c\right] \le E[F].$$

The assertion now follows by replacing c by $c - \varepsilon$ and letting ε tend to zero. ■

Ergodic theorems for Markov chains

Suppose that Θ is the shift on $\Omega = S^{\mathbb{Z}_+}$, and (X_n, P_μ) is a canonical time-homogeneous Markov chain with state space S, initial distribution μ and transition kernel π. Then Θ is measure-preserving w.r.t. P_μ if and only if μ is an invariant probability measure for π. Furthermore, by Theorem 2.2, the measure P_μ is ergodic if and only if any function $h \in \mathcal{L}^2(\mu)$ satisfying $\mathcal{L}h = 0$ μ-almost surely is μ-almost surely constant. Other necessary and sufficient conditions for ergodicity are summarized in the following exercise:

Exercise (Ergodicity of stationary Markov chains). Suppose that μ is an invariant probability measure for the transition kernel π of a canonical Markov chain (X_n, P_x) with state space $(S, \mathcal{B})$. Prove that the following statements are equivalent:

(i) P_μ is ergodic.

(ii) For any $F \in \mathcal{L}^1(P_\mu)$,

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \;\to\; E_\mu[F] \quad P_\mu\text{-almost surely.}$$

(iii) For any $f \in \mathcal{L}^1(\mu)$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \;\to\; \int f\, d\mu \quad P_\mu\text{-almost surely.}$$

(iv) For any $B \in \mathcal{B}$,

$$\frac{1}{n}\sum_{i=0}^{n-1} \pi^i(x, B) \to \mu(B) \quad \text{for } \mu\text{-a.e. } x \in S.$$

(v) For any $B \in \mathcal{B}$ such that $\mu(B) > 0$,

$$P_x[T_B < \infty] > 0 \quad \text{for } \mu\text{-a.e. } x \in S.$$

(vi) Any $B \in \mathcal{B}$ such that $\pi 1_B = 1_B$ μ-a.s. has measure $\mu(B) \in \{0, 1\}$.

In particular, we point out the following consequences of Birkhoff's ergodic theorem for ergodic stationary Markov chains:

a) **Law of large numbers**: For any function $f \in \mathcal{L}^1(S, \mu)$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \to \int f d\mu \quad P_\mu\text{-almost surely as } n \to \infty. \tag{2.13}$$

The law of large numbers for Markov chains is exploited in Markov chain Monte Carlo (MCMC) methods for the numerical estimation of integrals w.r.t. a given probability measure μ.

b) **Estimation of the transition kernel**: For any Borel sets $A, B \in \mathcal{B}$,

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_{A\times B}(X_i, X_{i+1}) \to E[1_{A\times B}(X_0, X_1)] = \int_A \mu(dx) p(x, B) \tag{2.14}$$

P_μ-a.s. as $n \to \infty$. This is applied in statistics of Markov chains for estimating both the invariant measure and the transition kernel of a Markov chain from observed values.

Both applications lead to new questions:

- How can the deviation of the ergodic average from its limit be quantified?
- What can be said if the initial distribution of the Markov chain is not a stationary distribution?

We return to the first question later - in particular in Sections 2.4 and 3.2. Let us now consider the second question. The next remark contains some preliminary observations:

Remark (Non-stationary initial distributions).

1) If ν is a probability measure on S that is absolutely continuous w.r.t. a stationary distribution μ then the law P_ν of the Markov chain with initial distribution ν is absolutely continuous w.r.t. P_μ. Therefore, in this case P_ν-almost sure convergence holds in Birkhoff's Theorem. More generally, P_ν-almost sure convergence holds whenever $\nu\pi^k$ is absolutely continuous w.r.t. μ for some $k \in \mathbb{N}$, since the limits of the ergodic averages coincide for the original Markov chain $(X_n)_{n\geq 0}$ and the chain $(X_{n+k})_{n\geq 0}$ with initial distribution $\nu\pi^k$.

2) Since $P_\mu = \int P_x\, \mu(dx)$, P_μ-almost sure convergence also implies P_x-almost sure convergence of the ergodic averages for μ-almost every x.

3) Nevertheless, P_ν-almost sure convergence does not hold in general. In particular, there are many Markov chains that have several stationary distributions. If ν and μ are different stationary distributions for the transition kernel π then the limits $E_\nu[F|\mathcal{J}]$ and $E_\mu[F|\mathcal{J}]$ of the ergodic averages A_nF w.r.t. P_ν and P_μ respectively do *not coincide*.

We conclude this section with two necessary and sufficient conditions guaranteeing that the ergodic averages converge for every initial distribution:

Corollary 2.7 (Liouville Theorem and ergodicity for arbitrary initial distributions). Suppose that μ is an invariant probability measure for the transition kernel π of a canonical Markov chain (X_n, P_x). Then the following statements are all equivalent:

(i) For any $x \in S$ and $B \in \mathcal{B}$ with $\mu(B) > 0$,

$$P_x\left[X_n \in B \text{ infinitely often}\right] \;=\; 1. \tag{2.15}$$

(ii) The constants are the only bounded harmonic functions for the generator $\mathcal{L} = \pi - I$.

(iii) For any bounded measurable function $F : \Omega \to \mathbb{R}$ and any $x \in S$,

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \;\to\; E_\mu[F] \quad P_x\text{-almost surely.}$$

Proof. We first note that by the exercise above, each of the conditions (i), (ii) and (iii) implies ergodicity of P_μ. We now show that "(i)⇒(ii)⇒(iii)⇒(i)".

(i)⇒(ii). Suppose that $h : S \to \mathbb{R}$ is a bounded harmonic function, and let $x \in S$. Then $h(X_n)$ is a bounded martingale w.r.t. P_x. By the L^1 martingale convergence theorem, $h(X_n)$ converges P_x-almost surely, and

$$h(x) \;=\; E_x[h(X_0)] \;=\; E_x[\lim h(X_n)]. \tag{2.16}$$

Now fix $c \in \mathbb{R}$ such that both $\mu(h \geq c)$ and $\mu(h \leq c)$ are strictly positive. For example, we may choose c as a median of $\mu \circ h^{-1}$. Then by (i), for any $x \in S$,

$$P_x[h(X_n) \geq c \text{ infinitely often}] \;=\; 1 \;=\; P_x[h(X_n) \leq c \text{ infinitely often}].$$

Hence by (2.16), $h(x) = c$ for any $x \in S$.

(ii)⇒(iii). Consider the event $A := \{\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \;\to\; E_\mu[F]\}$. Observe that by ergodicity, $P_\mu[A] = 1$. Moreover, A is shift-invariant, and hence $h(x) := P_x[A]$ is a bounded harmonic function. Therefore, by (ii), h is constant, and thus

$$h(x) \;=\; \int h\, d\mu \;=\; P_\mu[A] \;=\; 1 \qquad \text{for any } x \in S.$$

(iii)⇒(i). Let $B \in \mathcal{B}$ with $\mu(B) > 0$. Then applying (iii) with $F = 1_B(X_0)$ shows that (2.15) holds. ■

Remark. The assumption (2.15) is satisfied if the Markov chain is "globally Harris recurrent", see Meyn and Tweedie [40, Theorem 9.1.4]. Necessary and sufficient conditions for Harris recurrence are given in Theorem 1.11.

2.2. Ergodic theory in continuous time

We now extend the results in Section 2.1 to the continuous time case. Indeed we will see that the main results in continuous time can be deduced from those in discrete time.

Ergodic theorem

Let $(\Omega, \mathfrak{A}, P)$ be a probability space. Furthermore, suppose that we are given a product-measurable map

$$\begin{aligned} \Theta : [0,\infty) \times \Omega &\to \Omega \\ (t\ ,\ \omega) &\mapsto \Theta_t(\omega) \end{aligned}$$

satisfying the semigroup property

$$\Theta_0 = \mathrm{id}_\Omega \quad \text{and} \quad \Theta_t \circ \Theta_s = \Theta_{t+s} \quad \text{for any } t, s \geq 0. \tag{2.17}$$

The analogue in discrete time are the maps $\Theta_n(\omega) = \Theta^n(\omega)$. As in the discrete time case, the main example for the maps Θ_t are the time-shifts on the canonical probability space of a stochastic process:

Example (Stationary processes in continuous time). Suppose $\Omega = C([0,\infty), S)$ or $\Omega = \mathcal{D}([0,\infty), S)$ is the space of continuous, right-continuous or càdlàg functions from $[0,\infty)$ to S, $X_t(\omega) = \omega(t)$ is the evolution of a function at time t, and $\mathfrak{A} = \sigma(X_t : t \in [0,\infty))$. Then, by right continuity of $t \mapsto X_t(\omega)$, the time-shift $\Theta : [0,\infty) \times \Omega \to \Omega$ defined by

$$\Theta_t(\omega) = \omega(t + \cdot) \quad \text{for } t \in [0,\infty), \omega \in \Omega,$$

is product-measurable and satisfies the semigroup property (2.17). Suppose moreover that P is a probability measure on $(\Omega, \mathfrak{A})$. Then the continuous-time stochastic process $((X_t)_{t\in[0,\infty)}, P)$ is **stationary**, i.e.,

$$(X_{s+t})_{t\in[0,\infty)} \sim (X_t)_{t\in[0,\infty)} \quad \text{under } P \text{ for any } s \in [0,\infty),$$

if and only if P is **shift-invariant**, i.e., iff $P \circ \Theta_s^{-1} = P$ for any $s \in [0,\infty)$.

The σ-algebra of shift-invariant events is defined by

$$\mathcal{J} = \left\{ A \in \mathfrak{A} : A = \Theta_s^{-1}(A) \text{ for any } s \in [0,\infty) \right\}.$$

Verify for yourself that the definition is consistent with the one in discrete time, and that $\mathcal{J}$ is indeed a σ-algebra.

Theorem 2.8 (Ergodic theorem in continuous time). Suppose that P is a probability measure on $(\Omega, \mathfrak{A})$ satisfying $P \circ \Theta_s^{-1} = P$ for any $s \in [0,\infty)$. Then for any $p \in [1,\infty]$ and any random variable $F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P)$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t F \circ \Theta_s \, ds = E[F|\mathcal{J}] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P). \tag{2.18}$$

Similarly to the discrete time case, we use the notation

$$A_t F = \frac{1}{t} \int_0^t F \circ \Theta_s \, ds$$

for the ergodic averages. It is straightforward to verify that A_t is a contraction on $\mathcal{L}^p(\Omega, \mathfrak{A}, P)$ for any $p \in [1,\infty]$ provided the maps Θ_s are measure-preserving.

Proof.

Step 1: **Time discretization.** Suppose that F is uniformly bounded, and let

$$\hat{F} := \int_0^1 F \circ \Theta_s \, ds.$$

Since $(s, \omega) \mapsto \Theta_s(\omega)$ is product-measurable, $\hat{F}$ is a well-defined uniformly bounded random variable. Furthermore, by the semigroup property (2.17),

$$A_n F = \hat{A}_n \hat{F} \quad \text{for any } n \in \mathbb{N}, \text{ where} \quad \hat{A}_n \hat{F} := \frac{1}{n} \sum_{i=0}^{n-1} \hat{F} \circ \Theta_i$$

denotes the discrete time ergodic average of $\hat{F}$. If $t \in [0, \infty)$ is not an integer then we can estimate

$$\begin{aligned} |A_t F - \hat{A}_{\lfloor t \rfloor} \hat{F}| = |A_t F - A_{\lfloor t \rfloor} F| &\le \left| \frac{1}{t} \int_{\lfloor t \rfloor}^t F \circ \Theta_s \, ds \right| + \left(\frac{1}{\lfloor t \rfloor} - \frac{1}{t} \right) \cdot \left| \int_0^{\lfloor t \rfloor} F \circ \Theta_s \, ds \right| \\ &\le \frac{1}{t} \sup |F| + \left(\frac{t}{\lfloor t \rfloor} - 1 \right) \cdot \sup |F|. \end{aligned}$$

The right-hand side is independent of ω and converges to 0 as $t \to \infty$. Hence by the ergodic theorem in discrete time,

$$\lim_{t \to \infty} A_t F = \lim_{n \to \infty} \hat{A}_n \hat{F} = E[\hat{F} | \hat{\mathcal{J}}] \quad P\text{-a.s. and in } L^p \text{ for any } p \in [1, \infty), \tag{2.19}$$

where $\hat{\mathcal{J}} = \left\{ A \in \mathfrak{A} : \Theta_1^{-1}(A) = A \right\}$ is the collection of Θ_1-invariant events.

Step 2: **Identification of the limit.** Next we show that the limit in (2.19) coincides with the conditional expectation $E[F|\mathcal{J}]$ P-almost surely. To this end note that the limit superior of $A_t F$ as $t \to \infty$ is $\mathcal{J}$-measurable, since

$$(A_t F) \circ \Theta_s = \frac{1}{t} \int_0^t F \circ \Theta_u \circ \Theta_s \, du = \frac{1}{t} \int_0^t F \circ \Theta_{u+s} \, du = \frac{1}{t} \int_s^{s+t} F \circ \Theta_u \, du$$

has the same limit superior as $A_t F$ for any $s \in [0, \infty)$. Since L^1 convergence holds,

$$\lim_{t \to \infty} A_t F = E[\lim A_t F | \mathcal{J}] = \lim E[A_t F | \mathcal{J}] = \lim_{t \to \infty} \frac{1}{t} \int_0^t E[F \circ \Theta_s | \mathcal{J}] \, ds$$

P-almost surely. Since Θ_s is measure-preserving, it can be easily verified that $E[F \circ \Theta_s | \mathcal{J}] = E[F|\mathcal{J}]$ P-almost surely for any $s \in [0, \infty)$. Hence

$$\lim_{t \to \infty} A_t F = E[F|\mathcal{J}] \quad P\text{-almost surely.}$$

Step 3: **Extension to $\mathbf{F \in \mathcal{L}^p}$.** Since $\mathcal{F}_b(\Omega)$ is a dense subset of $\mathcal{L}^p(\Omega, \mathfrak{A}, P)$ and A_t is a contraction w.r.t. the L^p-norm, the L^p convergence in (2.18) holds for any $F \in \mathcal{L}^p$ by an $\varepsilon/3$-argument. In order to show that almost sure convergence holds for any $F \in \mathcal{L}^1$ we apply once more the maximal ergodic theorem 2.6. For $t \ge 1$,

$$|A_t F| \le \frac{1}{t} \int_0^{\lfloor t \rfloor + 1} |F \circ \Theta_s| \, ds = \frac{\lfloor t \rfloor + 1}{t} \hat{A}_{\lfloor t \rfloor + 1} |\hat{F}| \le 2 \hat{A}_{\lfloor t \rfloor + 1} |\hat{F}|.$$

Hence for any $c \in (0,\infty)$,

$$P\left[\sup_{t>1}|A_tF| \geq c\right] \leq P\left[\sup_{n\in\mathbb{N}}\hat{A}_n|\hat{F}| \geq c/2\right] \leq \frac{2}{c}E[|\hat{F}|] \leq \frac{2}{c}E[|F|].$$

Thus we have deduced a maximal inequality in continuous time from the discrete time maximal ergodic theorem. The proof of almost sure convergence of the ergodic averages can now be completed similarly to the discrete time case by approximating F by uniformly bounded functions, cf. the proof of Theorem 2.5 above. ■

The ergodic theorem implies the following alternative characterizations of ergodicity:

Corollary 2.9 (Ergodicity and decay of correlations). Suppose that $P \circ \Theta_s^{-1} = P$ for any $s \in [0,\infty)$. Then the following statements are equivalent:

(i) P is ergodic w.r.t. $(\Theta_s)_{s\geq 0}$.

(ii) For any $F \in \mathcal{L}^2(\Omega, \mathfrak{A}, P)$,

$$\mathrm{Var}\left(\frac{1}{t}\int_0^t F\circ\Theta_s\,ds\right) \to 0 \quad \text{as } t\to\infty.$$

(iii) For any $F \in \mathcal{L}^2(\Omega, \mathfrak{A}, P)$,

$$\frac{1}{t}\int_0^t \mathrm{Cov}\left(F\circ\Theta_s, F\right)\,ds \to 0 \quad \text{as } t\to\infty.$$

(iv) For any $A, B \in \mathfrak{A}$,

$$\frac{1}{t}\int_0^t P\left[A\cap\Theta_s^{-1}(B)\right]\,ds \to P[A]\,P[B] \quad \text{as } t\to\infty.$$

The proof is left as an exercise.

Applications

a) Flows of ordinary differential equations
Let $b : \mathbb{R}^d \to \mathbb{R}^d$ be a smooth (C^∞) vector field. The flow $(\Theta_t)_{t\in\mathbb{R}}$ of b is a dynamical system on $\Omega = \mathbb{R}^d$ defined by

$$\frac{d}{dt}\Theta_t(\omega) = b(\Theta_t(\omega)), \quad \Theta_0(\omega) = \omega \quad \text{for any } \omega\in\mathbb{R}^d. \tag{2.20}$$

For a smooth function $F : \mathbb{R}^d \to \mathbb{R}$ and $t \in \mathbb{R}$ let

$$(U_tF)(\omega) = F(\Theta_t(\omega)).$$

Then the flow equation (2.20) implies the **forward equation**

$$\frac{d}{dt}U_tF = \dot{\Theta}_t\cdot(\nabla F)\circ\Theta_t = (b\cdot\nabla F)\circ\Theta_t, \quad \text{i.e.,}$$

$$\text{(F)} \qquad \frac{d}{dt}U_tF = U_t\mathcal{L}F \qquad \text{where} \qquad \mathcal{L}F = b\cdot\nabla F$$

is the **infinitesimal generator** of the time-evolution. There is also a corresponding **backward equation** that follows from the identity $U_h U_{t-h} F = U_t F$. By differentiating w.r.t. h at $h = 0$ we obtain $\mathcal{L} U_t F - \frac{d}{dt} U_t F = 0$, and thus

$$\text{(B)} \qquad \frac{d}{dt} U_t F = \mathcal{L} U_t F = b \cdot \nabla(F \circ \Theta_t).$$

The backward equation can be used to identify **invariant measures** for the flow $(\Theta_t)_{t \in \mathbb{R}}$. Suppose that P is a positive measure on $\mathbb{R}^d$ with a smooth density ϱ w.r.t. Lebesgue measure λ, and let $F \in C_0^\infty(\mathbb{R}^d)$. Then

$$\frac{d}{dt} \int U_t F \, dP = \int b \cdot \nabla(F \circ \Theta_t) \varrho \, d\lambda = \int F \circ \Theta_t \operatorname{div}(\varrho b) \, d\lambda.$$

Hence we can conclude that if

$$\operatorname{div}(\varrho b) = 0$$

then $\int F \circ \Theta_t \, dP = \int U_t F \, dP = \int F \, dP$ for any $F \in C_0^\infty(\mathbb{R}^d)$ and $t \geq 0$, i.e.,

$$P \circ \Theta_t^{-1} = P \qquad \text{for any } t \in \mathbb{R}.$$

Example (Hamiltonian systems). In Hamiltonian mechanics, the state space of a system is $\Omega = \mathbb{R}^{2d}$ where a vector $\omega = (q, p) \in \Omega$ consists of the position variable $q \in \mathbb{R}^d$ and the momentum variable $p \in \mathbb{R}^d$. If we choose units such that the mass is equal to one then the total energy is given by the **Hamiltonian**

$$H(q, p) = \frac{1}{2} |p|^2 + V(q)$$

where $\frac{1}{2}|p|^2$ is the kinetic energy and $V(q)$ is the potential energy. Here we assume $V \in C^\infty(\mathbb{R}^d)$. The dynamics is given by the equations of motion

$$\begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p}(q, p) = p, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial q}(q, p) = -\nabla V(q). \end{aligned}$$

A simple example is the harmonic oscillator (pendulum) where $d = 1$ and $V(q) = \frac{1}{2} q^2$. Let $(\Theta_t)_{t \in \mathbb{R}}$ be the corresponding flow of the vector field

$$b(q, p) = \begin{pmatrix} \frac{\partial H}{\partial p}(q, p) \\ -\frac{\partial H}{\partial q}(q, p) \end{pmatrix} = \begin{pmatrix} p \\ -\nabla V(q) \end{pmatrix}.$$

The first important observation is that the system does not explore the whole state space, since the energy is conserved:

$$\frac{d}{dt} H(q, p) = \frac{\partial H}{\partial q}(q, p) \cdot \frac{dq}{dt} + \frac{\partial H}{\partial p}(q, p) \cdot \frac{dp}{dt} = (b \cdot \nabla H)(q, p) = 0 \tag{2.21}$$

where the dot stands both for the Euclidean inner product in $\mathbb{R}^d$ and in $\mathbb{R}^{2d}$. Thus $H \circ \Theta_t$ is constant, i.e., $t \mapsto \Theta_t(\omega)$ remains on a fixed energy shell.

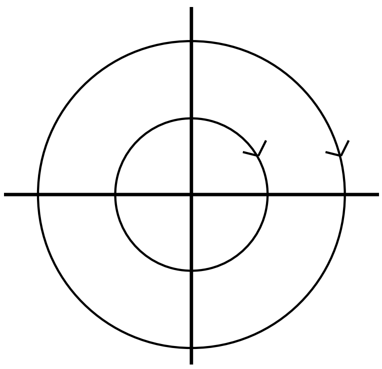


Figure 2.1.: Trajectories of harmonic oscillator

As a consequence, there are infinitely many invariant measures. Indeed, suppose that $\varrho(q,p) = g(H(q,p))$ for a smooth non-negative function g on $\mathbb{R}$. Then the measure

$$P(d\omega) = g(H(\omega))\,\lambda^{2d}(d\omega)$$

is invariant w.r.t. (Θ_t) because

$$\operatorname{div}(\varrho b) = b\cdot\nabla\varrho + \varrho\operatorname{div}(b) = (g'\circ H)\,(b\cdot\nabla H) + \varrho\left(\frac{\partial^2 H}{\partial q\partial p} - \frac{\partial^2 H}{\partial p\partial q}\right) = 0$$

by (2.21). What about ergodicity? For any Borel set $B \subseteq \mathbb{R}$, the event $\{H \in B\}$ is invariant w.r.t. (Θ_t) by conservation of the energy. Therefore, ergodicity can not hold if g is a smooth function. However, the example of the harmonic oscillator shows that ergodicity may hold if we replace g by a Dirac measure, i.e., if we restrict to a fixed energy shell.

Remark (Deterministic vs. stochastic dynamics). The flow of an ordinary differential equation can be seen as a very special Markov process - with a deterministic dynamics. More generally, the ordinary differential equation can be replaced by a stochastic differential equation to obtain Itô type diffusion processes, cf. below. In this case it is not possible any more to choose Ω as the state space of the system as we did above - instead Ω has to be replaced by the space of all trajectories with appropriate regularity properties.

b) Gaussian processes

Simple examples of non-Markovian stochastic processes can be found in the class of Gaussian processes. We consider the canonical model with $\Omega = \mathcal{D}([0,\infty),\mathbb{R})$, $X_t(\omega) = \omega(t)$, $\mathfrak{A} = \sigma(X_t : t \in \mathbb{R}_+)$, and $\Theta_t(\omega) = \omega(t+\cdot)$. In particular,

$$X_t\circ\Theta_s = X_{t+s} \qquad \text{for any } t,s \geq 0.$$

Let P be a probability measure on $(\Omega,\mathfrak{A})$. The stochastic process (X_t,P) is called a **Gaussian process** if and only if $(X_{t_1},\dots,X_{t_n})$ has a multivariate normal distribution for any $n \in \mathbb{N}$ and $t_1,\dots,t_n \in \mathbb{R}_+$ (Recall that it is not enough to assume that X_t is normally distributed for any t!). The law P of a Gaussian process is uniquely determined by the averages and covariances

$$m(t) = E[X_t], \quad c(s,t) = \operatorname{Cov}(X_s,X_t), \quad s,t\geq 0.$$

It can be shown (Exercise) that a Gaussian process is stationary if and only if $m(t)$ is constant, and

$$c(s,t) = r(|s-t|)$$

for some function $r : \mathbb{R}_+ \to \mathbb{R}$ (**auto-correlation function**). To obtain a necessary condition for ergodicity note that if (X_t,P) is stationary and ergodic then $\frac{1}{t}\int_0^t X_s\,ds$ converges to the constant average m, and hence

$$\operatorname{Var}\left(\frac{1}{t}\int_0^t X_s\,ds\right) \to 0 \qquad \text{as } t\to\infty.$$

On the other hand, by Fubini's theorem,

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{t}\int_0^t X_s\,ds\right) &= \operatorname{Cov}\left(\frac{1}{t}\int_0^t X_s\,ds, \frac{1}{t}\int_0^t X_u\,du\right)\\
&= \frac{1}{t^2}\int_0^t\int_0^t \operatorname{Cov}(X_s,X_u)\,duds = \frac{1}{2t^2}\int_0^t\int_0^s r(s-u)\,duds\\
&= \frac{1}{2t^2}\int_0^t (t-v)r(v)\,dv = \frac{1}{2t}\int_0^t\left(1-\frac{v}{t}\right)r(v)\,dv\\
&\sim \frac{1}{2t}\int_0^t r(v)dv \quad \text{asymptotically as } t\to\infty.
\end{aligned}$$

Hence ergodicity can only hold if

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t r(v)\,dv = 0.$$

It can be shown by Spectral analysis/Fourier transform techniques that this condition is also sufficient for ergodicity, cf. e.g. Lindgren, "Lectures on Stationary Stochastic Processes" [36].

c) Random Fields

We have stated the ergodic theorem for temporal, i.e., one-dimensional averages. There are corresponding results in the multi-dimensional case, i.e., $t \in \mathbb{Z}^d$ or $t \in \mathbb{R}^d$, cf. e.g. Stroock, "Probability Theory: An Analytic View" [52]. These apply for instance to ergodic averages of the form

$$A_t F = \frac{1}{(2t)^d} \int_{(-t,t)^d} F \circ \Theta_s \, ds, \quad t \in \mathbb{R}_+,$$

where $(\Theta_s)_{s\in\mathbb{R}^d}$ is a group of measure-preserving transformations on a probability space $(\Omega, \mathfrak{A}, P)$. Multi-dimensional ergodic theorems are important for the study of stationary random fields. Here we just mention briefly two typical examples:

Example (Massless Gaussian free field on $\mathbb{Z}^d$). Let $\Omega = \mathbb{R}^{\mathbb{Z}^d}$ where $d \geq 3$, and let $X_s(\omega) = \omega_s$ for $\omega = (\omega_s) \in \Omega$. The **massless Gaussian free field** is the probability measure P on Ω given informally by

$$\text{“}\, P(d\omega) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{\substack{s,t\in\mathbb{Z}^d \\ |s-t|=1}} |\omega_t - \omega_s|^2 \right) \prod_{s\in\mathbb{Z}^d} d\omega_s \,\text{”}. \tag{2.22}$$

The expression is not rigorous since the Gaussian free field on $\mathbb{R}^{\mathbb{Z}^d}$ does not have a density w.r.t. a product measure. Indeed, the density in (2.22) would be infinite for almost every ω. Nevertheless, P can be defined rigorously as the law of a centered Gaussian process (or random field) $(X_s)_{s\in\mathbb{Z}^d}$ with covariances

$$\mathrm{Cov}(X_s, X_t) = G(s,t) \quad \text{for any } s,t \in \mathbb{Z}^d,$$

where $G(s,t) = \sum_{n=0}^{\infty} p^n(s,t)$ is the Green's function of the Random Walk on $\mathbb{Z}^d$. The connection to the informal expression in (2.22) is made by observing that the generator of the random walk is the discrete Laplacian $\Delta_{\mathbb{Z}^d}$, and the informal density in (2.22) takes the form

$$Z^{-1} \exp\left(-\frac{1}{2} (\omega, \Delta_{\mathbb{Z}^d}\omega)_{l^2(\mathbb{Z}^d)} \right).$$

For $d \geq 3$, the random walk on $\mathbb{Z}^d$ is transient. Hence the Green's function is finite, and one can show that there is a unique centered Gaussian measure P on Ω with covariance function $G(s,t)$. Since $G(s,t)$ depends only on $s-t$, the measure P is stationary w.r.t. the shift $\Theta_s(\omega) = \omega(s+\cdot)$, $s \in \mathbb{Z}^d$. Furthermore, decay of correlations holds for $d \geq 3$ since

$$G(s,t) \sim |s-t|^{2-d} \quad \text{as } |s-t| \to \infty.$$

It can be shown that this implies ergodicity of P, i.e., the P-almost sure limits of spatial ergodic averages are constant. In dimensions $d = 1, 2$ the Green's function is infinite and the massless Gaussian free field does not exist. However, in any dimension $d \in \mathbb{N}$ it is possible to define in a similar way the Gaussian free field with mass $m \geq 0$, where G is replaced by the Green's function of the operator $m^2 - \Delta_{\mathbb{Z}^d}$.

Example (Markov chains in random environment). Suppose that $(\Theta_x)_{x\in\mathbb{Z}^d}$ is stationary and ergodic on a probability space $(\Omega, \mathfrak{A}, P)$, and let $q : \Omega \times \mathbb{Z}^d \to [0,1]$ be a stochastic kernel from Ω to $\mathbb{Z}^d$. Then random transition probabilities on $\mathbb{Z}^d$ can be defined by setting

$$\pi(\omega, x, y) = q\,(\Theta_x(\omega), y - x) \quad \text{for any } \omega \in \Omega \text{ and } x, y \in \mathbb{Z}^d.$$

For any fixed $\omega \in \Omega$, $\pi(\omega, \cdot)$ is the transition matrix of a Markov chain on $\mathbb{Z}^d$. The variable ω is called the **random environment** - it determines which transition matrix is applied. One is now considering a two-stage model where at first an environment ω is chosen at random, and then (given ω) a Markov chain is run in this environment. Typical questions that arise are the following:

- **Quenched asymptotics.** How does the Markov chain with transition kernel $\pi(\omega, \cdot, \cdot)$ behave asymptotically for a typical ω (i.e., for P-almost every $\omega \in \Omega$)?

- **Annealed asymptotics.** What can be said about the asymptotics if one is averaging over ω w.r.t. P?

For an introduction to these and other questions see e.g. Sznitman, "Ten lectures on Random media" [4].

Ergodic theory for Markov processes

We now return to our main interest in these notes: The application of ergodic theorems to Markov processes in continuous time. Suppose that $(p_t)_{t\in[0,\infty)}$ is a transition function of a time-homogeneous Markov process (X_t, P_μ) on $(\Omega, \mathfrak{A})$. We assume that $(X_t)_{t\in[0,\infty)}$ is the canonical process on $\Omega = \mathcal{D}([0,\infty), S)$, $\mathfrak{A} = \sigma(X_t : t \in [0,\infty))$, and μ is the law of X_0 w.r.t. P_μ. The measure μ is a stationary distribution for (p_t) iff $\mu p_t = \mu$ for any $t \in [0,\infty)$. The existence of stationary distributions can be shown by the theorems of Krylov-Bogoliubov and Foguel. The proof of the next theorem is left as an exercise.

Theorem 2.10 (Characterizations of ergodicity in continuous time). 1) The shift semigroup $\Theta_s(\omega) = \omega(t+\cdot)$, $t \geq 0$, preserves the measure P_μ if and only if μ is a stationary distribution for $(p_t)_{t\geq 0}$.

2) In this case, the following statements are all equivalent

(i) P_μ is ergodic.

(ii) For any $f \in \mathcal{L}^2(S, \mu)$,

$$\frac{1}{t}\int_0^t f(X_s)\,ds \to \int f\,d\mu \quad P_\mu\text{-a.s. as } t \to \infty.$$

(iii) For any $f \in \mathcal{L}^2(S, \mu)$,

$$\mathrm{Var}_{P_\mu}\left(\frac{1}{t}\int_0^t f(X_s)\,ds\right) \to 0 \quad \text{as } t \to \infty.$$

(iv) For any $f, g \in \mathcal{L}^2(S, \mu)$,

$$\frac{1}{t}\int_0^t \mathrm{Cov}_{P_\mu}\left(g(X_0), f(X_s)\right)\,ds \to 0 \quad \text{as } t \to \infty.$$

(v) For any $A, B \in \mathcal{B}$,

$$\frac{1}{t}\int_0^t P_\mu\left[X_0 \in A, X_s \in B\right]\,ds \to \mu(A)\mu(B) \quad \text{as } t \to \infty.$$

(vi) For any $B \in \mathcal{B}$,

$$\frac{1}{t}\int_0^t p_s(x, B)\,ds \to \mu(B) \quad \mu\text{-a.e. as } t \to \infty.$$

(vii) For any $B \in \mathcal{B}$ with $\mu(B) > 0$,

$$P_x[T_B < \infty] > 0 \quad \text{for } \mu\text{-a.e. } x \in S.$$

(viii) For any $B \in \mathcal{B}$ such that $p_t 1_B = 1_B$ μ-a.e. for any $t \geq 0$,

$$\mu(B) \in \{0, 1\}.$$

(ix) Any function $h \in \mathcal{F}_b(S)$ satisfying $p_t h = h$ μ-a.e. for any $t \geq 0$ is constant up to a set of μ-measure zero.

Again, ergodicity can be verified by the strong Feller property:

Corollary 2.11. Suppose that one of the transition kernels p_t, $t > 0$, is strong Feller. Then P_μ is stationary and ergodic for any stationary distribution μ of $(p_t)_{t\geq 0}$ that has connected support.

The proof is similar to that of the corresponding result in discrete time, cf. Corollary 2.4.

Example (Brownian motion on $\mathbb{R}/\mathbb{Z}$). A Brownian motion (X_t) on the circle $\mathbb{R}/\mathbb{Z}$ can be obtained by considering a Brownian motion (B_t) on $\mathbb{R}$ modulo the integers, i.e.,

$$X_t = B_t - \lfloor B_t \rfloor \in [0, 1) \subseteq \mathbb{R}/\mathbb{Z}.$$

Since Brownian motion on $\mathbb{R}$ has the smooth transition density

$$p_t^{\mathbb{R}}(x, y) = (2\pi t)^{-1/2} \exp(-|x - y|^2/(2t)),$$

the transition density of Brownian motion on $\mathbb{R}/\mathbb{Z}$ w.r.t. the uniform distribution is given by

$$p_t(x, y) = \sum_{n\in\mathbb{Z}} p_t^{\mathbb{R}}(x, y + n) = \frac{1}{\sqrt{2\pi t}} \sum_{n\in\mathbb{Z}} e^{-\frac{|x-y-n|^2}{2t}} \qquad \text{for any } t > 0 \text{ and } x, y \in [0, 1).$$

Since p_t is a smooth function with bounded derivatives of all orders, the transition kernels are strong Feller for any $t > 0$. The uniform distribution on $\mathbb{R}/\mathbb{Z}$ is stationary for $(p_t)_{t\geq 0}$. Therefore, by Corollary 2.11, Brownian motion on $\mathbb{R}/\mathbb{Z}$ with uniform initial distribution is a stationary and ergodic Markov process.

A similar reasoning as in the last example can be carried out for general non-degenerate diffusion processes on $\mathbb{R}^d$. These are Markov processes generated by a second order differential operator of the form

$$\mathcal{L} = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x) \frac{\partial}{\partial x_i}.$$

By PDE theory it can be shown that if the coefficients are locally Hölder continuous, the matrix $(a_{ij}(x))$ is non-degenerate for any x, and appropriate growth conditions hold at infinity then there is a unique transition semigroup $(p_t)_{t\geq 0}$ with a smooth transition density corresponding to $\mathcal{L}$, cf. e.g. $[XXX]$. Therefore, Corollary 2.11 can be applied to prove that the law of a corresponding Markov process with stationary initial distribution is stationary and ergodic.

2.3. Structure of invariant measures

In this section we apply the ergodic theorem to study the structure of the set of all invariant measures w.r.t. a given one-parameter family of transformations $(\Theta_t)_{t\geq 0}$, as well as the structure of the set of all stationary distributions of a given transition semigroup $(p_t)_{t\geq 0}$.

The convex set of Θ-invariant probability measures

Let $\Theta : \mathbb{R}_+ \times \Omega \to \Omega$, $(t,\omega) \mapsto \Theta_t(\omega)$ be product-measurable on $(\Omega, \mathfrak{A})$ satisfying the semigroup property

$$\Theta_0 = \mathrm{id}_\Omega, \quad \Theta_t \circ \Theta_s = \Theta_{t+s} \quad \text{for any } t,s \geq 0,$$

and let $\mathcal{J} = \left\{A \in \mathfrak{A} : \Theta_t^{-1}(A) = A \text{ for any } t \geq 0\right\}$. Alternatively, the results will also hold in the discrete time case, i.e., $\mathbb{R}_+$ may be replaced by $\mathbb{Z}_+$. We denote by

$$\mathcal{S}(\Theta) = \left\{P \in \mathcal{P}(\Omega) : P \circ \Theta_t^{-1} = P \text{ for any } t \geq 0\right\}$$

the set of all (Θ_t)-invariant (stationary) probability measures on $(\Omega, \mathfrak{A})$.

Lemma 2.12 (Singularity of ergodic probability measures). *Suppose $P, Q \in \mathcal{S}(\Theta)$ are distinct ergodic probability measures. Then P and Q are singular on the σ-algebra $\mathcal{J}$, i.e., there exists an event $A \in \mathcal{J}$ such that $P[A] = 1$ and $Q[A] = 0$.*

Proof. This is a direct consequence of the ergodic theorem. If $P \neq Q$ then there is a random variable $F \in \mathcal{F}_b(\Omega)$ such that $\int F\,dP \neq \int F\,dQ$. The event

$$A := \left\{\limsup_{t\to\infty} A_t F = \int F\,dP\right\}$$

is contained in $\mathcal{J}$, and by the ergodic theorem, $P[A] = 1$ and $Q[A] = 0$. ■

Recall that an element x in a convex set C is called an extreme point of C if x can not be represented in a non-trivial way as a convex combination of elements in C. The set C_e of all extreme points in C is hence given by

$$C_e = \{x \in C : \nexists\, x_1, x_2 \in C\backslash\{x\}, \alpha \in (0,1) : x = \alpha x_1 + (1-\alpha)x_2\}\,.$$

Theorem 2.13 (Structure and extremals of $\mathcal{S}(\Theta)$). 1) The set $\mathcal{S}(\Theta)$ is convex.

2) A (Θ_t)-invariant probability measure P is extremal in $\mathcal{S}(\Theta)$ if and only if P is ergodic.

3) If Ω is a polish space and $\mathfrak{A}$ is the Borel σ-algebra then any (Θ_t)-invariant probability measure P on $(\Omega, \mathfrak{A})$ can be represented as a convex combination of extremal (ergodic) elements in $\mathcal{S}(\Theta)$, i.e., there exists a probability measure ϱ on $\mathcal{S}(\Theta)_e$ such that

$$P = \int_{\mathcal{S}(\Theta)_e} Q\,\varrho(dQ).$$

Proof. 1) If P_1 and P_2 are (Θ_t)-invariant probability measures then any convex combination $\alpha P_1 + (1-\alpha)P_2$, $\alpha \in [0,1]$, is (Θ_t)-invariant, too.

2) Suppose first that $P \in \mathcal{S}(\Theta)$ is ergodic and $P = \alpha P_1 + (1-\alpha)P_2$ for some $\alpha \in (0,1)$ and $P_1, P_2 \in \mathcal{S}(\Theta)$. Then P_1 and P_2 are both absolutely continuous w.r.t. P. Hence P_1 and P_2 are ergodic, i.e., they only take the values 0 and 1 on sets in $\mathcal{J}$. Since distinct ergodic measures are singular by Lemma 2.12 we can conclude that $P_1 = P = P_2$, i.e., the convex combination is trivial. This shows $P \in \mathcal{S}(\Theta)_e$.
Conversely, suppose that $P \in \mathcal{S}(\Theta)$ is not ergodic, and let $A \in \mathcal{J}$ such that $P[A] \in (0,1)$. Then P can be represented as a non-trivial combination by conditioning on $\sigma(A)$:

$$P = P[\,\cdot\,|A]\,P[A] + P[\,\cdot\,|A^c]\,P[A^c].$$

As A is in $\mathcal{J}$, the conditional distributions $P[\,\cdot\,|A]$ and $P[\,\cdot\,|A^c]$ are both (Θ_t)-invariant again. Hence $P \notin \mathcal{S}(\Theta)_e$.

3) This part is a bit tricky, and we only sketch the main idea. For more details see e.g. Varadhan, "Probability Theory" [53]. Since $(\Omega, \mathfrak{A})$ is a polish space with Borel σ-algebra, there is a regular version $p_{\mathcal{J}}(\omega,\cdot)$ of the conditional distributions $P[\,\cdot\,|\mathcal{J}](\omega)$ given the σ-algebra $\mathcal{J}$. Furthermore, it can be shown that $p_{\mathcal{J}}(\omega,\cdot)$ is **stationary** and **ergodic** for P-almost every $\omega \in \Omega$ (The idea in the background is that we "divide out" the non-trivial invariant events by conditioning on $\mathcal{J}$). Assuming the ergodicity of $p_{\mathcal{J}}(\omega,\cdot)$ for P-a.e. ω, we obtain the representation

$$\begin{aligned} P(d\omega) &= \int p_{\mathcal{J}}(\omega,\cdot)\,P(d\omega) \\ &= \int_{\mathcal{S}(\Theta)_e} Q\,\varrho(dQ) \end{aligned}$$

where ϱ is the law of $\omega \mapsto p_{\mathcal{J}}(\omega,\cdot)$ under P. Here we have used the definition of a regular version of the conditional distribution and the transformation theorem for Lebesgue integrals.
To prove ergodicity of $p_{\mathcal{J}}(\omega,\cdot)$ for almost every ω one can use that a measure is ergodic if and only if all limits of ergodic averages of indicator functions are almost surely constant. For a fixed event $A \in \mathfrak{A}$,

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t}\int_0^t 1_A \circ \Theta_s\,ds = P[A|\mathcal{J}] \quad P\text{-almost surely, and thus} \\ &\lim_{t\to\infty} \frac{1}{t}\int_0^t 1_A \circ \Theta_s\,ds = p_{\mathcal{J}}(\omega, A) \quad p_{\mathcal{J}}(\omega,\cdot)\text{-almost surely for } P\text{-a.e. } \omega. \end{aligned}$$

The problem is that the exceptional set in "P-almost every" depends on A, and there are uncountably many events $A \in \mathfrak{A}$ in general. To resolve this issue, one can use that the Borel σ-algebra on a Polish space is generated by countably many sets A_n. The convergence above then holds simultaneously with the same exceptional set for all A_n. This is enough to prove ergodicity of $p_{\mathcal{J}}(\omega,\cdot)$ for P-almost every ω. ■

The set of stationary distributions of a transition semigroup

We now specialize again to Markov processes. Let $p = (p_t)_{t\geq 0}$ be a transition semigroup on $(S, \mathcal{B})$, and let (X_t, P_x) be a corresponding canonical Markov process on $\Omega = \mathcal{D}(\mathbb{R}_+, S)$. We now denote by $\mathcal{S}(p)$ the collection of all stationary distributions for $(p_t)_{t\geq 0}$, i.e.,

$$\mathcal{S}(p) = \{\mu \in \mathcal{P}(S) : \mu = \mu p_t \text{ for any } t \geq 0\}\,.$$

As usually in this setup, $\mathcal{J}$ is the σ-algebra of events in $\mathfrak{A} = \sigma(X_t : t \geq 0)$ that are invariant under time-shifts $\Theta_t(\omega) = \omega(t + \cdot)$.

Exercise (Shift-invariants events for Markov processes). Show that for any $A \in \mathcal{J}$ there exists a Borel set $B \in \mathcal{B}$ such that $p_t 1_B = 1_B$ μ-almost surely for any $t \geq 0$, and

$$A = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} \{X_m \in B\} = \{X_0 \in B\} \quad P\text{-almost surely.}$$

The next result is an analogue to Theorem 2.13 for Markov processes. It can be either deduced from Theorem 2.13 or proven independently.

Theorem 2.14 (Structure and extremals of $\mathcal{S}$(p)). 1) The set $\mathcal{S}(p)$ is convex.

2) A stationary distribution μ of (p_t) is extremal in $\mathcal{S}(p)$ if and only if any set $B \in \mathcal{B}$ such that $p_t 1_B = 1_B$ μ-a.s. for any $t \geq 0$ has measure $\mu(B) \in \{0, 1\}$.

3) Any stationary distribution μ of (p_t) can be represented as a convex combination of extremal elements in $\mathcal{S}(p)$.

Remark (Phase transitions). The existence of several stationary distributions can correspond to the occurrence of a phase transition. For instance we will see in Section 5.4 below that for the heat bath dynamics of the Ising model on $\mathbb{Z}^d$ there is only one stationary distribution above the critical temperature but there are several stationary distributions in the phase transition regime below the critical temperature.

2.4. Central limit theorem for ergodic averages

Let $(p_t)_{t \geq 0}$ be the transition semigroup of a Markov process $((X_t)_{t \in \mathbb{Z}_+}, P_x)$ in discrete time or a right-continuous Markov process $((X_t)_{t \in \mathbb{R}_+}, P_x)$ in continuous time with state space $(S, \mathcal{B})$. In discrete time, $p_t = \pi^t$ where π is the one-step transition kernel. Suppose that μ is a stationary distribution of $(p_t)_{t \geq 0}$. If ergodicity holds then by the ergodic theorem, for any $f \in \mathcal{L}^1(\mu)$, the averages

$$A_t f = \frac{1}{t} \sum_{i=0}^{t-1} f(X_i), \quad A_t f = \frac{1}{t} \int_0^t f(X_s)\, ds \quad \text{respectively,}$$

converge to $\mu(f) = \int f d\mu$ P_μ-almost surely and in $L^2(P_\mu)$. In this section, we study the asymptotics of the fluctuations of $A_t f$ around $\mu(f)$ as $t \to \infty$ for $f \in \mathcal{L}^2(\mu)$.

Bias and variance of stationary ergodic averages

Theorem 2.15 (Bias, variance and asymptotic variance of ergodic averages). Let $f \in \mathcal{L}^2(\mu)$ and let $f_0 = f - \mu(f)$. The following statements hold:

1) For any $t > 0$, $A_t f$ is an unbiased estimator for $\mu(f)$ w.r.t. P_μ, i.e.,

$$E_{P_\mu}[A_t f] \;=\; \mu(f).$$

2) The variance of $A_t f$ in stationarity is given by

$$\mathrm{Var}_{P_\mu}[A_t f] \;=\; \frac{1}{t} \mathrm{Var}_\mu(f) + \frac{2}{t} \sum_{k=1}^{t} \left(1 - \frac{k}{t}\right) \mathrm{Cov}_\mu(f, \pi^k f) \quad \text{in discrete time,}$$

$$\mathrm{Var}_{P_\mu}[A_t f] \;=\; \frac{2}{t} \int_0^t \left(1 - \frac{r}{t}\right) \mathrm{Cov}_\mu(f, p_r f) dr \quad \text{in continuous time, respectively.}$$

3) Suppose that the series $Gf_0 = \sum_{k=0}^{\infty} \pi^k f_0$ or the integral $Gf_0 = \int_0^\infty p_s f_0 \, ds$ (in discrete/ continuous time respectively) converges in $\mathcal{L}^2(\mu)$. Then the asymptotic variance of $\sqrt{t} A_t f$ is given by

$$\lim_{t\to\infty} t \cdot \mathrm{Var}_{P_\mu}[A_t f] \;=\; \sigma_f^2, \qquad \text{where}$$

$$\sigma_f^2 \;=\; \mathrm{Var}_\mu(f) + 2\sum_{k=1}^{\infty} \mathrm{Cov}_\mu(f, \pi^k f) \;=\; 2(f_0, Gf_0)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)}$$

in the discrete time case, and

$$\sigma_f^2 \;=\; 2\int_0^\infty \mathrm{Cov}_\mu(f, p_s f) ds = 2(f_0, Gf_0)_{L^2(\mu)}$$

in the continuous time case, respectively.

Remark. 1) The asymptotic variance equals

$$\sigma_f^2 \;=\; \mathrm{Var}_{P_\mu}[f(X_0)] + 2\sum_{k=1}^{\infty} \mathrm{Cov}_{P_\mu}[f(X_0), f(X_k)],$$

$$\sigma_f^2 \;=\; 2\int_0^\infty \mathrm{Cov}_{P_\mu}[f(X_0), f(X_s)] ds, \quad \text{respectively.}$$

If Gf_0 exists then the variance of the ergodic averages behaves asymptotically as σ_f^2/t.

2) The statements hold under the assumption that the Markov process is started in stationarity. Bounds for ergodic averages of Markov processes with non-stationary initial distribution are given in Section 3.2 below.

Proof (of Theorem 2.15.). We prove the results in the continuous time case. The analogue discrete time case is left as an exercise. Note first that by right-continuity of $(X_t)_{t\geq 0}$, the process $(s, \omega) \mapsto f(X_s(\omega))$ is product measurable and square integrable on $[0, t] \times \Omega$ w.r.t. $\lambda \otimes P_\mu$ for any $t \in \mathbb{R}_+$.

1) By Fubini's theorem and stationarity,

$$E_{P_\mu}\left[\frac{1}{t}\int_0^t f(X_s)\, ds\right] = \frac{1}{t}\int_0^t E_{P_\mu}[f(X_s)]\, ds = \mu(f) \qquad \text{for any } t > 0.$$

2) Similarly, by Fubini's theorem, stationarity and the Markov property,

$$\begin{aligned}
\mathrm{Var}_{P_\mu}[A_t f] &= \mathrm{Cov}_{P_\mu}\left[\frac{1}{t}\int_0^t f(X_s)\, ds, \frac{1}{t}\int_0^t f(X_u)\, du\right] \\
&= \frac{2}{t^2}\int_0^t \int_0^u \mathrm{Cov}_{P_\mu}\left[f(X_s), f(X_u)\right]\, ds\, du \\
&= \frac{2}{t^2}\int_0^t \int_0^u \mathrm{Cov}_\mu(f, p_{u-s} f)\, ds\, du \\
&= \frac{2}{t^2}\int_0^t (t-r)\,\mathrm{Cov}_\mu(f, p_r f)\, dr.
\end{aligned}$$

3) Note that by stationarity, $\mu(p_r f) = \mu(f)$, and hence

$$\mathrm{Cov}_\mu(f, p_r f) = \int f_0\, p_r f_0\, d\mu \qquad \text{for any } r \geq 0.$$

Therefore, by 2) and Fubini's theorem,

$$\begin{aligned} t \cdot \mathrm{Var}_{P_\mu}[A_t f] &= 2\int_0^t \left(1-\frac{r}{t}\right)\int f_0\, p_r f_0\, d\mu\, dr \\ &= 2\left(f_0, \int_0^t \left(1-\frac{r}{t}\right) p_r f_0\, dr\right)_{L^2(\mu)} \\ &\to 2\left(f_0, \int_0^\infty p_r f_0\, dr\right)_{L^2(\mu)} \qquad \text{as } t\to\infty \end{aligned}$$

provided the integral $\int_0^\infty p_r f_0\, dr$ converges in $L^2(\mu)$. Here the last conclusion holds since $L^2(\mu)$-convergence of $\int_0^t p_r f_0\, dr$ as $t \to \infty$ implies that

$$\int_0^t \frac{r}{t} p_r f_0\, dr = \frac{1}{t}\int_0^t \int_0^r p_r f_0\, ds dr = \frac{1}{t}\int_0^t \int_s^t p_r f_0\, dr ds \to 0 \text{ in } L^2(\mu) \text{ as } t\to\infty. \qquad \blacksquare$$

Remark (Potential operator, existence of asymptotic variance). The theorem states that the asymptotic variance of $\sqrt{t}A_t f$ exists if the series/integral Gf_0 converges in $L^2(\mu)$. Notice that G is a linear operator that is defined in the same way as the Green's function. However, the Markov process is recurrent due to stationarity, and therefore $G1_B = \infty$ μ-a.s. on B for any Borel set $B \subseteq S$. Nevertheless, Gf_0 often exists because f_0 has mean $\mu(f_0) = 0$. Some sufficient conditions for the existence of Gf_0 (and hence of the asymptotic variance) are given below. If Gf_0 exists for any $f \in \mathcal{L}^2(\mu)$ then G induces a linear operator on the Hilbert space

$$L_0^2(\mu) \;=\; \{f \in L^2(\mu) : \mu(f) = 0\},$$

i.e., on the orthogonal complement of the constant functions in $L^2(\mu)$. This linear operator is called the **potential operator**. It is the inverse of the negative generator restricted to the orthogonal complement of the constant functions. Indeed, in discrete time,

$$-\mathcal{L}Gf_0 \;=\; (I-\pi)\sum_{n=0}^\infty \pi^n f_0 \;=\; f_0$$

whenever Gf_0 converges. Similarly, in continuous time, if Gf_0 exists then

$$\begin{aligned} -\mathcal{L}Gf_0 \;&=\; -\lim_{h\downarrow 0}\frac{p_h - I}{h}\int_0^\infty p_t f_0\, dt \;=\; \lim_{h\downarrow 0}\frac{1}{h}\left(\int_0^\infty p_t f_0\, dt - \int_0^\infty p_{t+h} f_0\, dt\right) \\ &=\; \lim_{h\downarrow 0}\frac{1}{h}\int_0^h p_t f_0\, dt \;=\; f_0. \end{aligned}$$

The last conclusion holds by strong continuity of $t \mapsto p_t f_0$, cf. Theorem 4.4 below.

Exercise (Sufficient conditions for existence of the asymptotic variance). Prove that in the continuous time case, $Gf_0 = \int_0^\infty p_t f_0\, dt$ converges in $L^2(\mu)$ if one of the following conditions is satisfied:

(i) *Decay of correlations:* $\int_0^\infty \left|\mathrm{Cov}_{P_\mu}[f(X_0), f(X_t)]\right| dt < \infty$.

(ii) *L^2 bound:* $\int_0^\infty \|p_t f_0\|_{L^2(\mu)} dt < \infty$.

Deduce non-asymptotic (t finite) and asymptotic ($t \to \infty$) bounds for the variances of ergodic averages under the assumption that either the correlations $|\,\mathrm{Cov}_{P_\mu}[f(X_0), f(X_t)]|$ or the $L^2(\mu)$ norms $\|p_t f_0\|_{L^2(\mu)}$ are bounded by an integrable function $r(t)$.

We now restrict ourselves to the *discrete time case*. In order to apply Theorem 2.15 in practice, quantitative bounds for the asymptotic variances are required. One possibility for deriving such bounds is to estimate the contraction coefficients of the transition kernels on the orthogonal complement

$$L^2_0(\mu) \;=\; \{f \in L^2(\mu) : \mu(f) = 0\}$$

of the constants in the Hilbert space $L^2(\mu)$. Indeed, note that by invariance of μ, π is a linear contraction that maps functions in $L^2_0(\mu)$ to $L^2_0(\mu)$. Let $\gamma(\pi) = \|\pi\|_{L^2_0(\mu)\to L^2_0(\mu)}$ denote the operator norm of π on $L^2_0(\mu)$. Recall that

$$\gamma(\pi) \;=\; \sup_{f\perp 1} \frac{\|\pi f\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}} \;=\; \sup_{f\perp 1} \frac{(f,\pi^*\pi f)^{1/2}_{L^2(\mu)}}{(f,f)^{1/2}_{L^2(\mu)}} \;=\; \varrho(\pi^*\pi|_{L^2_0(\mu)})^{1/2},$$

i.e., $\gamma(\pi)$ is the square root of the *spectral radius* of the linear operator $\pi^*\pi$ restricted to $L^2_0(\mu)$.

Theorem 2.16 (Upper bound for asymptotic variances). Suppose that $\gamma(\pi^t) < 1$ for some $t \in \mathbb{N}$. Then the series $Gf_0 = \sum_{n=0}^\infty \pi^n f_0$ converges for every $f \in \mathcal{L}^2(\mu)$, and the asymptotic variances are bounded by

$$\sigma^2_f \;\le\; (2c-1)\,\mathrm{Var}_\mu(f)\,, \qquad \text{where} \qquad c \;:=\; \sum_{n=0}^\infty \gamma(\pi^n) \;\le\; \frac{t}{1-\gamma(\pi^t)}. \tag{2.23}$$

Proof. By multiplicativity of the operator norm, we obtain

$$c \;=\; \sum_{i=0}^{t-1}\sum_{n=0}^\infty \gamma(\pi^{nt+i}) \;\le\; t\sum_{n=0}^\infty \gamma(\pi^t)^n \;=\; \frac{t}{1-\gamma(\pi^t)} \;<\; \infty. \tag{2.24}$$

Hence, the series Gf_0 is absolutely convergent, and

$$\sigma^2_f \;=\; 2(f_0,Gf_0)_{L^2(\mu)} - (f_0,f_0)_{L^2(\mu)} \;\le\; (2c-1)\|f_0\|^2_{L^2(\mu)} \;=\; (2c-1)\,\mathrm{Var}_\mu(f). \tag{2.25}$$

The claim follows from (2.24) and (2.25). ■

Remark (Relationship to spectral gap in the reversible case). Suppose that π satisfies the detailed balance condition w.r.t. μ, i.e.,

$$\mu(dx)\,\pi(x,dy) \;=\; \mu(dy)\,\pi(y,dx).$$

Then π is a self-adjoint linear operator on the Hilbert space $L^2_0(\mu)$. Therefore,

$$\gamma(\pi) \;=\; \varrho\left(\pi^2|_{L^2_0(\mu)}\right)^{1/2} \;=\; \sup\{|\lambda| : \lambda \in \mathrm{spec}(\pi)\} \;=\; \sup_{f\perp 1}\frac{\left|(f,\pi f)_{L^2(\mu)}\right|}{(f,f)_{L^2(\mu)}}.$$

The difference $1-\gamma(\pi)$ is called the **absolute spectral gap**, since it measures the gap between the largest eigenvalue 1 (corresponding to the constant functions) and the moduli of the other elements in the spectrum of π. If, additionally, the self-adjoint linear operator π is non-negative (i.e., its spectrum is contained in $\mathbb{R}_+$), then $1-\gamma(\pi)$ coincides with the **spectral gap** of the generator $\mathcal{L} = \pi - I$, which is defined by

$$\mathrm{Gap}(\mathcal{L}) \;=\; \inf\mathrm{spec}\left(-\mathcal{L}|_{L^2_0(\mu)}\right) \;=\; \inf_{f\perp 1}\frac{(f,-\mathcal{L}f)_{L^2(\mu)}}{(f,f)_{L^2(\mu)}}.$$

$\mathrm{Gap}(\mathcal{L})$ is the gap in the spectrum of $-\mathcal{L}$ between the eigenvalue 0 corresponding to the constant functions, and the infimum of the spectrum on the complement of the constants. By (2.23),

$$\sigma_f^2 \;\le\; \left(2\mathrm{Gap}(\mathcal{L})^{-1} - 1\right)\,\mathrm{Var}_\mu(f) \qquad \text{for all } f \in \mathcal{L}^2(\mu). \tag{2.26}$$

This bound is sharp since G is the inverse of the operator $-\mathcal{L}$ on $L_0^2(\mu)$. If π is non-negative then the spectra of π and $\mathcal{L}$ are contained in $[0,1]$, and thus the best possible bound occurs in (2.26) in the i.i.d. case where $\pi(x,dy) = \mu(dy)$ and $\mathrm{Gap}(\mathcal{L}) = 1$.

If π is not non-negative, one can still apply a similar argument as above with π replaced by π^2. In contrast, in the non-reversible case, the asymptotic variances are not related to the spectral gap of π.

Exercise (Rotation on the discrete circle). Consider the Markov chain on $\mathbb{Z}_k = \mathbb{Z}/(k\mathbb{Z})$ with deterministic transition step $x \mapsto x+1$. Show that for every initial distribution and every function $f : \mathbb{Z}_k \to \mathbb{R}$, the variance of $A_t f$ is of order $O(1/t^2)$. Conclude that the asymptotic variance $\sigma_f^2 = \lim_{t\to\infty} t \cdot \mathrm{Var}[A_t f]$ vanishes.

Central limit theorem for Markov chains

We again restrict ourselves to the discrete time case. Corresponding results in continuous time will be given in Section 9.7. Let $f \in \mathcal{L}^2(\mu)$, and suppose that the asymptotic variance

$$\sigma_f^2 = \lim_{n\to\infty} n\,\mathrm{Var}_{P_\mu}[A_n f]$$

exists and is finite. Without loss of generality we assume $\mu(f) = 0$, otherwise we may consider f_0 instead of f. Our goal is to prove a central limit theorem of the form

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(X_i) \;\stackrel{\mathcal{D}}{\longrightarrow}\; N(0, \sigma_f^2) \tag{2.27}$$

where "$\stackrel{\mathcal{D}}{\rightarrow}$" stands for convergence in distribution. The key idea is to use the martingale problem in order to reduce (2.27) to a central limit theorem for martingales. If g is a function in $\mathcal{L}^2(\mu)$ then $g(X_n) \in \mathcal{L}^2(P_\mu)$ for any $n \ge 0$, and hence

$$g(X_n) - g(X_0) \;=\; M_n + \sum_{k=0}^{n-1} (\mathcal{L}g)(X_k) \tag{2.28}$$

where (M_n) is a square-integrable $(\mathcal{F}_n^X)$ martingale with $M_0 = 0$ w.r.t. P_μ, and $\mathcal{L}g = \pi g - g$. Now suppose that there exists a function $g \in \mathcal{L}^2(\mu)$ such that $\mathcal{L}g = -f$ μ-a.e. Note that this is always the case with $g = Gf$ if $Gf = \sum_{n=0}^\infty \pi^n f$ converges in $L^2(\mu)$. Then by (2.28),

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k) = \frac{M_n}{\sqrt{n}} + \frac{g(X_0) - g(X_n)}{\sqrt{n}}. \tag{2.29}$$

As $n \to \infty$, the second summand converges to 0 in $L^2(P_\mu)$. Therefore, (2.27) is equivalent to a central limit theorem for the martingale (M_n). Explicitly,

$$M_n \;=\; \sum_{i=1}^n Y_i \qquad \text{for any } n \ge 0,$$

where the martingale increments Y_i are given by

$$Y_i \;=\; M_i - M_{i-1} \;=\; g(X_i) - g(X_{i-1}) - (\mathcal{L}g)(X_{i-1}) \;=\; g(X_i) - (\pi g)(X_{i-1}).$$

These increments form a stationary sequence w.r.t. P_μ. Thus we can apply the following theorem:

Theorem 2.17 (CLT for martingales with stationary increments). Let $(\mathcal{F}_n)$ be a filtration on a probability space $(\Omega, \mathfrak{A}, P)$. Suppose that $M_n = \sum_{i=1}^{n} Y_i$ is an $(\mathcal{F}_n)$ martingale on $(\Omega, \mathfrak{A}, P)$ with stationary increments $Y_i \in \mathcal{L}^2(P)$, and let $\sigma \in \mathbb{R}_+$. If

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \;\to\; \sigma^2 \quad \text{in } L^1(P) \;\text{ as } n \to \infty \tag{2.30}$$

then

$$\frac{1}{\sqrt{n}} M_n \;\xrightarrow{\mathcal{D}}\; N(0, \sigma^2) \quad \text{w.r.t. } P. \tag{2.31}$$

The proof of Theorem 2.17 will be given below. Note that by the ergodic theorem, the condition (2.30) is satisfied with $\sigma^2 = E[Y_i^2]$ if the process (Y_i, P) is ergodic. As a consequence of Theorem 2.17 and the considerations above, we obtain:

Corollary 2.18 (CLT for stationary Markov chains). Let (X_n, P_μ) be a stationary and ergodic Markov chain with initial distribution μ and one-step transition kernel π, and let $f \in \mathcal{L}^2(\mu)$. Suppose that there exists a function $g \in \mathcal{L}^2(\mu)$ solving the Poisson equation

$$-\mathcal{L}g \;=\; f - \mu(f). \tag{2.32}$$

Let $\sigma_f^2 := 2\,\mathrm{Cov}_\mu(f, g) - \mathrm{Var}_\mu(f)$. Then as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} (f(X_k) - \mu(f)) \;\xrightarrow{\mathcal{D}}\; N(0, \sigma_f^2).$$

Remark. Recall that (2.32) is satisfied with $g = G(f - \mu(f))$ if it exists.

Proof. Let $Y_i = g(X_i) - (\pi g)(X_{i-1})$. Then under P_μ, (Y_i) is a stationary sequence of square-integrable martingale increments. By the ergodic theorem for the process (X_n, P_μ),

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \to E_\mu[Y_1^2] \quad \text{in } L^1(P_\mu) \text{ as } n \to \infty.$$

The limiting expectation can be identified as the asymptotic variance σ_f^2 by an explicit computation:

$$\begin{aligned}
E_\mu[Y_1^2] &= E_\mu[(g(X_1) - (\pi g)(X_0))^2] \\
&= \int \mu(dx) E_x[g(X_1)^2 - 2g(X_1)(\pi g)(X_0) + (\pi g)(X_0)^2] \\
&= \int (\pi g^2 - 2(\pi g)^2 + (\pi g)^2) d\mu = \int g^2\, d\mu - \int (\pi g)^2 d\mu \\
&= (g - \pi g, g + \pi g)_{L^2(\mu)} = 2(f_0, g)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)} = \sigma_f^2.
\end{aligned}$$

Here $f_0 := f - \mu(f) = -\mathcal{L}g = g - \pi g$ by assumption. The martingale CLT 2.17 now implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{\mathcal{D}} N(0, \sigma_f^2),$$

and hence

$$\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1}(f(X_i)-\mu(f)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i + \frac{g(X_0)-g(X_n)}{\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,\sigma_f^2)$$

as well, because $g(X_0) - g(X_n)$ is bounded in $L^2(P_\mu)$. ■

We conclude this section with a proof of the CLT for martingales with stationary increments:

Central limit theorem for martingales

Let $M_n = \sum_{i=1}^n Y_i$ where (Y_i) is a stationary sequence of square-integrable random variables on a probability space $(\Omega, \mathfrak{A}, P)$ satisfying

$$E[Y_i|\mathcal{F}_{i-1}] = 0 \quad P\text{-a.s. for any } i \in \mathbb{N} \tag{2.33}$$

w.r.t. a filtration $(\mathcal{F}_n)$. We now prove the central limit theorem 2.17, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 \to \sigma^2 \text{ in } L^1(P) \;\Rightarrow\; \frac{1}{\sqrt{n}}M_n \xrightarrow{\mathcal{D}} N(0,\sigma^2). \tag{2.34}$$

Proof (of Theorem 2.17). Since the characteristic function $\varphi(p) = \exp\left(-\sigma^2 p^2/2\right)$ of $N(0,\sigma^2)$ is continuous, it suffices to show that for any fixed $p \in \mathbb{R}$,

$$\begin{aligned} &E\left[e^{ipM_n/\sqrt{n}}\right] \to \varphi(p) \quad \text{as } n\to\infty, \text{ or, equivalently,} \\ &E\left[e^{ipM_n/\sqrt{n}+\sigma^2p^2/2} - 1\right] \to 0 \quad \text{as } n\to\infty. \end{aligned} \tag{2.35}$$

Let

$$Z_{n,k} := \exp\left(i\frac{p}{\sqrt{n}}M_k + \frac{\sigma^2p^2}{2}\frac{k}{n}\right), \quad k = 0,1,\dots,n.$$

Then the left-hand side in (2.35) is given by

$$\begin{aligned} E[Z_{n,n} - Z_{n,0}] &= \sum_{k=1}^{n} E[Z_{n,k} - Z_{n,k-1}] \\ &= \sum_{k=1}^{n} E\left[Z_{n,k-1}\cdot E\left[\exp\left(\frac{ip}{\sqrt{n}}Y_k + \frac{\sigma^2p^2}{2n}\right) - 1|\mathcal{F}_{k-1}\right]\right]. \end{aligned} \tag{2.36}$$

The random variables $Z_{n,k-1}$ are uniformly bounded independently of n and k, and by a Taylor approximation and (2.33),

$$\begin{aligned} E\left[\exp\left(\frac{ip}{\sqrt{n}}Y_k + \frac{\sigma^2p^2}{2n}\right) - 1|\mathcal{F}_{k-1}\right] &= E\left[\frac{ip}{\sqrt{n}}Y_k - \frac{p^2}{2n}\left(Y_k^2-\sigma^2\right)|\mathcal{F}_{k-1}\right] + R_{n,k} \\ &= -\frac{p^2}{2n}E[Y_k^2-\sigma^2|\mathcal{F}_{k-1}] + R_{n,k} \end{aligned}$$

with a remainder $R_{n,k}$ of order $o(1/n)$. Hence by (2.36),

$$E\left[e^{ipM_n/\sqrt{n}+\sigma^2p^2/2} - 1\right] = -\frac{p^2}{2n}\sum_{k=1}^{n} E\left[Z_{n,k-1}\cdot(Y_k^2-\sigma^2)\right] + r_n$$

where $r_n = \sum_{k=1}^{n} E[Z_{n,k-1}\, R_{n,k}]$. It can be verified that $r_n \to 0$ as $n \to \infty$, so we are only left with the first term. To control this term, we divide the positive integers into blocks of size l where $l \to \infty$ below, and we apply (2.30) after replacing $Z_{n,k-1}$ by $Z_{n,jl}$ on the j-th block. We first estimate

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n} E[Z_{n,k-1}(Y_k^2-\sigma^2)]\right| \\
&\leq \frac{1}{n}\sum_{j=0}^{\lfloor n/l\rfloor}\left(\left|E\left[Z_{n,jl}\sum_{\substack{jl<k\leq(j+1)l\\ k\leq n}}(Y_k^2-\sigma^2)\right]\right| + l\cdot\sup_{\substack{jl<k\leq(j+1)l\\ k\leq n}} E[|Z_{n,k-1}-Z_{n,jl}|\cdot|Y_k^2-\sigma^2|]\right) \\
&\leq c_1\cdot E\left[\left|\frac{1}{l}\sum_{k=1}^{l}(Y_k^2-\sigma^2)\right|\right] + c_2 \sup_{1\leq k\leq l} E\left[\left|Z_{n,k-1}-1\right|\cdot\left|Y_k^2-\sigma^2\right|\right]. \qquad (2.37)
\end{aligned}$$

with finite constants c_1, c_2. Here we have used that the random variables $Z_{n,k}$ are uniformly bounded,

$$|Z_{n,k-1}-Z_{n,jl}| \leq |Z_{n,jl}|\cdot\left|\exp\left(i\frac{p}{\sqrt{n}}(M_{k-1}-M_{jl}) + \frac{\sigma^2p^2}{2}\frac{k-jl}{n}\right)-1\right|,$$

and, by stationarity of the sequence (Y_k),

$$\sum_{jl<k\leq(j+1)l}(Y_k^2-\sigma^2) \sim \sum_{k=1}^{l}(Y_k^2-\sigma^2), \quad\text{and}\quad (M_{k-1}-M_{jl}, Y_k) \sim (M_{k-1-jl}, Y_{k-jl}).$$

By the assumption (2.30), the first term on the right-hand side of (2.37) can be made arbitrarily small by choosing l sufficiently large. Moreover, for any fixed $l \in \mathbb{N}$, the second summand converges to 0 as $n \to \infty$ by dominated convergence. Hence the left-hand side in (2.37) also converges to 0 as $n \to \infty$, and thus (2.30) holds. ■

3. Couplings and mixing times

So far, we have mainly studied the limits of ergodic averages for stationary processes, i.e., Markov processes that are started with an equilibrium distribution. In this chapter, we will derive general conditions that ensure for other initial distributions that after a certain time, the law of the Markov process is close to the one of a stationary process. This allows us to extend results in the previous chapter from stationary to arbitrary initial laws. We will also discuss different approaches to quantify the distance to stationarity.

3.1. Couplings and transportation metrics

Additional reference: Villani: Optimal transport – old and new [54].

Let S be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$. An invariant probability measure of a Markov kernel π on $(S, \mathcal{B})$ is a fixed point of the map $\mu \mapsto \mu\pi$ acting on an appropriate subspace of $\mathcal{P}(S)$. Therefore, one approach for studying convergence to equilibrium of Markov chains is to apply the Banach fixed point theorem and variants thereof. To obtain useful results in this way we need adequate metrics on probability measures.

Wasserstein distances

We fix a metric $d : S \times S \to [0, \infty)$ on the state space S that generates the topology on S. Some of the statements below also hold under the weaker assumption that d is lower semicontinuous, which applies for example to the degenerate metric $d(x, y) = 1_{x \neq y}$. For $p \in [1, \infty)$, the space of all probability measures on S with finite p-th moment is defined by

$$\mathcal{P}^p(S) = \left\{ \mu \in \mathcal{P}(S) : \int d(x_0, y)^p \mu(dy) < \infty \right\},$$

where x_0 is an arbitrary given point in S. Note that by the triangle inequality, the definition is indeed independent of x_0. A natural distance on $\mathcal{P}^p(S)$ can be defined via couplings.

Definition 3.1 (Coupling of probability measures). A *coupling* of measures $\mu, \nu \in \mathcal{P}(S)$ is a probability measure $\gamma \in \mathcal{P}(S \times S)$ with marginals μ and ν. The coupling γ is *realized* by random variables $X, Y : \Omega \to S$ defined on a common probability space $(\Omega, \mathcal{A}, P)$ such that $(X, Y) \sim \gamma$.

We denote the set of all couplings of given probability measures μ and ν by $\Pi(\mu, \nu)$.

Definition 3.2 (Wasserstein distance, Kantorovich distance). For $p \in [1, \infty)$, the L^p *Wasserstein distance* of probability measures $\mu, \nu \in \mathcal{P}(S)$ is defined by

$$\mathcal{W}^p(\mu, \nu) \;=\; \inf_{\gamma \in \Pi(\mu, \nu)} \left(\int d(x, y)^p \gamma(dx dy) \right)^{1/p} \;=\; \inf_{\substack{X \sim \mu \\ Y \sim \nu}} E\left[d(X, Y)^p \right]^{1/p}, \tag{3.1}$$

where the second infimum is over all random variables X, Y defined on a common probability space with laws μ and ν. The *Kantorovich distance* of μ and ν is the L^1 Wasserstein distance $\mathcal{W}^1(\mu, \nu)$.

Remark (Optimal transport). The minimization in (3.1) is a particular case of an optimal transportation problem. Given a cost function $c : S \times S \longrightarrow [0, \infty]$, one is either looking for a map $T : S \longrightarrow S$ minimizing the average cost $\int c(x, T(x))\, \mu(dx)$ under the constraint $\nu = \mu \circ T^{-1}$ (*Monge problem*), or, less restrictively, for a coupling $\gamma \in \Pi(\mu, \nu)$ minimizing $\int c(x, y)\, \gamma(dxdy)$ (*Kantorovich problem*).

Note that the definition of the $\mathcal{W}^p$ distance depends in an essential way on the metric d considered on S. In particular, we can create different distances on probability measures by modifying the underlying metric. For example, if $f : [0, \infty) \longrightarrow [0, \infty)$ is increasing and *concave* with $f(0) = 0$ and $f(r) > 0$ for any $r > 0$ then $f \circ d$ is again a metric, and we can consider the corresponding Kantorovich distance

$$\mathcal{W}_f(\mu, \nu) = \inf_{\substack{X \sim \mu \\ Y \sim \nu}} E\left[f(d(X, Y))\right].$$

The distances $\mathcal{W}_f$ obtained in this way are in some sense converse to $\mathcal{W}^p$ distances for $p > 1$ which are obtained by applying the convex function $r \mapsto r^p$ to $d(x, y)$.

Example (Couplings and Wasserstein distances for probability measures on $\mathbb{R}^1$). Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with distribution functions F_μ and F_ν, and let

$$F_\mu^{-1}(u) = \inf\{c \in \mathbb{R} : F_\mu(c) \geq u\}, \quad u \in (0, 1),$$

denote the *left-continuous generalized inverse* of the distribution function. If $U \sim \text{Unif}(0, 1)$ then $F_\mu^{-1}(U)$ is a random variable with law μ. This can be used to determine optimal couplings of μ and ν for Wasserstein distances based on the Euclidean metric $d(x, y) = |x - y|$ explicitly:

(i) **Coupling by monotone rearrangement.** A straightforward coupling of μ and ν is given by

$$X = F_\mu^{-1}(U) \text{ and } Y = F_\nu^{-1}(U), \quad \text{where } U \sim \text{Unif}(0, 1).$$

This coupling is a monotone rearrangement, i.e., it couples the lower lying parts of the mass of μ with the lower lying parts of the mass of ν. If F_μ and F_ν are both one-to-one, then it maps u-quantiles of μ to u-quantiles of ν. It can be shown that the coupling is *optimal w.r.t. the* $\mathcal{W}^{\mathbf{p}}$ *distance* for every $p \geq 1$, i.e.,

$$\mathcal{W}^p(\mu, \nu) \;=\; E\left[|X - Y|^p\right]^{1/p} \;=\; \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p(0,1)},$$

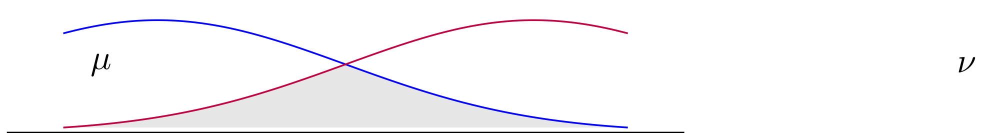
see e.g. Rachev&Rueschendorf [44]. On the other hand, the coupling by monotone rearrangement is *not optimal w.r.t.* $\mathcal{W}_f$ *if* f *is strictly concave*. Indeed, consider for example $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ and $\nu = \frac{1}{2}(\delta_0 + \delta_{-1})$. Then the coupling above satisfies $X \sim \mu$ and $Y = X - 1$, hence

$$E[f(|X - Y|)] = f(1).$$

On the other hand, we may couple by antimonotone rearrangement choosing $X \sim \mu$ and $\widetilde{Y} = -X$. In this case the average distance is smaller, since by Jensen's inequality,

$$E[f(|X - \widetilde{Y}|)] \;=\; E[f(2X)] \;<\; f(E[2X]) \;=\; f(1).$$

(ii) **Maximal coupling with antimonotone rearrangement** We now give a coupling that is optimal w.r.t. $\mathcal{W}_f$ for every concave function f provided an additional condition is satisfied. The idea is to keep the common mass of μ and ν in place, and to apply an antimonotone rearrangement to the remaining mass.



Suppose that $S = S^+ \dot\cup S^-$ and $\mu - \nu = (\mu-\nu)^+ - (\mu-\nu)^-$ is a Hahn-Jordan decomposition of the finite signed measure $\mu - \nu$ into a difference of positive measures such that $(\mu-\nu)^+(A\cap S^-) = 0$ and $(\mu-\nu)^-(A\cap S^+) = 0$ for any $A \in \mathcal{B}$, cf. also Section 3.3 below. Let

$$\mu \wedge \nu \;=\; \mu - (\mu-\nu)^+ \;=\; \nu - (\mu-\nu)^-.$$

If $p = (\mu\wedge\nu)(S)$ is the total shared mass of the measures μ and ν, then we can write μ and ν as mixtures

$$\begin{aligned} \mu &= (\mu\wedge\nu) + (\mu-\nu)^+ \;=\; p\alpha + (1-p)\beta, \\ \nu &= (\mu\wedge\nu) + (\mu-\nu)^- \;=\; p\alpha + (1-p)\delta \end{aligned}$$

of probability measures α, β and δ. Hence a coupling (X,Y) of μ and ν as described above is given by setting

$$(X,Y) = \begin{cases} \left(F_\alpha^{-1}(U), F_\alpha^{-1}(U)\right) & \text{if } B = 1, \\ \left(F_\beta^{-1}(U), F_\delta^{-1}(1-U)\right) & \text{if } B = 0, \end{cases}$$

with independent random variables $B \sim$Bernoulli(p) and $U \sim$Unif$(0,1)$. It can be shown that if S^+ and S^- are intervals then (X,Y) is an optimal coupling w.r.t. $\mathcal{W}_f$ for every concave function f, see McCann [39].

In contrast to the one-dimensional case, it is not easy to describe optimal couplings on $\mathbb{R}^d$ for $d > 1$ explicitly. On the other hand, the existence of optimal couplings holds on an arbitrary Polish space S by Prokhorov's Theorem.

Theorem 3.3 (Existence of optimal couplings). For any $\mu, \nu \in \mathcal{P}(S)$ and any $p \in [1,\infty)$ there exists a coupling $\gamma \in \Pi(\mu,\nu)$ such that

$$\mathcal{W}^p(\mu,\nu)^p = \int d(x,y)^p \gamma(dxdy).$$

Proof. Let $I(\gamma) := \int d(x,y)^p \gamma(dxdy)$. By definition of $\mathcal{W}^p(\mu,\nu)$, there exists a minimizing sequence (γ_n) of couplings in $\Pi(\mu,\nu)$ such that

$$I(\gamma_n) \to \mathcal{W}^p(\mu,\nu)^p \quad \text{as } n\to\infty.$$

Moreover, such a sequence is automatically tight in $\mathcal{P}(S\times S)$. Indeed, let $\varepsilon > 0$ be given. Then, since S is a Polish space, there exists a compact set $K \subset S$ such that

$$\mu(S\setminus K) < \varepsilon/2 \quad \text{and} \quad \nu(S\setminus K) < \varepsilon/2,$$

and hence for every $n \in \mathbb{N}$,

$$\gamma_n\left((x,y)\notin K\times K\right) \;\le\; \gamma_n(x\notin K) + \gamma_n(y\notin K) \;=\; \mu(S\setminus K) + \nu(S\setminus K) \;<\; \varepsilon.$$

Prokhorov's Theorem now implies that there is a subsequence (γ_{n_k}) that converges weakly to a limit $\gamma \in \mathcal{P}(S\times S)$. It is straightforward to verify that γ is again a coupling of μ and ν, and, since $d(x,y)^p$ is (lower semi-)continuous,

$$I(\gamma) \;=\; \int d(x,y)^p \gamma(dxdy) \;\le\; \liminf_{k\to\infty} \int d(x,y)^p \gamma_{n_k}(dxdy) \;=\; \mathcal{W}^p(\mu,\nu)^p$$

by the Portemanteau Theorem. ■

The proof shows that the statement of Theorem 3.3 still holds if the metric d does not generate the topology on S, provided $d : S \times S \to \mathbb{R}$ is a lower semi-continuous function.

Lemma 3.4 (Triangle inequality). *$\mathcal{W}^p$ is a metric on $\mathcal{P}^p(S)$.*

Proof. Let $\mu, \nu, \varrho \in \mathcal{P}^p(S)$. We prove the triangle inequality

$$\mathcal{W}^p(\mu, \varrho) \ \leq \ \mathcal{W}^p(\mu, \nu) + \mathcal{W}^p(\nu, \varrho). \tag{3.2}$$

The other properties of a metric can be verified easily. To prove (3.2) let γ and $\widetilde{\gamma}$ be couplings of μ and ν, ν and ϱ, respectively. We show

$$\mathcal{W}^p(\mu, \varrho) \ \leq \ \left(\int d(x,y)^p \gamma(dxdy) \right)^{1/p} + \left(\int d(y,z)^p \widetilde{\gamma}(dydz) \right)^{1/p}. \tag{3.3}$$

The claim then follows by taking the infimum over all $\gamma \in \Pi(\mu, \nu)$ and $\widetilde{\gamma} \in \Pi(\nu, \varrho)$.

Since S is a Polish space we can disintegrate

$$\gamma(dxdy) = \mu(dx)p(x, dy) \ \text{ and } \ \widetilde{\gamma}(dydz) = \nu(dy)\widetilde{p}(y, dz),$$

where p and $\widetilde{p}$ are regular versions of conditional distributions of the second components w.r.t. $\gamma, \widetilde{\gamma}$ given the first components. The disintegration enables us to "glue" the couplings γ and $\widetilde{\gamma}$ to a joint coupling

$$\hat{\gamma}(dxdydz) \ := \ \mu(dx)p(x, dy)\widetilde{p}(y, dz)$$

of the measures μ, ν and ϱ such that under $\hat{\gamma}$, $(x, y) \sim \gamma$ and $(y, z) \sim \widetilde{\gamma}$. Therefore, by the triangle inequality for the L^p norm, we obtain

$$\begin{aligned} \mathcal{W}^p(\mu, \varrho) \ &\leq \ \left(\int d(x,z)^p \hat{\gamma}(dxdydz) \right)^{1/p} \ \leq \ \left(\int d(x,y)^p \hat{\gamma}(dxdydz) \right)^{1/p} + \left(\int d(y,z)^p \hat{\gamma}(dxdydz) \right)^{1/p} \\ &= \ \left(\int d(x,y)^p \gamma(dxdy) \right)^{1/p} + \left(\int d(y,z)^p \widetilde{\gamma}(dydz) \right)^{1/p}. \end{aligned}$$

■

Exercise (Couplings on $\mathbb{R}^d$). Let $W : \Omega \to \mathbb{R}^d$ be a random variable on $(\Omega, \mathcal{A}, P)$, and let μ_a denote the law of $a + W$.

a) *Synchronous coupling:* Let $X = a + W$ and $Y = b + W$ for $a, b \in \mathbb{R}^d$. Show that

$$\mathcal{W}^2(\mu_a, \mu_b) \ = \ |a - b| \ = \ E(|X - Y|^2)^{1/2},$$

i.e., (X, Y) is an optimal coupling w.r.t. $\mathcal{W}^2$.

b) *Reflection coupling:* Assume that the law of W is a rotationally symmetric probability measure on $\mathbb{R}^d$. Let $\widetilde{Y} = \widetilde{W} + b$ where $\widetilde{W} = W - 2e \cdot W e$ with $e = \frac{a-b}{|a-b|}$. Prove that $(X, \widetilde{Y})$ is a coupling of μ_a and μ_b, and if $|W| \leq \frac{|a-b|}{2}$ almost surely, then

$$E\left[f(|X - \widetilde{Y}|\right] \ \leq \ f(|a - b|) \ = \ E\left[f(|X - Y|) \right]$$

for any concave, increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f(0) = 0$.

Kantorovich-Rubinstein duality

The *Lipschitz semi-norm* of a function $g : S \to \mathbb{R}$ is defined by

$$\|g\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|g(x) - g(y)|}{d(x, y)}.$$

Bounds in Wasserstein distances can be used to bound differences of integrals of Lipschitz continuous functions w.r.t. different probability measures. Indeed, one even has the following dual description of the L^1 Wasserstein distance.

Theorem 3.5 (Kantorovich-Rubinstein duality). For any $\mu, \nu \in \mathcal{P}^1(S)$,

$$\mathcal{W}^1(\mu, \nu) = \sup_{\|g\|_{\mathrm{Lip}} \leq 1} \left(\int g d\mu - \int g d\nu \right). \tag{3.4}$$

There is also a corresponding dual description of $\mathcal{W}^p$ for $p > 1$ but it takes a more complicated form, see for example Villani [54].

Proof. We only prove the easy "≥" part. For different proofs of the converse inequality see Rachev and Rueschendorf [44], Villani [55, 54], or Chen [8]. For instance, one can approximate μ and ν by finite convex combinations of Dirac measures for which (3.4) is a consequence of the standard duality principle of linear programming [8].

To prove "≥" let $\mu, \nu \in \mathcal{P}^1(S)$ and $g : S \to \mathbb{R}$ Lipschitz continuous. If γ is a coupling of μ and ν then

$$\int g \, d\mu - \int g \, d\nu = \int (g(x) - g(y)) \, \gamma(dxdy) \leq \|g\|_{\mathrm{Lip}} \int d(x, y) \, \gamma(dxdy).$$

Hence, by taking the infimum over $\gamma \in \Pi(\mu, \nu)$, we obtain $\int g \, d\mu - \int g \, d\nu \leq \|g\|_{\mathrm{Lip}} \mathcal{W}^1(\mu, \nu)$. ■

As a consequence of the "≥" part of (3.4), we see that if $(\mu_n)_{n \in \mathbb{N}}$ is a sequence of probability measures such that $\mathcal{W}^1(\mu_n, \mu) \to 0$ then $\int g d\mu_n \to \int g d\mu$ for every Lipschitz continuous function $g : S \to \mathbb{R}$. In particular, $\mu_n \to \mu$ weakly. Conversely, it is easy to see that weak convergence implies a bound on the Wasserstein distance of the limit measures.

Lemma 3.6. *Suppose that (μ_n) and (ν_n) are weakly convergent sequences of probability measures with limits μ and ν, respectively. Then*

$$\mathcal{W}^p(\mu, \nu) \leq \liminf_{n \to \infty} \mathcal{W}^p(\mu_n, \nu_n) \tag{3.5}$$

Proof. Without loss of generality, we assume that the sequence $\mathcal{W}^p(\mu_n, \nu_n)$ is convergent (otherwise, we can consider a subsequence that converges to the limit inferior). For each n, we choose an optimal coupling γ_n of μ_n and ν_n w.r.t. $\mathcal{W}^p$. Since the sequences (μ_n) and (ν_n) are weakly convergent and S is a Polish space, the family $\{\mu_n, \nu_n : n \in \mathbb{N}\}$ is tight. Thus for a given $\varepsilon > 0$, there is a compact set K such that for all n, $\mu_n(S \setminus K) < \varepsilon/2$, $\nu_n(S \setminus K) < \varepsilon/2$, and thus $\gamma_n((S \times S) \setminus (K \times K)) < \varepsilon$. Hence (γ_n) is a tight sequence of probability measures on $S \times S$, and thus by Prokhorov, there is a weakly convergent subsequence (γ_{n_k}). Since for every n, γ_n is a coupling of μ_n and ν_n, the limit γ is a coupling of μ and ν. Therefore, and by continuity of the distance function,

$$\mathcal{W}^p(\mu, \nu)^p \leq \int d(x, y)^p \, \gamma(dx\,dy) \leq \liminf_{k \to \infty} \int d(x, y)^p \, \gamma_{n_k}(dx\,dy) = \liminf_{k \to \infty} \mathcal{W}^p(\mu_{n_k}, \nu_{n_k}).$$

This proves the claim, since we have assumed that the sequence of Wasserstein distances on the right hand side is convergent. ■

The following general statement connects convergence in Wasserstein distances and weak convergence:

Theorem 3.7 ($\mathcal{W}^p$ convergence and weak convergence). Let $p \in [1, \infty)$.

1) The metric space $(\mathcal{P}^p(S), \mathcal{W}^p)$ is complete and separable.

2) A sequence (μ_n) in $\mathcal{P}^p(S)$ converges to a limit μ w.r.t. the $\mathcal{W}^p$ distance if and only if

$$\int g d\mu_n \to \int g d\mu \qquad \text{for any } g \in C(S) \text{ satisfying } g(x) \leq C \cdot (1 + d(x, x_o)^p)$$
$$\text{for a finite constant } C \text{ and some } x_0 \in S.$$

Among other things, the proof relies again on Prokhorov's Theorem. We refer to [54] for details.

Contraction coefficients

When studying convergence to equilibrium, we are particularly interested in contraction properties of transition kernels w.r.t. Wasserstein distances. Let $\pi(x, dy)$ be a Markov kernel on $(S, \mathcal{B})$, and fix $p \in [1, \infty)$. In applications, we will be mainly consider the case $p = 1$.

Definition 3.8 (Wasserstein contraction coefficient of a transition kernel). The global *contraction coefficient* of π w.r.t. the distance $\mathcal{W}^p$ is defined as

$$\alpha_p(\pi) = \sup \left\{ \frac{\mathcal{W}^p(\mu\pi, \nu\pi)}{\mathcal{W}^p(\mu, \nu)} : \mu, \nu \in \mathcal{P}^p(S) \text{ s.t. } \mu \neq \nu \right\}.$$

In other words, $\alpha_p(\pi)$ is the Lipschitz norm of the map $\mu \mapsto \mu\pi$ w.r.t. the $\mathcal{W}^p$ distance. By applying the Banach fixed point theorem, we obtain:

Theorem 3.9 (Geometric ergodicity for Wasserstein contractions). If $\alpha_p(\pi) < 1$ then there exists a unique invariant probability measure μ of π in $\mathcal{P}^p(S)$. Moreover, for every initial distribution $\nu \in \mathcal{P}^p(S)$, $\nu\pi^n$ converges to μ with a geometric rate:

$$\mathcal{W}^p(\nu\pi^n, \mu) \leq \alpha_p(\pi)^n \mathcal{W}^p(\nu, \mu).$$

Proof. The Banach fixed point theorem can be applied by Theorem 3.7. ■

The assumption $\alpha_p(\pi) < 1$ seems restrictive. However, one should bear in mind that the underlying metric on S can be chosen adequately. In particular, in applications it is often possible to find a concave function f such that $\mu \mapsto \mu\pi$ is a contraction w.r.t. the $\mathcal{W}^1$ distance based on the modified metric $f \circ d$. Furthermore, even if π is not contractive, it is often the case that $\alpha_p(\pi^k) < 1$ for k sufficiently large.

The next theorem is crucial for bounding $\alpha_p(\pi)$ in applications:

Theorem 3.10 (Bounds for contraction coefficients, Path coupling).

1) Suppose that the transition kernel $\pi(x, dy)$ is Feller. Then

$$\alpha_p(\pi) = \sup_{x \neq y} \frac{\mathcal{W}^p(\pi(x,\cdot), \pi(y,\cdot))}{d(x,y)}. \tag{3.6}$$

2) Moreover, suppose that S is a **geodesic graph** with edge set E in the sense that for any $x, y \in S$, there exists a path $x_0 = x, x_1, x_2, \ldots, x_{n-1}, x_n = y$ from x to y such that $\{x_{i-1}, x_i\} \in E$ for $i = 1, \ldots, n$ and $d(x,y) = \sum_{i=1}^n d(x_{i-1}, x_i)$. Then

$$\alpha_p(\pi) = \sup_{\{x,y\} \in E} \frac{\mathcal{W}^p(\pi(x,\cdot), \pi(y,\cdot))}{d(x,y)}. \tag{3.7}$$

The application of the second assertion of the lemma to prove upper bounds for $\alpha_p(\pi)$ is known as the **path coupling method** of Bubley and Dyer [6, 24].

Proof. 1) Let $\beta := \sup_{x \neq y} \frac{\mathcal{W}^p(\pi(x,\cdot),\pi(y,\cdot))}{d(x,y)}$. We have to show that

$$\mathcal{W}^p(\mu\pi, \nu\pi) \leq \beta \mathcal{W}^p(\mu, \nu) \tag{3.8}$$

holds for arbitrary probability measures $\mu, \nu \in \mathcal{P}(S)$. By definition of β and since $\mathcal{W}^p(\delta_x, \delta_y) = d(x,y)$, (3.8) is satisfied if μ and ν are Dirac measures. Next suppose that

$$\mu = \sum_{x \in C} \mu(x)\delta_x \quad \text{and} \quad \nu = \sum_{x \in C} \nu(x)\delta_y$$

are convex combinations of Dirac measures, where $C \subset S$ is a countable subset. Then for any $x, y \in C$, we can choose a coupling γ_{xy} of $\delta_x\pi$ and $\delta_y\pi$ such that

$$\left(\int d(x', y')^p \gamma_{xy}(dx'dy') \right)^{1/p} = \mathcal{W}^p(\delta_x\pi, \delta_y\pi) \leq \beta d(x,y). \tag{3.9}$$

Let $\xi(dxdy)$ be an arbitrary coupling of μ and ν. Then a coupling $\gamma(dx'dy')$ of $\mu\pi$ and $\nu\pi$ is given by

$$\gamma(B) := \int \gamma_{xy}(B)\,\xi(dxdy) \qquad \text{for any measurable set } B \subseteq S \times S.$$

Therefore, by (3.9),

$$\begin{aligned} \mathcal{W}^p(\mu\pi, \nu\pi) &\leq \left(\int d(x', y')^p \gamma(dx'dy') \right)^{1/p} \\ &= \left(\int \int d(x', y')^p \gamma_{xy}(dx'dy')\xi(dxdy) \right)^{1/p} \leq \beta \left(\int d(x,y)^p \xi(dxdy) \right)^{1/p}. \end{aligned}$$

By taking the infimum over all couplings $\xi \in \Pi(\mu, \nu)$, we see that μ and ν satisfy (3.8).

Finally, in order to show that (3.8) holds for arbitrary $\mu, \nu \in \mathcal{P}(S)$, note that since S is separable, there is a countable dense subset C, and the convex combinations of Dirac measures based in C are dense in $\mathcal{W}^p$. Hence μ and ν are $\mathcal{W}^p$ limits of corresponding convex combinations μ_n and ν_n ($n \in \mathbb{N}$). By the Feller property, the sequences $\mu_n\pi$ and $\nu_n\pi$ converge weakly to $\mu\pi$ and $\nu\pi$, respectively. Hence by Lemma 3.6,

$$\mathcal{W}^p(\mu\pi, \nu\pi) \leq \liminf \mathcal{W}^p(\mu_n\pi, \nu_n\pi) \leq \beta \liminf \mathcal{W}^p(\mu_n, \nu_n) = \beta \mathcal{W}^p(\mu, \nu).$$

2) Let $\widetilde{\beta} := \sup_{(x,y)\in E} \frac{\mathcal{W}^p(\pi(x,\cdot),\pi(y,\cdot))}{d(x,y)}$. We show that

$$\mathcal{W}^p(\pi(x,\cdot),\pi(y,\cdot)) \;\leq\; \widetilde{\beta} d(x,y)$$

holds for arbitrary $x, y \in S$. Indeed, let $x_0 = x, x_1, x_2, \ldots, x_n = y$ be a geodesic from x to y such that $\{x_{i-1}, x_i\} \in E$ for $i = 1, \ldots, n$. Then by the triangle inequality,

$$\mathcal{W}^p(\pi(x,\cdot),\pi(y,\cdot)) \;\leq\; \sum_{i=1}^{n} \mathcal{W}^p(\pi(x_{i-1},\cdot),\pi(x_i,\cdot)) \;\leq\; \widetilde{\beta} \sum_{i=1}^{n} d(x_{i-1},x_i) \;=\; \widetilde{\beta}\, d(x,y),$$

where we have used for the last equality that $x_0, \ldots, x_n$ is a geodesic. ■

The following exercise shows that the assumption of the Feller property in Theorem 3.10 can be avoided if the statement is modified slightly.

Exercise ($\mathcal{W}^1$ convergence by couplings of transition probabilities). Let $\overline{\pi}$ be a transition kernel on $S \times S$ such that $\overline{\pi}((x,y), dx'dy')$ is a coupling of $\pi(x, dx')$ and $\pi(y, dy')$ for every $x, y \in S$. Prove that if there exists a distance function $d : S \times S \to [0,\infty)$ and a constant $\alpha \in (0,1)$ such that $\overline{\pi} d \leq \alpha d$, then there is a unique invariant probability measure μ of π, and

$$\mathcal{W}^1_d(\nu\pi^n, \mu) \;\leq\; \alpha^n \mathcal{W}^1_d(\nu, \mu) \quad \text{for any } \nu \in \mathcal{P}^1(S).$$

3.2. Markov Chain Monte Carlo

Let μ be a probability measure on $(S, \mathcal{B})$. In Markov chain Monte Carlo methods one is approximating integrals $\mu(f) = \int f \, d\mu$ by ergodic averages of the form

$$A_{b,n} f = \frac{1}{n} \sum_{i=b}^{b+n-1} f(X_i),$$

where (X_n, P) is a time-homogeneous Markov chain with a transition kernel π satisfying $\mu = \mu\pi$, and $b, n \in \mathbb{N}$ are sufficiently large integers. Here the random variables X_i serve as approximate samples from μ. The constant b is called the **burn-in time** - it should be chosen in such a way that the law of the Markov chain after b steps is sufficiently close to the stationary distribution μ.

How can we construct transition kernels π such that μ is invariant for π ? A sufficient condition for invariance that is easy to fulfil is the **detailed balance condition**

$$\mu(dx)\, \pi(x, dy) \;=\; \mu(dy)\, \pi(y, dx). \tag{3.10}$$

It says that the measure $\mu \otimes \pi$ is invariant under the (time) reversal map $R(x, y) = (y, x)$. Clearly, if (3.10) holds, then by Fubini's Theorem,

$$(\mu\pi)(B) \;=\; \int \mu(dx)\, \pi(x, B) \;=\; \int\int 1_B(y)\, \mu(dx)\, \pi(x, dy) \;=\; \mu(B)$$

for any measurable set $B \subseteq S$.

There are several possibilities to construct Markov kernels for which a given probability measure μ satisfies the detailed balance condition. These lead to different types of Markov Chain Monte Carlo methods. Important classes of Markov chains used in MCMC methods are Metropolis-Hastings chains and Gibbs samplers.

Metropolis-Hastings methods

A simple way to construct a Markov chain satisfying detailed balance w.r.t. a given probability measure μ is to modify a given transition kernel $p(x, dy)$ by rejecting moves in such a way that (3.10) holds. Let λ be a positive reference measure on $(S, \mathcal{B})$, e.g. Lebesgue measure on $\mathbb{R}^d$ or the counting measure on a countable space. Suppose that μ is absolutely continuous w.r.t. λ, and denote the density by $\mu(x)$ as well. Then a Markov transition kernel π with invariant probability measure μ can be constructed by proposing moves according to an absolutely continuous proposal kernel

$$p(x, dy) = p(x, y)\,\lambda(dy)$$

with strictly positive density $p(x, y)$, and accepting a proposed move from x to y with probability

$$\alpha(x, y) = \min\left(1, \frac{\mu(y)p(y, x)}{\mu(x)p(x, y)}\right).$$

If a proposed move is not accepted then the Markov chain stays at its current position x. The transition kernel is hence given by

$$\pi(x, dy) = \alpha(x, y)p(x, dy) + r(x)\delta_x(dy)$$

where $r(x) = 1 - \int \alpha(x, y)p(x, dy)$ is the rejection probability for the next move from x. Typical examples of Metropolis-Hastings methods are Random Walk Metropolis algorithms where p is the transition kernel of a random walk. Note that if p is symmetric then the acceptance probability simplifies to

$$\alpha(x, y) = \min(1, \mu(y)/\mu(x)).$$

Lemma 3.11 (Detailed balance). *The transition kernel π of a Metropolis-Hastings chain satisfies the detailed balance condition*

$$\mu(dx)\pi(x, dy) = \mu(dy)\pi(y, dx). \tag{3.11}$$

In particular, μ is an invariant probability measure for π.

Proof. On $\{(x, y) \in S \times S : x \neq y\}$, the measure $\mu(dx)\pi(x, dy)$ is absolutely continuous w.r.t. $\lambda \otimes \lambda$ with density

$$\mu(x)\alpha(x, y)p(x, y) = \min\big(\mu(x)p(x, y), \mu(y)p(y, x)\big).$$

The detailed balance condition (3.11) follows, since this expression is a symmetric function of x and y. ■

A disadvantage of Metropolis-Hastings algorithms is that the rejection of proposal moves can slow down convergence. Therefore, it is important to choose a proposal kernel that is adjusted properly to the target distribution μ. In particular, this is crucial in high dimensional applications where the acceptance probabilities tend to degenerate.

Glauber dynamics, Gibbs sampler

On product spaces, a rejection-free alternative to Metropolis-Hastings methods is the Gibbs sampler. It assumes that one is able to draw samples from the conditional distribution of each coordinate given the other coordinates. If this is not possible, the Gibbs sampler can be combined with a Metropolis-Hastings algorithm targeting the conditional law.

Let μ be a probability measure on a product space

$$S = T^V = \{\eta : V \to T\}.$$

We assume that V is a finite set (for example a finite graph) and T is a Polish space (e.g. $T = \mathbb{R}^d$). Depending on the model considered, the elements in T are called types, states, spins, colours etc., whereas we call the elements of S configurations. There is a natural transition mechanism on S that leads to a Markov chain which is reversible w.r.t. μ. The transition step from a configuration $\xi \in S$ to the next configuration ξ' is given in the following way:

- Choose an element $x \in V$ uniformly at random.
- Set $\xi'(y) = \xi(y)$ for any $y \neq x$, and sample $\xi'(x)$ from the conditional law w.r.t. μ of $\eta(x)$ given that $\eta(y) = \xi(y)$ for all $y \neq x$.

To make this precise, we fix for every $x \in V$ a regular version $\pi_x(\xi, d\xi')$ of the conditional law $\mu(d\xi' \,|\, \xi' = \xi \text{ on } V \setminus \{x\})$, and we define the transition kernel π by

$$\pi \;=\; \frac{1}{|V|}\sum_{x\in V} \pi_x \,.$$

Definition 3.12 (Random scan Gibbs sampler). A time-homogeneous Markov chain with transition kernel π is called a *random scan Gibbs sampler* with stationary distribution μ.

Besides the random scan Gibbs sampler, there is also the *systematic scan Gibbs sampler*, where the transition kernels π_x, $x \in V$, are applied in a fixed deterministic order.

That μ is indeed invariant w.r.t. π_x and π is shown in the next lemma.

Lemma 3.13. *The transition kernels π_x $(x \in V)$ and π satisfy the detailed balance conditions*

$$\mu(d\xi)\pi_x(\xi, d\xi') = \mu(d\xi')\pi_x(\xi', d\xi),$$
$$\mu(d\xi)\pi(\xi, d\xi') = \mu(d\xi')\pi(\xi', d\xi).$$

In particular, μ is an invariant probability measure for π.

Proof. Let $x \in V$, and let $\hat\eta(x) := (\eta(y))_{y\neq x}$ denote the configuration restricted to $V \setminus \{x\}$. Disintegration of the measure μ into the law $\hat\mu_x$ of $\hat\eta(x)$ and the conditional law $\mu_x(\cdot|\hat\eta(x))$ of $\eta(x)$ given $\hat\eta(x)$ yields

$$\begin{aligned}
\mu\,(d\xi)\,\pi_x\,(\xi, d\xi') \;&=\; \hat\mu_x\left(d\hat\xi(x)\right)\mu_x\left(d\xi(x)\,\middle|\,\hat\xi(x)\right)\delta_{\hat\xi(x)}\left(d\hat\xi'(x)\right)\,\mu_x\left(d\xi'(x)\,\middle|\,\hat\xi(x)\right)\\
&=\; \hat\mu_x\left(d\hat\xi'(x)\right)\mu_x\left(d\xi(x)\,\middle|\,\hat\xi'(x)\right)\delta_{\hat\xi'(x)}\left(d\hat\xi(x)\right)\,\mu_x\left(d\xi'(x)\,\middle|\,\hat\xi'(x)\right)\\
&=\; \mu\,(d\xi')\,\pi_x\,(\xi', d\xi)\,.
\end{aligned}$$

Hence the detailed balance condition is satisfied with respect to π_x for all $x \in V$, and, by averaging over x, also with respect to π. ■

Exercise (Systematic scan Gibbs sampler). Suppose that $V = \{1, \ldots, d\}$, and consider the systematic Gibbs sampler which during every scan is updating all coordinates in sequential order. The total update during one scan is then given by the transition kernel $\pi = \pi_1\pi_2\cdots\pi_d$.

a) Show that μ is invariant for π but the detailed balance condition is not satisfied in general.

b) Show that the detailed balance condition is satisfied for the modified transition kernel $\widetilde{\pi} = \pi_1\pi_2\cdots\pi_{d-1}\pi_d\pi_d\pi_{d-1}\cdots\pi_2\pi_1$ which corresponds to updating the coordinates at first in forward and then in backward order.

Example (Some basic models from statistical physics). In the following examples we assume that V is the vertex set of a finite graph with edge set E.

1) **Random colourings.** Here T is a finite set (the set of possible colours of a vertex), and μ is the uniform distribution on all admissible colourings of the vertices in V such that no two neighbouring vertices have the same colour:

$$\mu = \mathrm{Unif}\left(\{\eta \in T^V : \eta(x) \neq \eta(y)\ \forall\{x, y\} \in E\}\right).$$

In each step, the Gibbs sampler selects a vertex at random and changes its colour randomly to one of the colours that are different from all colours of neighbouring vertices.

2) **Hard core model.** Here $T = \{0,1\}$ where $\eta(x) = 1$ stands for the presence of a particle at the vertex x. The hard core model with fugacity $\lambda \in \mathbb{R}_+$ is the probability measure μ_λ on $\{0,1\}^V$ satisfying

$$\mu_\lambda(\eta) \;=\; \tfrac{1}{Z_\lambda}\lambda^{\sum_{x\in V}\eta(x)} \qquad \text{if } \eta(x)\eta(y) = 0 \text{ for all } \{x,y\} \in E,$$

and $\mu_\lambda(\eta) = 0$ otherwise. Here Z_λ is a finite normalization constant. In each step, the Gibbs sampler updates $\xi(x)$ for a randomly chosen vertex x according to

$$\xi'(x) \sim \text{Bernoulli}\left(\frac{\lambda}{1+\lambda}\right) \qquad \text{if } \xi(y) = 0 \text{ for all } y \sim x,$$

and $\xi'(x) = 0$ otherwise.

3) **Ising model.** Here $T = \{-1,+1\}$ where -1 and $+1$ stand for spin directions. The ferromagnetic Ising model at inverse temperature $\beta > 0$ is given by

$$\mu_\beta(\eta) \;=\; \tfrac{1}{Z_\beta}\exp\left(-\beta H(\eta)\right) \qquad \text{for all } \eta \in \{-1,+1\}^V,$$

where Z_β is again a normalizing constant, and the Ising Hamiltonian H is given by

$$H(\eta) \;=\; \frac{1}{2}\sum_{\{x,y\}\in E}|\eta(x)-\eta(y)|^2 \;=\; |E| - \sum_{\{x,y\}\in E}\eta(x)\eta(y).$$

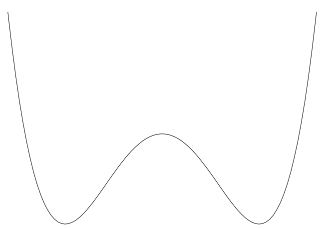
Thus μ_β favours configurations where neighbouring spins coincide, and this preference gets stronger as the temperature β^{-1} decreases. The heat bath dynamics (Gibbs sampler) updates a randomly chosen spin $\xi(x)$ to $\xi'(x)$ with probability proportional to $\exp\left(\beta\eta(x)\sum_{y\sim x}\eta(y)\right)$.

The **mean field Ising model** is the Ising model on the complete graph with n vertices, i.e., every spin is interacting with every other spin. In this case the update probability only depends on $\eta(x)$ and the "mean field" $\frac{1}{n}\sum_{y\in V}\eta(y)$. Mean field models are studied in more detail in Section 5.2.

4) **Continuous spin systems.** Here $T = \mathbb{R}$, and

$$\mu_\beta(dy) \;=\; \frac{1}{Z_\beta}\exp\left(-\frac{1}{2}\sum_{\{x,y\}\in E}|\eta(x)-\eta(y)|^2 + \beta\sum_{x\in V}U(\eta(x))\right)\prod_{x\in V}d\eta(x).$$

The function $U : \mathbb{R} \to [0,\infty)$ is a given potential, and Z_β is a normalizing constant. For $U \equiv 0$, the measure is called the *massless Gaussian free field* over V. Although the density is similar to the Ising model, the measure is very different, since the configuration space is $\mathbb{R}^V$ instead of $\{-1,+1\}^V$. In particular, phase transitions occur for Ising models, but not for the Gaussian free field which is a Gaussian measure. On the other hand, if U is a double-well potential, then μ_β is a continuous counterpart to the Ising model.



Example (Bayesian posterior distributions.). Gibbs samplers are applied frequently to sample from posterior distributions in Bayesian statistical models. Here is a typical example of a hierarchical Bayesian model, see Jones [25]. One assumes that the data are realizations of conditionally independent random variables Y_{ij} ($i = 1, \dots, k, j = 1, \dots, m_i$) with conditional laws

$$Y_{ij}|(\theta_1, \dots, \theta_k, \lambda_e) \sim \mathcal{N}(\theta_i, \lambda_e^{-1}).$$

The parameters $\theta_1, \dots, \theta_k$ and λ_e are again assumed to be conditionally independent random variables with

$$\theta_i|(\alpha, \lambda_\theta) \sim \mathcal{N}(\alpha, \lambda_\theta^{-1}) \quad \text{and} \quad \lambda_e|(\alpha, \lambda_\theta) \sim \Gamma(a_2, b_2).$$

Finally, α and λ_θ are independent random variables with

$$\alpha \sim \mathcal{N}(m, v) \quad \text{and} \quad \lambda_\theta \sim \Gamma(a_1, b_1),$$

where $a_1, b_1, a_2, b_2, v \in \mathbb{R}_+$ and $m \in \mathbb{R}$ are given constants.

The posterior distribution μ of the parameter vector $(\theta_1, \dots, \theta_k, \mu, \lambda_e, \lambda_\theta)$ on $\mathbb{R}^{k+3}$ given observations $Y_{ij} = y_{ij}$ is given by Bayes' formula. Although the density is explicit up to a normalizing constant involving a possibly high-dimensional integral, it is not clear how to generate exact samples from μ, and how to compute expectation values w.r.t. μ. On the other hand, it is not difficult to see that all the conditional laws w.r.t. μ of one of the parameters $\theta_1, \dots, \theta_k, \alpha, \lambda_e, \lambda_\theta$ given all the other parameters are either normal or Gamma distributions with parameters depending on the observed data. Therefore, it is easy to run a Gibbs sampler with target distribution μ on a computer. If the corresponding Markov chain converges sufficiently rapidly to its stationary distribution then its values after a sufficiently large number of steps can be used as approximate samples from μ, and long time averages of the values of a function applied to the Markov chain provide estimators for the integral of this function. It is then an obvious question for how many steps the Gibbs sampler has to be run to obtain sufficiently good approximations, see for example Roberts and Rosenthal [47].

Convergence bounds for Gibbs samplers

Returning to the general setup on the product space T^V, we fix a metric ϱ on T, and we denote by d the corresponding l^1 metric on the configuration space T^V, i.e.,

$$d(\xi, \eta) \;=\; \sum_{x \in V} \varrho\,(\xi(x), \eta(x))\,, \quad \xi, \eta \in T^V.$$

A frequent choice is $\varrho(s,t) = 1_{s \neq t}$. In this case, the corresponding metric

$$d(\xi, \eta) = |\{x \in V : \xi(x) \neq \eta(x)\}|$$

on the configuration space is called the **Hamming distance**.

Lemma 3.14. *Let $n = |V|$. Then for the Gibbs sampler,*

$$\mathcal{W}_d^1\,(\pi(\xi, \cdot), \pi(\eta, \cdot)) \;\leq\; \left(1 - \frac{1}{n}\right) d(\xi, \eta) + \frac{1}{n} \sum_{x \in V} \mathcal{W}_\varrho^1\left(\mu_x(\cdot|\hat{\xi}(x)), \mu_x(\cdot|\hat{\eta}(x))\right)$$

for any $\xi, \eta \in T^V$.

Proof. For every $x \in V$, let γ_x be an optimal coupling w.r.t. $\mathcal{W}_\varrho^1$ of the conditional measures $\mu_x(\cdot|\hat{\xi}(x))$ and $\mu_x(\cdot|\hat{\eta}(x))$. Then we can construct a coupling of $\pi(\xi, d\xi')$ and $\pi(\eta, d\eta')$ in the following way:

- Draw $U \sim \mathrm{Unif}(V)$.
- Given U, choose $(\xi'(U), \eta'(U)) \sim \gamma_U$, and set $\xi'(x) = \xi(x)$ and $\eta'(x) = \eta(x)$ for all $x \neq U$.

For this coupling we obtain

$$\begin{aligned} E[d(\xi',\eta')] &= \sum_{x\in V} E[\varrho(\xi'(x),\eta'(x))] \;=\; d(\xi,\eta) + E\left[\varrho(\xi'(U),\eta'(U)) - \varrho(\xi(U),\eta(U))\right] \\ &= d(\xi,\eta) + \frac{1}{n}\sum_{x\in V}\left(\int \varrho(s,t)\gamma_x(dsdt) - \varrho(\xi(x),\eta(x))\right) \\ &= \left(1-\frac{1}{n}\right) d(\xi,\eta) + \frac{1}{n}\sum_{x\in V} \mathcal{W}^1_\varrho\left(\mu_x(\cdot|\hat{\xi}(x)), \mu_x(\cdot|\hat{\eta}(x))\right). \end{aligned}$$

Here we have used in the last step the optimality of the coupling γ_x. The claim follows since $E[d(\xi',\eta')]$ is an upper bound for $\mathcal{W}^1_d(\pi(\xi,\cdot),\pi(\eta,\cdot))$. ■

The lemma shows that we obtain contractivity w.r.t. $\mathcal{W}^1_d$ if the conditional distributions at $x \in V$ do not depend too strongly on the values of the configuration at other vertices.

Theorem 3.15 (Geometric ergodicity of the Gibbs sampler for weak interactions).

1) Suppose that there exists a constant $c \in (0,1)$ such that

$$\sum_{x\in V} \mathcal{W}^1_\varrho\left(\mu_x(\cdot|\hat{\xi}(x))\,,\,\mu_x(\cdot|\hat{\eta}(x))\right) \;\le\; c\,d(\xi,\eta) \qquad \text{for any } \xi,\eta \in T^V. \tag{3.12}$$

Then

$$\mathcal{W}^1_d(\nu\pi^t,\mu) \;\le\; \exp\left(-\frac{1-c}{n}t\right)\mathcal{W}^1_d(\nu,\mu) \qquad \text{for any } \nu \in \mathcal{P}(T^V) \text{ and } t \in \mathbb{Z}_+. \tag{3.13}$$

2) If T is a graph and ϱ is geodesic then it is sufficient to verify (3.12) for neighbouring configurations $\xi,\eta \in T^V$, i.e., for configurations satisfying $\xi = \eta$ on $V \setminus \{x\}$ and $\xi(x) \sim \eta(x)$ for some $x \in V$.

Proof. 1) If (3.12) holds then by Lemma 3.14,

$$\mathcal{W}_d\left(\pi(\xi,\cdot),\pi(\eta,\cdot)\right) \;\le\; \left(1-\frac{1-c}{n}\right) d(\xi,\eta) \;\le\; \exp\left(-\frac{1-c}{n}\right) d(\xi,\eta) \qquad \text{for all } \xi,\eta \in T^V.$$

Hence (3.13) holds by Theorem 3.10.

2) If (T,ϱ) is a geodesic graph and d is the l^1 distance based on ϱ, then (T^V,d) is again a geodesic graph. Indeed, a geodesic path between two configurations ξ and η w.r.t. the l^1 distance is given by changing one component after the other along a geodesic path on T. Therefore, the claim follows by path coupling, see Theorem 3.10, 2). ■

The results in Theorem 3.15 can be applied to many basic models including random colourings, hard core models and Ising models at low temperature.

Example (Random colourings). Suppose that V is a regular graph of degree Δ. Then T^V is geodesic w.r.t. the Hamming distance d. Suppose that ξ and η are admissible random colourings such that $d(\xi,\eta) = 1$, and let $y \in V$ be the unique vertex such that $\xi(y) \neq \eta(y)$. Then

$$\mu_x(\cdot|\hat{\xi}(x)) = \mu_x(\cdot|\hat{\eta}(x)) \qquad \text{for } x = y, \text{ and for any } x \text{ such that } \{x,y\} \notin E.$$

Moreover, for $\{x,y\} \in E$ and $\varrho(s,t) = 1_{s\neq t}$ we have

$$\mathcal{W}^1_\varrho\left(\mu_x(\cdot|\hat{\xi}(x)), \mu_x(\cdot|\hat{\eta}(x))\right) \;\le\; \frac{1}{|T|-\Delta}$$

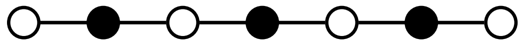
since there are at least $|T| - \Delta$ possible colours available, and the possible colours at x given ξ respectively η on $V \setminus \{x\}$ differ only in one colour. Hence

$$\sum_{x \in V} \mathcal{W}_{\varrho}^{1}\left(\mu_x(\cdot|\hat{\xi}(x)), \mu_x(\cdot|\hat{\eta}(x))\right) \;\leq\; \frac{\Delta}{|T| - \Delta} d(\xi, \eta),$$

and therefore, (3.13) holds with

$$1 - c \;=\; 1 - \frac{\Delta}{|T| - \Delta} \;=\; \frac{|T| - 2\Delta}{|T| - \Delta}.$$

Thus for $|T| > 2\Delta$ we have an exponential decay of the $\mathcal{W}_d^1$ distance to equilibrium with a rate of order $O(n^{-1})$. On the other hand, it is obvious that mixing can break down completely if there are too few colours - consider for example two colours on a linear graph:



Corresponding applications of Theorem 3.15 to Ising and hard core models will be considered in the exercises in Section 3.5 below.

Quantitative bounds for ergodic averages

Suppose that (X_n, P_x) is a Markov chain with transition kernel π. A central problem in the mathematical study of Markov Chain Monte Carlo methods for the estimation of integrals w.r.t. μ is the derivation of bounds for the approximation errors $A_{b,n}f - \mu(f)$. Typically, the initial distribution of the chain is not the stationary distribution, and the number n of steps is large but finite. Thus one is interested in non-asymptotic error bounds for ergodic averages of non-stationary Markov chains. In order to derive such bounds we assume contractivity in an appropriate Kantorovich distance. Suppose that there exists a distance d on S, and constants $\alpha \in (0,1)$ and $\overline{\sigma} \in \mathbb{R}_+$ such that

(A1) $\mathcal{W}_d^1(\nu\pi, \widetilde{\nu}\pi) \leq \alpha \mathcal{W}_d^1(\nu, \widetilde{\nu})$ for any $\nu, \widetilde{\nu} \in \mathcal{P}(S)$, and

(A2) $\mathrm{Var}_{\pi(x,\cdot)}(f) \leq \overline{\sigma}^2 \|f\|_{\mathrm{Lip}(d)}^2$ for any $x \in S$, and any Lipschitz continuous function $f : S \to \mathbb{R}$.

We have seen above an approach for verifying Condition (A1) in applications. Furthermore, Condition (A2) is satisfied with

$$\overline{\sigma}^2 \;=\; 2 \sup_{x \in S} \int d(x,y)^2 \pi(x, dy)$$

provided the supremum is finite. Indeed, for every Lipschitz continuous function f and $x \in S$,

$$\mathrm{Var}_{\pi(x,\cdot)}(f) \;=\; \frac{1}{2} \int\int (f(y) - f(z))^2 \pi(x, dy)\pi(x, dz) \;\leq\; \frac{1}{2} \|f\|_{\mathrm{Lip}(d)}^2 \int\int d(y,z)^2 \pi(x, dy)\pi(x, dz),$$

and thus (A2) follows by the triangle inequality.

Lemma 3.16 (Decay of correlations). *If (A1) and (A2) hold, then the following non-asymptotic bounds hold for every $n, k \in \mathbb{Z}_+$ and every Lipschitz continuous function $f : S \to \mathbb{R}$:*

$$\mathrm{Var}_{P_x}[f(X_n)] \;\leq\; \sum_{k=0}^{n-1} \alpha^{2k} \overline{\sigma}^2 \|f\|_{\mathrm{Lip}(d)}^2, \quad \textit{and} \tag{3.14}$$

$$\left|\mathrm{Cov}_{P_x}[f(X_n), f(X_{n+k})]\right| \;\leq\; \frac{\alpha^k}{1 - \alpha^2} \overline{\sigma}^2 \|f\|_{\mathrm{Lip}(d)}^2. \tag{3.15}$$

Proof. The inequality (3.14) follows by induction on n. It holds true for $n = 0$, and if (3.14) holds for some $n \in \mathbb{Z}_+$ then by the Markov property, (A1) and (A2),

$$\begin{aligned}
\mathrm{Var}_{P_x}[f(X_{n+1})] &= E_x\Big[\,\mathrm{Var}_{P_x}[f(X_{n+1})|\mathcal{F}_n^X]\Big] + \mathrm{Var}_{P_x}\Big[E_x[f(X_{n+1})|\mathcal{F}_n^X]\Big] \\
&= E_x\Big[\,\mathrm{Var}_{\pi(X_n,\cdot)}(f)\Big] + \mathrm{Var}_{P_x}\big[(\pi f)(X_n)\big] \\
&\le \overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)} + \sum_{k=0}^{n-1}\alpha^{2k}\overline{\sigma}^2\|\pi f\|^2_{\mathrm{Lip}(d)} \;\le\; \sum_{k=0}^{n}\alpha^{2k}\overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)}.
\end{aligned}$$

Here we have used in the last step that $\|\pi f\|_{\mathrm{Lip}(d)} \le \alpha\|f\|_{\mathrm{Lip}(d)}$ holds, since by the Kantorovich duality,

$$|(\pi f)(x) - (\pi f)(y)| \;\le\; \|f\|_{\mathrm{Lip}(d)}\mathcal{W}_d^1(\pi(x,\cdot),\pi(y,\cdot)) \;\le\; \alpha\|f\|_{\mathrm{Lip}(d)}\, d(x,y) \qquad \text{for all } x,y \in S.$$

Noting that

$$\sum_{k=0}^{n-1}\alpha^{2k} \;\le\; \frac{1}{1-\alpha^2} \qquad \text{for any } n \in \mathbb{N},$$

the bound (3.15) for the correlations follows from (3.14), since

$$\begin{aligned}
\big|\mathrm{Cov}_{P_x}\left[f(X_n), f(X_{n+k})\right]\big| &= \big|\mathrm{Cov}_{P_x}\left[f(X_n), (\pi^k f)(X_n)\right]\big| \;\le\; \mathrm{Var}_{P_x}\left[f(X_n)\right]^{1/2}\,\mathrm{Var}_{P_x}\left[(\pi^k f)(X_n)\right]^{1/2} \\
&\le \frac{1}{1-\alpha^2}\overline{\sigma}^2\|f\|_{\mathrm{Lip}(d)}\|\pi^k f\|_{\mathrm{Lip}(d)} \;\le\; \frac{\alpha^k}{1-\alpha^2}\overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)}
\end{aligned}$$

by Assumption (A1). ■

As a consequence of Lemma 3.16 we obtain non-asymptotic upper bounds for mean square errors of ergodic averages. The following result and further extensions have been proven in [26].

Theorem 3.17 (Quantitative bounds for ergodic averages of non-stationary Markov chains).
Suppose that (A1) and (A2) hold, and let μ be an invariant probability measure for the transition kernel π. Then the following upper bounds for the bias and variance of ergodic averages hold for all $b, n \in \mathbb{Z}_+$, every initial distribution $\nu \in \mathcal{P}(S)$, and every Lipschitz continuous function $f : S \to \mathbb{R}$:

$$\left|E_\nu\left[A_{b,n}f\right] - \mu(f)\right| \;\le\; \frac{1}{n}\|f\|_{\mathrm{Lip}(d)}\,\frac{\alpha^b}{1-\alpha}\mathcal{W}_d^1(\nu,\mu), \tag{3.16}$$

$$\mathrm{Var}_{P_\nu}\left[A_{b,n}f\right] \le \frac{1}{n}\|f\|^2_{\mathrm{Lip}(d)}\cdot\frac{1}{(1-\alpha)^2}\left(\overline{\sigma}^2 + \frac{\alpha^{2b}}{n}\,\mathrm{Var}(\nu)\right), \tag{3.17}$$

where $\mathrm{Var}(\nu) := \frac{1}{2}\int\int d(x,y)^2\nu(dx)\nu(dy)$.

Proof. 1) By definition of the averaging operator, $E_\nu[A_{b,n}f] = \frac{1}{n}\sum_{i=b}^{b+n-1}(\nu\pi^i)(f)$, and thus

$$\begin{aligned}
\left|E_\nu[A_{b,n}f] - \mu(f)\right| &\le \frac{1}{n}\sum_{i=b}^{b+n-1}|(\nu\pi^i)(f) - \mu(f)| \;\le\; \frac{1}{n}\sum_{i=b}^{b+n-1}\mathcal{W}_d^1(\nu\pi^i,\mu)\,\|f\|_{\mathrm{Lip}(d)} \\
&\le \frac{1}{n}\sum_{i=b}^{b+n-1}\alpha^i\,\mathcal{W}_d^1(\nu,\mu)\,\|f\|_{\mathrm{Lip}(d)}.
\end{aligned}$$

2) By the correlation bound in Lemma 3.16,

$$\begin{aligned}\mathrm{Var}_{P_x}[A_{b,n}f] &= \frac{1}{n^2}\sum_{i,j=b}^{b+n-1}\mathrm{Cov}_{P_x}[f(X_i),f(X_j)] \;\le\; \frac{1}{n^2}\sum_{i,j=b}^{b+n-1}\frac{\alpha^{|i-j|}}{1-\alpha^2}\,\overline{\sigma}^2\,\|f\|^2_{\mathrm{Lip}(d)}\\ &\le \frac{1}{n}\,\frac{\overline{\sigma}^2}{1-\alpha^2}\left(1+2\sum_{k=1}^{\infty}\alpha^k\right)\|f\|^2_{\mathrm{Lip}(d)} \;=\; \frac{1}{n}\,\frac{\overline{\sigma}^2}{(1-\alpha)^2}\,\|f\|^2_{\mathrm{Lip}(d)}.\end{aligned}$$

Therefore, for an arbitrary initial distribution $\nu\in\mathcal{P}(S)$,

$$\begin{aligned}\mathrm{Var}_{P_\nu}[A_{b,n}f] &= E_\nu\left[\mathrm{Var}_{P_\nu}\left[A_{b,n}f|X_0\right]\right]+\mathrm{Var}_{P_\nu}\left[E_\nu\left[A_{b,n}f|X_0\right]\right]\\ &= \int \mathrm{Var}_{P_x}\left[A_{b,n}f\right]\nu(dx)+\mathrm{Var}_\nu\left[\frac{1}{n}\sum_{i=b}^{b+n-1}\pi^i f\right]\\ &\le \frac{1}{n}\,\frac{\overline{\sigma}^2}{(1-\alpha)^2}\|f\|^2_{\mathrm{Lip}(d)}+\left(\frac{1}{n}\sum_{i=b}^{b+n-1}\mathrm{Var}_\nu(\pi^i f)^{1/2}\right)^2.\end{aligned}$$

The assertion now follows since

$$\mathrm{Var}_\nu(\pi^i f)\;\le\;\frac{1}{2}\|\pi^i f\|^2_{\mathrm{Lip}(d)}\iint d(x,y)^2\nu(dx)\nu(dy)\;\le\;\alpha^{2i}\,\|f\|^2_{\mathrm{Lip}(d)}\,\mathrm{Var}(\nu)$$

holds by the Kantorovich duality. ■

3.3. Geometric ergodicity

In this section, we derive different bounds for convergence to equilibrium w.r.t. the total variation distance. In particular, we prove a version of Harris' theorem which states that geometric ergodicity follows from a local minorization combined with a global Lyapunov condition.

Total variation norms

The variation $|\eta|$ of an additive set function $\eta:\mathcal{B}\to\mathbb{R}$ is the non-negative set function defined by

$$|\eta|(B) := \sup\left\{\sum_{i=1}^{n}|\eta(A_i)| : n\in\mathbb{N}, A_1,\dots,A_n\in\mathcal{B}\text{ disjoint with }\bigcup_{i=1}^{n}A_i\subseteq B\right\}\qquad\text{for } B\in\mathcal{B}.$$

If η is a signed measure then $|\eta|$ is a positive measure.

Definition 3.18 (Total variation norm). The *total variation norm* of η is

$$\|\eta\|_{\mathrm{TV}} \;=\; \frac{1}{2}|\eta|(S).$$

More generally, let $V:S\to(0,\infty)$ be a measurable non-negative function. Then the *weighted total variation norm of η w.r.t. V* is defined by

$$\|\eta\|_V \;=\; \int V\,d|\eta|.$$

Note that the definition of the unweighted total variation norm differs from the usual convention in analysis by a factor $\frac{1}{2}$. The reason for introducing the factor $\frac{1}{2}$ will become clear by Lemma 3.19 below.

Now let us assume that η is a finite signed measure on S, and suppose that η is absolutely continuous with density ϱ with respect to some positive reference measure λ. Then there is an explicit Hahn-Jordan decomposition of the state space S and the measure η given by $S = S^+ \dot\cup S^-$ and $\eta = \eta^+ - \eta^-$, where

$$S^+ = \{\varrho \geq 0\}, \ \ S^- = \{\varrho < 0\}, \ \ d\eta^+ = \varrho^+ d\lambda, \ \ d\eta^- = \varrho^- d\lambda.$$

The measures η^+ and η^- are finite positive measures with

$$\eta^+(B \cap S^-) = 0 \quad \text{and} \quad \eta^-(B \cap S^+) = 0 \quad \text{for any } B \in \mathcal{B}.$$

Hence the variation of η is the measure $|\eta|$ given by

$$|\eta| = \eta^+ + \eta^-, \quad \text{i.e.,} \quad d|\eta| = \varrho \cdot d\lambda.$$

In particular, up to a factor $1/2$, the total variation norm of η is the L^1 norm of ϱ:

$$\|\eta\|_{\mathrm{TV}} \ = \ \frac{1}{2}\int |\varrho| dx \ = \ \frac{1}{2}\|\varrho\|_{L^1(\lambda)}. \tag{3.18}$$

Lemma 3.19 (Equivalent descriptions of the TV distance of probability measures). *Let $\mu, \nu \in \mathcal{P}(S)$ and $\lambda \in \mathcal{M}_+(S)$ such that μ and ν are both absolutely continuous w.r.t. λ. Then the following identities hold:*

$$\begin{aligned}
\|\mu - \nu\|_{\mathrm{TV}} \ &= \ (\mu - \nu)^+(S) = (\mu - \nu)^-(S) = 1 - (\mu \wedge \nu)(S) \\
&= \ \frac{1}{2}\left\| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right\|_{L^1(\lambda)} \\
&= \ \frac{1}{2} \sup \left\{ |\mu(f) - \nu(f)| : f \in \mathcal{F}_b(S) \ s.t. \ \|f\|_{\sup} \leq 1 \right\} && (3.19) \\
&= \ \sup \{ |\mu(B) - \nu(B)| : B \in \mathcal{B} \} && (3.20) \\
&= \ \inf \{ P[X \neq Y] : X \sim \mu, Y \sim \nu \} && (3.21)
\end{aligned}$$

In particular, $\|\mu - \nu\|_{TV} \in [0, 1]$.

Remark. 1) The last identity shows that the total variation distance of μ and ν is the Kantorovich distance $\mathcal{W}_d^1(\mu, \nu)$ based on the trivial metric $d(x, y) = 1_{x \neq y}$ on S.

2) The assumption $\mu, \nu \ll \lambda$ can always be satisfied by choosing λ appropriately. For example, we may choose $\lambda = \mu + \nu$.

Proof. Since μ and ν are both probability measures, $(\mu - \nu)(S) = \mu(S) - \nu(S) = 0$. Therefore, $(\mu - \nu)^+(S) = (\mu - \nu)^-(S)$, and

$$\|\mu - \nu\|_{\mathrm{TV}} \ = \ \frac{1}{2}|\mu - \nu|(S) \ = \ (\mu - \nu)^+(S) \ = \ \mu(S) - (\mu \wedge \nu)(S) \ = \ (\mu - \nu)^-(S).$$

The identity $\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\left\| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right\|_{L^1(\lambda)}$ holds by (3.18). Moreover, for $f \in \mathcal{F}_b(S)$ with $\|f\|_{\sup} \leq 1$,

$$|\mu(f) - \nu(f)| \ \leq \ |(\mu - \nu)^+(f)| + |(\mu - \nu)^-(f)| \ \leq \ (\mu - \nu)^+(S) + (\mu - \nu)^-(S) \ = \ 2\|\mu - \nu\|_{\mathrm{TV}}$$

with equality for $f = 1_{S^+} - 1_{S^-}$. This proves the representation (3.19) of $\|\mu - \nu\|_{\mathrm{TV}}$. Furthermore, for $B \in \mathcal{B}$ and $f := 1_B - 1_{B^c}$, we have

$$|\mu(B) - \nu(B)| \ = \ \frac{1}{2}|(\mu - \nu)(B) - (\mu - \nu)(B^c)| \ = \ \frac{1}{2}|\mu(f) - \nu(f)| \ \leq \ \|\mu - \nu\|_{\mathrm{TV}}$$

with equality for $B = S^+$. Thus the representation (3.20) holds.

Finally, to prove (3.21) note that if (X,Y) is a coupling of μ and ν, then

$$|\mu(f) - \nu(f)| \;=\; |E[f(X) - f(Y)]| \;\leq\; 2P[X \neq Y]$$

holds for any bounded measurable function f with $\|f\|_{\sup} \leq 1$. Hence by (3.19),

$$\|\mu - \nu\|_{\mathrm{TV}} \;\leq\; \inf_{\substack{X\sim\mu\\ Y\sim\nu}} P[X \neq Y].$$

To show the converse inequality we choose a coupling (X,Y) that maximizes the probability that X and Y agree. The maximal coupling can be constructed by noting that

$$\mu \;=\; (\mu\wedge\nu) + (\mu-\nu)^+ \;=\; p\alpha + (1-p)\beta, \tag{3.22}$$
$$\nu \;=\; (\mu\wedge\nu) + (\mu-\nu)^- \;=\; p\alpha + (1-p)\delta \tag{3.23}$$

with $p = (\mu\wedge\nu)(S)$ and probability measures $\alpha,\beta,\delta \in \mathcal{P}(S)$. We choose independent random variables $U \sim \alpha, V \sim \beta, W \sim \delta$ and $Z \sim \text{Bernoulli}(p)$, and we define

$$(X,Y) \;=\; \begin{cases} (U,U) & \text{on } \{Z=1\},\\ (V,W) & \text{on } \{Z=0\}.\end{cases}$$

Then by (3.22) and (3.23), (X,Y) is a coupling of μ and ν, and

$$P[X\neq Y] \;\leq\; P[Z=0] \;=\; 1-p \;=\; 1-(\mu\wedge\nu)(S) \;=\; \|\mu-\nu\|_{\mathrm{TV}}. \qquad \blacksquare$$

Most of the representations of total variation norms in Lemma 3.19 have analogues for weighted total variation norms.

Exercise (Equivalent descriptions for weighted total variation norms). Let $V : S \to (0,\infty)$ be a measurable function, and let $d_V(x,y) := (V(x)+V(y))1_{x\neq y}$. Show that the following identities hold for probability measures μ,ν on $(S,\mathcal{B})$:

$$\begin{aligned}
\|\mu-\nu\|_V \;&=\; \left\|\frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda}\right\|_{L^1(V\,d\lambda)}\\
&=\; \sup\{|\mu(f)-\nu(f)| : f\in\mathcal{F}(S) \text{ s.t. } |f|\leq V\}\\
&=\; \sup\{|\mu(f)-\nu(f)| : f\in\mathcal{F}(S) \text{ s.t. } |f(x)-f(y)|\leq d_V(x,y)\ \forall\, x,y\}\\
&=\; \inf\{E[d_V(X,Y)] : X\sim\mu, Y\sim\nu\}
\end{aligned}$$

The last equation is a Kantorovich-Rubinstein duality w.r.t. the underlying metric d_V. The exercise shows that this duality holds although the metric space (S,d_V) is not necessarily separable.

Exercise (Total variation distance of product measures). Let $\nu = \bigotimes_{i=1}^d \nu_i$ and $\mu = \bigotimes_{i=1}^d \mu_i$ be two product probability measures on S^d. Show in at least two different ways that

$$\|\nu-\mu\|_{TV} \;\leq\; \sum_{i=1}^d \|\nu_i-\mu_i\|_{TV}.$$

Does a similar statement hold if the total variation distance is replaced by a general transportation metric $\mathcal{W}^1$?

TV contraction coefficients

Let p be a transition kernel on $(S, \mathcal{B})$. We define the **local contraction coefficient** $\alpha(p, K)$ of p on a set $K \subseteq S$ w.r.t. the total variation distance by

$$\alpha(p,K) \;=\; \sup_{x,y\in K} \|p(x,\cdot) - p(y,\cdot)\|_{\mathrm{TV}} \;=\; \sup_{\substack{x,y\in K\\ x\neq y}} \frac{\|\delta_x p - \delta_y p\|_{\mathrm{TV}}}{\|\delta_x - \delta_y\|_{\mathrm{TV}}}. \tag{3.24}$$

The next result is analogous to Theorem 3.10, 1), but here we do not assume the Feller property of the transition kernel.

Corollary 3.20.

$$\alpha(p,K) \;=\; \sup_{\substack{\mu,\nu\in\mathcal{P}(K)\\ \mu\neq\nu}} \frac{\|\mu p - \nu p\|_{\mathrm{TV}}}{\|\mu - \nu\|_{\mathrm{TV}}}.$$

Proof. Let $\mu, \nu \in \mathcal{P}(K)$. We want to show that $\|\mu p - \nu p\|_{\mathrm{TV}} \le \alpha(p,K)\|\mu - \nu\|_{\mathrm{TV}}$. To this end, we fix a maximal coupling γ of μ and ν, i.e., $\|\mu - \nu\|_{\mathrm{TV}} = \gamma(x \neq y)$. Then by Lemma 3.19, for every measurable function $f : S \to [-1, 1]$,

$$\begin{aligned} \int f\, d(\mu p) - \int f\, d(\nu p) \;&=\; \int_K (pf)\, d\mu - \int_K (pf)\, d\nu \;=\; \int_{K\times K} (pf(x) - pf(y))\, \gamma(dx\, dy) \\ &\le\; 2\,\alpha(p,K)\,\gamma(x \neq y) \;=\; 2\,\alpha(p,K)\, \|\mu - \nu\|_{\mathrm{TV}}. \end{aligned}$$

By Lemma 3.19, the assertion follows by taking the supremum over all functions f as above. ∎

An important consequence of the corollary is that for a Markov semigroup (p_t), the total variation distance $\|\nu p_t - \mu p_t\|_{\mathrm{TV}}$ is a non-increasing function of t.

Exercise (Monotonicity of TV distance to invariant measure). Show that the total variation distance of the law of a Markov process to an invariant probability measure is a non-increasing function of time. Does a similar statement hold if the total variation distance is replaced by a general transportation metric $\mathcal{W}^1$?

Note that in contrast to more general Wasserstein contraction coefficients, we always have

$$\alpha(p,K) \;\le\; 1.$$

Moreover, for $\varepsilon > 0$, we have $\alpha(p,K) \le 1 - \varepsilon$ if p satisfies the following condition:

Local minorization condition. There exists a probability measure ν on S such that

$$p(x,B) \ge \varepsilon \nu(B) \qquad \text{for all } x \in K \text{ and } B \in \mathcal{B}. \tag{3.25}$$

Now suppose that $(p_t)_{t\ge 0}$ is a Markov semigroup. Doeblin's classical theorem states that if $\alpha(p_t, S) < 1$ for some $t \in (0,\infty)$, then *uniform ergodicity* holds in the following sense:

Exercise (Doeblin's Theorem). Suppose that $\alpha(p_t, S) < 1$ for some $t \in (0,\infty)$. Prove that there exists a unique invariant probability measure μ of (p_t), and

$$\sup_{x\in S} \|p_s(x,\cdot) - \mu\|_{\mathrm{TV}} \to 0 \qquad \text{as } s \to \infty. \tag{3.26}$$

If the state space is infinite, a global contraction condition w.r.t. the total variation norm as assumed in Doeblin's Theorem can not be expected to hold in general.

Example (Autoregressive process AR(1)). Suppose that $X_0 = x$ and

$$X_{n+1} \;=\; \alpha X_n + W_{n+1}$$

with $\alpha \in (-1,1), x \in \mathbb{R}$, and independent random variables $W_n \sim N(0,1)$. By induction, one easily verifies that

$$X_n \;=\; \alpha^n x + \sum_{i=0}^{n-1} \alpha^i W_{n-i} \;\sim\; N\left(\alpha^n x, \frac{1-\alpha^{2n}}{1-\alpha^2}\right),$$

i.e., the n-step transition kernel is given by $p_n(x,\cdot) = N\left(\alpha^n x, (1-\alpha^{2n})/(1-\alpha^2)\right)$. As $n \to \infty$, $p_n(x,\cdot)$ converges in total variation to the unique invariant probability measure $\mu = N\left(0, 1/(1-\alpha^2)\right)$. However, the convergence is not uniform in x, since

$$\sup_{x\in\mathbb{R}} \|p_n(x,\cdot) - \mu\|_{\mathrm{TV}} \;=\; 1 \qquad \text{for every } n \in \mathbb{N}.$$

Harris' Theorem

The example above demonstrates the need of a weaker notion of convergence to equilibrium than uniform ergodicity, and of a weaker assumption than the global minorization condition.

Definition 3.21 (Geometric ergodicity). A time-homogeneous Markov process (X_t, P_x) with transition function (p_t) is called *geometrically ergodic with invariant probability measure* μ iff there exist a constant $c > 0$ and a non-negative function $M : S \to \mathbb{R}$ such that

$$\|p_t(x,\cdot) - \mu\|_{\mathrm{TV}} \;\le\; M(x)\exp(-ct) \qquad \text{for } \mu\text{-almost every } x \in S.$$

Harris' Theorem states that geometric ergodicity is a consequence of a *local minorization condition* combined with a *global Lyapunov condition* of the following form:

(LG) There exist a function $V \in \mathcal{F}_+(S)$ and constants $\lambda > 0$ and $C < \infty$ such that

$$\mathcal{L}V(x) \le C - \lambda V(x) \quad \text{for all } x \in S. \tag{3.27}$$

We consider the time-discrete case. Let π be the one step transition kernel of a time-homogeneous Markov chain. Then condition (LG) states that

$$\pi V(x) \le C + \gamma V(x), \qquad \text{where} \qquad \gamma = 1 - \lambda < 1. \tag{3.28}$$

Below, we follow the approach of M. Hairer and J. Mattingly [22] to give a simple proof of a quantitative version of the Harris Theorem, cf. e.g. [21]. The key idea is to replace the total variation distance by the Kantorovich distance

$$\mathcal{W}_\beta(\mu,\nu) \;=\; \inf_{\substack{X\sim\mu \\ Y\sim\nu}} E[d_\beta(\mu,\nu)]$$

based on a distance function on S of the form

$$d_\beta(x,y) \;=\; (1 + \beta V(x) + \beta V(y))\, 1_{x\neq y}$$

with $\beta > 0$. Note that $\|\mu - \nu\|_{\mathrm{TV}} \le \mathcal{W}_\beta(\mu,\nu)$ with equality for $\beta = 0$.

Theorem 3.22 (Quantitative Harris Theorem). Suppose that there exists a function $V \in \mathcal{F}_+(S)$ such that the condition in (LG) is satisfied with constants $C, \lambda \in (0, \infty)$, and

$$\alpha(\pi, \{V \le r\}) < 1 \quad \text{for some } r > 2C/\lambda. \tag{3.29}$$

Then there exists a constant $\beta \in \mathbb{R}_+$ such that $\alpha_\beta(\pi) < 1$. In particular, there is a unique invariant probability measure μ of π satisfying $\int V d\mu < \infty$, and geometric ergodicity holds:

$$\|\pi^n(x, \cdot) - \mu\|_{\mathrm{TV}} \;\le\; \mathcal{W}_\beta\left(\pi^n(x, \cdot), \mu\right) \;\le\; \left(1 + \beta V(x) + \beta \int V d\mu\right) \alpha_\beta(\pi)^n$$

for any $n \in \mathbb{N}$ and $x \in S$.

There are explicit expressions for the constants β and $\alpha_\beta(\pi)$, see the proof below.

Proof. Fix $x, y \in S$ with $x \neq y$, and let (X, Y) be a *maximal coupling* of $\pi(x, \cdot)$ and $\pi(y, \cdot)$ w.r.t. the total variation distance, i.e.,

$$P[X \neq Y] \;=\; \|\pi(x, \cdot) - \pi(y, \cdot)\|_{\mathrm{TV}}.$$

Then for every $\beta \ge 0$,

$$\begin{aligned} \mathcal{W}_\beta(\pi(x, \cdot), \pi(y, \cdot)) \;&\le\; E[d_\beta(X, Y)] \;\le\; P[X \neq Y] + \beta E[V(X)] + \beta E[V(Y)] \\ &=\; \|\pi(x, \cdot) - \pi(y, \cdot)\|_{\mathrm{TV}} + \beta(\pi V)(x) + \beta(\pi V)(y) \\ &\le\; \|\pi(x, \cdot) - \pi(y, \cdot)\|_{\mathrm{TV}} + 2C\beta + (1 - \lambda)\beta(V(x) + V(y)), \end{aligned} \tag{3.30}$$

where we have used (3.28) in the last step. We now fix r as in (3.29), and distinguish cases.

(i) If $V(x) + V(y) \ge r$ then the Lyapunov condition ensures contractivity. Indeed, by (3.30),

$$\mathcal{W}_\beta(\pi(x, \cdot), \pi(y, \cdot)) \;\le\; d_\beta(x, y) + 2C\beta - \lambda\beta \cdot (V(x) + V(y)). \tag{3.31}$$

Since $d_\beta(x, y) = 1 + \beta V(x) + \beta V(y)$, the expression on the right hand side in (3.31) is bounded from above by $(1 - \delta) d_\beta(x, y)$ for some constant $\delta > 0$ provided $2C\beta + \delta \le (\lambda - \delta)\beta r$. This condition is satisfied if we choose

$$\delta \;:=\; \frac{\lambda\beta r - 2C\beta}{1 + \beta r} \;=\; \frac{\lambda r - 2C}{1 + \beta r}\beta,$$

which is positive since $r > 2C/\lambda$.

(ii) If $V(x) + V(y) < r$ then contractivity follows from (3.29). Indeed, (3.30) implies that for $\varepsilon := \min\left(\frac{1 - \alpha(\pi, \{V \le r\})}{2}, \lambda\right)$ and $\beta \le \frac{1 - \alpha(\pi, \{V \le r\})}{4C}$,

$$\mathcal{W}_\beta(\pi(x, \cdot), \pi(y, \cdot)) \;\le\; \alpha(\pi, \{V \le r\}) + 2C\beta + (1 - \lambda)\beta(V(x) + V(y)) \;\le\; (1 - \varepsilon)\, d_\beta(x, y).$$

Choosing $\delta, \varepsilon, \beta > 0$ as in (i) and (ii), we obtain

$$\mathcal{W}_\beta(\pi(x, \cdot), \pi(y, \cdot)) \;\le\; (1 - \min(\delta, \varepsilon))\, d_\beta(x, y) \quad \text{for all } x, y \in S,$$

i.e., the *global* contraction coefficient $\alpha_\beta(\pi)$ w.r.t. $\mathcal{W}_\beta$ is strictly smaller than one. Hence there exists a unique invariant probability measure $\mu \in \mathcal{P}^1_\beta(S) = \left\{\mu \in \mathcal{P}(S) : \int V d\mu < \infty\right\}$, and

$$\begin{aligned} \mathcal{W}_\beta(\pi^n(x, \cdot), \mu) \;&=\; \mathcal{W}_\beta\left(\delta_x \pi^n, \mu\pi^n\right) \;\le\; \alpha_\beta(\pi)^n\, \mathcal{W}_\beta(\delta_x, \mu) \\ &=\; \alpha_\beta(\pi)^n \left(1 + \beta V(x) + \beta \int V d\mu\right). \end{aligned}$$

■

Remark (Doeblin's Theorem). If $\alpha(\pi, S) < 1$ then by choosing $V \equiv 0$, we recover Doeblin's Theorem:

$$\|\pi^n(x,\cdot) - \mu\|_{\mathrm{TV}} \;\leq\; \alpha(\pi, S)^n \;\to\; 0 \quad \text{uniformly in } x \in S.$$

Example (Euler scheme). Consider the Markov chain with state space $\mathbb{R}^d$ and transition step

$$x \mapsto x + h\, b(x) + \sqrt{h}\, \sigma(x) W,$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are continuous functions, $W \sim N(0, I_d)$, and h is a positive constant (the step size of the discretization). Choosing $V(x) = |x|^2$, we obtain

$$\mathcal{L}V(x) \;=\; 2h\, x \cdot b(x) + h^2 |b(x)|^2 + h\, \mathrm{tr}(\sigma^T \sigma)(x) \;\leq\; C - \lambda V(x)$$

for some $C, \lambda \in (0, \infty)$ provided

$$\limsup_{|x|\to\infty} \frac{2x \cdot b(x) + \mathrm{tr}(\sigma^T\sigma)(x) + h|b(x)|^2}{|x|^2} \;<\; 0\,.$$

Noting that for any $r \in (0, \infty)$,

$$\alpha(\pi, \{V \leq r\}) = \sup_{|x|\leq\sqrt{r}} \sup_{|y|\leq\sqrt{r}} \left\| N\left(x + hb(x), h(\sigma\sigma^T)(x)\right) - N\left(y + hb(y), h(\sigma\sigma^T)(y)\right) \right\|_{\mathrm{TV}} < 1,$$

we see that the conditions in Harris' Theorem are satisfied in this case.

Example (Gibbs Sampler in Bayesian Statistics). For several concrete Bayesian posterior distributions on moderately high dimensional spaces, Theorem 3.22 can be applied to show that the total variation distance between the law of the Gibbs sampler after n steps and the stationary target distribution is small after a feasible number of iterations, see Roberts&Rosenthal [47].

3.4. Couplings of Markov processes and convergence rates

On infinite state spaces, convergence to equilibrium may hold only at a **subgeometric** (i.e., slower than exponential) rate. Roughly, subgeometric convergence occurs if the drift is not strong enough to push the Markov chain rapidly back towards the center of the state space. One approach for proving convergence to equilibrium at subgeometric rates is to extend Harris' Theorem. This is possible provided a Lyapunov condition of the form

$$\mathcal{L}V \;\leq\; C - \phi \circ V$$

holds with a concave increasing function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\phi(0) = 0$, see e.g. Hairer [21] and Meyn&Tweedie [40]. Alternatively, couplings of Markov processes can be applied directly to prove both geometric and subgeometric convergence bounds. Both approaches eventually lead to similar conditions. We focus now on the second approach, which is also relevant in the geometrically ergodic case.

Couplings of Markov processes

So far, we have only considered couplings of the transition probabilities of a Markov process. One can often prove stronger results and gain flexibility by considering couplings of the complete laws of two Markov processes on the path space.

Definition 3.23 (Couplings of stochastic processes).

1) A **coupling** of two stochastic processes $((X_t), P)$ and $((Y_t), Q)$ with state spaces S and T is given by a process $\left((\widetilde{X}_t, \widetilde{Y}_t), \widetilde{P}\right)$ with state space $S \times T$ such that

$$(\widetilde{X}_t)_{t\geq 0} \sim (X_t)_{t\geq 0} \quad \text{and} \quad (\widetilde{Y}_t)_{t\geq 0} \sim (Y_t)_{t\geq 0}.$$

2) The coupling is called **Markovian** iff the process $(\widetilde{X}_t, \widetilde{Y}_t)_{t\geq 0}$ is a right continuous strong Markov process on the product space $S \times T$.

We recall that in discrete time, every Markov process has the strong Markov property, and every solution of the martingale problem for the generator is a Markov process. In continuous time, these statements are not always true. However, the strong Markov property holds if the process is right-continuous and the transition semigroup is Feller, or if uniqueness in an appropriate sense holds for the corresponding martingale problem, see Sections 4.2 and 4.4.

Example (Construction of Markovian couplings for Markov chains). A Markovian coupling of two time homogeneous Markov chains can be constructed from a coupling of the one step transition kernels. Suppose that p and q are Markov kernels on measurable spaces $(S, \mathcal{B})$ and $(T, \mathcal{C})$, and $\overline{p}$ is a Markov kernel on $(S \times T, \mathcal{B} \otimes \mathcal{C})$ such that $\overline{p}\,((x,y), dx'dy')$ is a coupling of the measures $p(x, dx')$ and $q(y, dy')$ for every $x \in S$ and $y \in T$. Then for $x \in S$ and $y \in T$, the canonical Markov chain $\big((X_n, Y_n), P_{x,y}\big)$ with one step transition kernel $\overline{p}$ and initial distribution $\delta_{(x,y)}$ is a Markovian coupling of Markov chains with transition kernels p and q and initial distributions δ_x and δ_y. More generally, if γ is a coupling of μ and ν, then $\big((X_n, Y_n), P_\gamma\big)$ is a coupling of Markov chains with transition kernels p, q and initial distributions μ and ν.

In a similar way as in the example above, couplings of Markov processes in continuous time can be constructed from couplings of the generators, cf. e.g. [8]. However, it is often more convenient to construct the couplings in a direct way, see the examples further below.

The following simple but fundamental result is the basis for many convergence bounds in total variation distance:

Theorem 3.24 (Coupling lemma). Suppose that $((X_t, Y_t)_{t\geq 0}, P)$ is a *Markovian coupling* of two time-homogeneous Markov processes with common transition function (p_t) and initial distributions μ and ν. Then

$$\|\mu p_t - \nu p_t\|_{\mathrm{TV}} \;\leq\; \|\mathrm{Law}[(X_s)_{s\geq t}] - \mathrm{Law}[(Y_s)_{s\geq t}]\|_{\mathrm{TV}} \;\leq\; P[T > t],$$

where T is the **coupling time** defined by

$$T \;=\; \inf\{t \geq 0 : X_t = Y_t\}.$$

In particular, the theorem shows that if $T < \infty$ almost surely, then

$$\lim_{t\to\infty} \|\mathrm{Law}(X_{t+\cdot}) - \mathrm{Law}(Y_{t+\cdot})\|_{\mathrm{TV}} \;=\; 0.$$

Proof. 1) By right continuity of the coupling process, X_T and Y_T coincide almost surely. We now first show that without loss of generality, we may even assume that almost surely, $X_t = Y_t$ for any $t \geq T$. Indeed, if this is not the case then we can define a modified coupling $(X_t, \widetilde{Y}_t)$ with *the same coupling time* by setting

$$\widetilde{Y}_t \;:=\; \begin{cases} Y_t & \text{for } t < T, \\ X_t & \text{for } t \geq T. \end{cases}$$

The fact that $(X_t, \widetilde{Y}_t)$ is again a coupling of the same Markov processes follows from the strong Markov property: T is a stopping time w.r.t. the filtration $(\mathcal{F}_t)$ generated by the process (X_t, Y_t), and hence on $\{T < \infty\}$ and under the conditional law given $\mathcal{F}_T$, both $Y_{T+\cdot}$ and $\widetilde{Y}_{T+\cdot} = X_{T+\cdot}$ are Markov processes with transition function (p_t), and their initial values Y_T and X_T coincide almost surely. Therefore, both processes have the same law. Since $Y_t = \widetilde{Y}_t$ for $t < T$, we can conclude that $(Y_t)_{t\geq 0} \sim (\widetilde{Y}_t)_{t\geq 0}$, whence $(X_t, \widetilde{Y}_t)_{t\geq 0}$ is indeed another coupling. Moreover, by construction, the coupling time is the same as before, and $X_t = \widetilde{Y}_t$ for $t \geq T$.

2) Now suppose that $X_t = Y_t$ for $t \geq T$. Then also $X_{t+\cdot} = Y_{t+\cdot}$ for $t \geq T$, and thus we obtain

$$\|\mathrm{Law}\,(X_{t+\cdot}) - \mathrm{Law}\,(Y_{t+\cdot})\|_{\mathrm{TV}} \;\leq\; P\,[X_{t+\cdot} \neq Y_{t+\cdot}] \;\leq\; P[T > t]. \qquad \blacksquare$$

The following example shows that in some cases, the coupling lemma yields sharp bounds if the coupling is chosen adequately.

Example (Brownian motion on $\mathbb{R}^d$). Let $x, y \in \mathbb{R}^d$ such that $x \neq y$. The *reflection coupling* of two Brownian motions with initial values $X_0 = x$ and $Y_0 = y$ is the Markovian coupling given by

$$\begin{aligned} X_t &= x + B_t, \\ Y_t &= y + R(x,y)B_t, \end{aligned}$$

where (B_t) is a Brownian motion with $B_0 = 0$, and

$$R(x,y) \;=\; I_d - 2\,\frac{(x-y)(x-y)^T}{|x-y|^2} \;\in\; O(d)$$

is the reflection at the hyperplane between x and y. Since $R(x,y)$ is an orthogonal transformation, Y_t is indeed a Brownian motion starting at y. Moreover,

$$X_t - Y_t \;=\; x - y + 2\,\frac{x-y}{|x-y|}\,\frac{x-y}{|x-y|}\cdot B_t \;=\; \frac{x-y}{|x-y|}\,(|x-y| + 2W_t), \tag{3.32}$$

where $W_t = \frac{x-y}{|x-y|} \cdot B_t$ is a *one-dimensional* Brownian motion starting at 0. By (3.32), the coupling time is the hitting time of $-|x-y|/2$ for (W_t). Therefore, by the reflection principle and the coupling lemma,

$$\begin{aligned} P[T \leq t] \;&=\; 2P[W_t \leq -|x-y|/2] \;=\; 2\left(1 - \Phi(|x-y|/(2\sqrt{t}))\right), \qquad \text{and} \\ \|p_t(x,\cdot) - p_t(y,\cdot)\|_{TV} \;&\leq\; P[T > t] \;=\; 2\left(\Phi(|x-y|/(2\sqrt{t})) - 1/2\right), \end{aligned}$$

where Φ is the distribution function of the standard normal distribution.

On the other hand, in this simple example we can compute the total variation distance explicitly in order to verify that the bound is sharp. Indeed, by an elementary computation,

$$\|p_t(x,\cdot) - p_t(y,\cdot)\|_{\mathrm{TV}} \;=\; \|N(x,tI_d) - N(y,tI_d)\|_{\mathrm{TV}} \;=\; 2\left(\Phi(|x-y|/(2\sqrt{t})) - 1/2\right).$$

It should be stressed, however, that it is not always possible to find a Markovian coupling such that the bound in the coupling lemma is sharp [7, 31].

Convergence rates

Let $I = \mathbb{Z}_+$ or $I = \mathbb{R}_+$. If μ is an invariant probability measure for a transition function $(p_t)_{t\in I}$ then $\mu p_t = \mu$ and $\mathrm{Law}(X_{t+\cdot}) = P_\mu$ for any $t \geq 0$. Hence the coupling lemma provides upper bounds for the total variation distance to stationarity. As an immediate consequence we note:

Corollary 3.25 (Convergence rates by coupling). Let T be the coupling time for a Markovian coupling of time-homogeneous Markov processes with transition function (p_t) and initial distributions μ and ν. Suppose that

$$E[\psi(T)] < \infty$$

for some non-decreasing, right-continuous function $\psi : I \to \mathbb{R}_+$ with $\lim_{t\to\infty} \psi(t) = \infty$. Then

$$\|\mu p_t - \nu p_t\|_{\mathrm{TV}} \;=\; O\left(\frac{1}{\psi(t)}\right), \tag{3.33}$$

and even

$$\int_0^\infty \|\mu p_t - \nu p_t\|_{\mathrm{TV}}\, d\psi(t) \;<\; \infty \qquad \text{in continuous time, and} \tag{3.34}$$

$$\sum_{n=0}^\infty (\psi(n+1) - \psi(n))\, \|\mu p^n - \nu p^n\|_{\mathrm{TV}} \;<\; \infty \qquad \text{in discrete time, respectively.} \tag{3.35}$$

Proof. By the coupling lemma and by Markov's inequality,

$$\|\mu p_t - \nu p_t\|_{\mathrm{TV}} \;\le\; P[T > t] \;\le\; \frac{1}{\psi(t)} E[\psi(T)] \qquad \text{for all } t > 0.$$

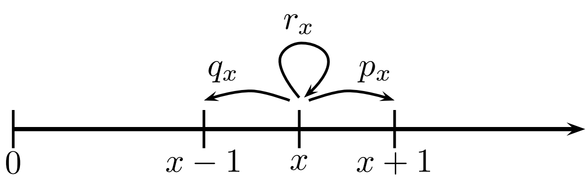
Furthermore, by Fubini's Theorem,

$$\int_0^\infty \|\mu p_t - \nu p_t\|_{\mathrm{TV}}\, d\psi(t) \;\le\; \int_0^\infty P[T \ge t]\, d\psi(t) \;=\; E\left[\int_0^\infty 1_{T\ge t}\, d\psi(t)\right] \;=\; E[\psi(T) - \psi(0)].$$

The assertion in discrete time follows similarly. ■

The corollary shows that convergence to equilibrium happens with a polynomial rate of order $O(n^{-k})$ if there is a coupling with the stationary Markov chain such that the coupling time has a finite k-th moment. If an exponential moment exists then the convergence is geometric.

Example (Markov chains on $\mathbb{Z}_+$). We consider a Markov chain on $\mathbb{Z}_+$ with transition probabilities $\pi(x, x+1) = p_x$, $\pi(x, x-1) = q_x$ and $\pi(x, x) = r_x$. We assume that $p_x + q_x + r_x = 1$, $q_0 = 0$, and $p_x, q_x > 0$ for $x \ge 1$. For simplicity we also assume $r_x = 1/2$ for all x (i.e., the Markov chain is "*lazy*").



For $f \in \mathcal{F}_b(\mathbb{Z}_+)$, the generator is given by

$$(\mathcal{L}f)(x) \;=\; p_x\, (f(x+1) - f(x)) + q_x\, (f(x-1) - f(x)), \qquad x \in \mathbb{Z}_+.$$

By solving the system of equations $\mu\mathcal{L} = \mu - \mu\pi = 0$ explicitly, one shows that there is a two-parameter family of invariant measures given by

$$\mu(x) \;=\; a + b \cdot \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} \qquad (a, b \in \mathbb{R}).$$

In particular, an invariant probability measure exists if and only if

$$Z \;:=\; \sum_{x=0}^\infty \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} \;<\; \infty.$$

For example, this is the case if there exists an $\varepsilon > 0$ such that

$$p_x \leq \left(1 - \frac{1+\varepsilon}{x}\right) q_{x+1} \quad \text{for large } x.$$

Now suppose that an invariant probability measure μ exists. To obtain an upper bound on the rate of convergence to μ, we consider the straightforward Markovian coupling $\big((X_n, Y_n), P_{x,y}\big)$ of two chains with transition kernel π determined by the transition step

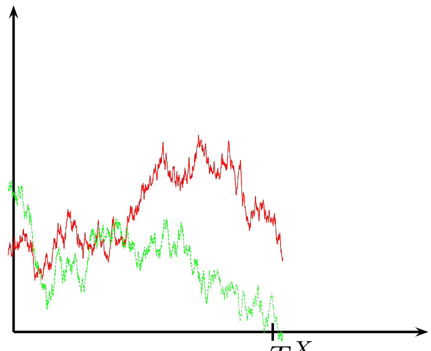
$$(x,y) \mapsto \begin{cases} (x+1,y) & \text{with probability } p_x, \\ (x-1,y) & \text{with probability } q_x, \\ (x,y+1) & \text{with probability } p_y, \\ (x,y-1) & \text{with probability } q_y. \end{cases}$$

Since at each transition step only one of the chains (X_n) and (Y_n) is moving one unit, the processes (X_n) and (Y_n) meet before the trajectories cross each other. In particular, if $X_0 \geq Y_0$ then the coupling time T is bounded from above by the first hitting time

$$T_0^X \;=\; \min\{n \geq 0 : X_n = 0\}.$$

Since an invariant probability measure exists and the chain is irreducible, all states are positive recurrent. Hence

$$E[T] \;\leq\; E[T_0^X] \;<\; \infty.$$



Therefore, by Corollary 3.25, the total variation distance to equilibrium is *always* decaying at least at linear order:

$$\|\pi^n(x,\cdot) - \mu\|_{\mathrm{TV}} \;=\; O(n^{-1}), \qquad \sum_{n=0}^{\infty} \|\pi^n(x,\cdot) - \mu\|_{\mathrm{TV}} < \infty.$$

To prove a stronger decay, one can construct appropriate Lyapunov functions for bounding higher moments of T. For instance suppose that there exist $a > 0$ and $\gamma \in (-1, 0]$ such that

$$p_x - q_x \sim -a x^{\gamma} \quad \text{as} \quad x \to \infty.$$

(i) If $\gamma \in (-1, 0)$ then as $x \to \infty$, the function $V(x) = x^n$ $(n \in \mathbb{N})$ satisfies

$$\begin{aligned} \mathcal{L}V(x) \;&=\; p_x\left((x+1)^n - x^n\right) + q_x\left((x-1)^n - x^n\right) \;\sim\; n(p_x - q_x)x^{n-1} \;\sim\; -nax^{n-1+\gamma} \\ &\leq\; -naV(x)^{1-\frac{1-\gamma}{n}}. \end{aligned}$$

It can now be shown in a similar way as in the proofs of Theorem 1.2 or Theorem 1.6 that

$$E[T^k] \;\leq\; E[(T_0^X)^k] \;<\; \infty \qquad \text{for any } k < \frac{n}{1-\gamma}.$$

Since n can be chosen arbitrarily large, we see that the convergence rate is faster than any polynomial rate:

$$\|\pi^n(x,\cdot) - \mu\|_{\mathrm{TV}} \;=\; O(n^{-k}) \qquad \text{for all } k \in \mathbb{N}.$$

Indeed, by choosing faster growing Lyapunov functions one can show that the convergence rate is $O(\exp(-n^{\beta}))$ for some $\beta \in (0,1)$ depending on γ.

(ii) If $\gamma = 0$ then even geometric convergence holds. Indeed, in this case, for large x, the function $V(x) = e^{\lambda x}$ satisfies

$$\mathcal{L}V(x) \;=\; \left(p_x\left(e^{\lambda}-1\right)+q_x\left(e^{-\lambda}-1\right)\right)V(x) \;\leq\; -c\cdot V(x)$$

for some constant $c > 0$ provided $\lambda > 0$ is chosen sufficiently small. Hence geometric ergodicity follows either by Harris' Theorem, or, alternatively, by applying Corollary 3.25 with $\psi(n) = e^{cn}$.

A CLT for non-stationary Markov chains

Suppose that (X_n, P) is a time-homogeneous Markov chain with transition kernel π, invariant probability measure μ, and initial distribution ν. As in Section 3.2, we consider the ergodic averages

$$A_{b,n}f \;=\; \frac{1}{n}\sum_{i=b}^{b+n-1} f(X_i),$$

The approximation error in the ergodic theorem is

$$A_{b,n}f - \mu(f) \;=\; A_{b,n}f_0,$$

where $f_0 = f - \mu(f)$. By a coupling argument, both the ergodic theorem and the central limit theorem can be extended from the stationary to the non-stationary case.

Theorem 3.26 (Ergodic theorem and CLT for non-stationary Markov chains). Let $b \in \mathbb{Z}_+$, and suppose that $\|\nu\pi^n - \mu\|_{\mathrm{TV}} \to 0$ as $n \to \infty$. Then for any $f \in \mathcal{L}^1(\mu)$, as $n \to \infty$,

$$A_{b,n}f \;\to\; \mu(f) \quad P\text{-almost surely.}$$

Moreover, for any $f \in \mathcal{L}^2(\mu)$ such that $Gf_0 = \sum_{n=0}^{\infty}\pi^n f_0$ converges in $L^2(\mu)$,

$$\sqrt{n}\left(A_{b,n}f - \mu(f)\right) \;\overset{\mathcal{D}}{\to}\; N(0,\sigma_f^2),$$

where $\sigma_f^2 = 2(f_0, Gf_0)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)}$ is the asymptotic variance for the ergodic averages from the stationary case.

Proof. Let $\varepsilon > 0$ be given. Then by the assertion, there exists $t \in \mathbb{N}$ such that

$$\|\nu\pi^t - \mu\pi^t\|_{\mathrm{TV}} \;=\; \|\nu\pi^t - \mu\|_{\mathrm{TV}} \;<\; \varepsilon.$$

Moreover, one can show that there exists a coupling $((X_n, Y_n)_{n\geq 0}, P)$ of Markov chains with initial laws ν and μ and transition kernel π such that

$$P\left[X_{t+n} = Y_{t+n} \text{ for all } n \geq 0\right] \;>\; 1-\varepsilon,$$

see the exercise below. Now let

$$B_{b,n}f \;=\; \frac{1}{n}\sum_{i=b}^{b+n-1} f(Y_i),$$

denote the ergodic averages of the stationary Markov chain $(Y_{b+k})_{k\in\mathbb{Z}_+}$. Then

$$A_{b,n}f - B_{b,n}f \;=\; \frac{1}{n}\sum_{i=b}^{b+n-1}(f(X_i) - f(Y_i))1_{i<t}.$$

On the event $C = \{X_{t+n} = Y_{t+n} \text{ for all } n \geq 0\}$, this difference is of order $O(1/n)$. Since by the ergodic theorem in the stationary case, $B_{b,n}f \to \mu(f)$ almost surely, we can conclude that

$$P\left[\lim_{n\to\infty} A_{b,n}f = \mu(f)\right] \;\geq\; P\left[\lim_{n\to\infty}\left(A_{b,n}f - B_{b,n}f\right) = 0\right] \;\geq\; P[C] \;>\; 1-\varepsilon.$$

Since $\varepsilon > 0$ has been chosen arbitrarily, this proves the first assertion.

Furthermore, let

$$S_n := \sqrt{n}(A_{b,n}f - \mu(f)),\;\; T_n := \sqrt{n}(B_{b,n}f - \mu(f)),\;\; \text{and}\;\; R_n := \sqrt{n}(A_{b,n}f - B_{b,n}f).$$

Then $S_n = T_n + R_n$. By the CLT for stationary Markov chains, we know that $T_n \xrightarrow{\mathcal{D}} N(0,\sigma_f^2)$. Moreover, R_n converges to 0 on the event C, and thus with probability greater than $1-\varepsilon$. Noting that $\varepsilon > 0$ has been chosen arbitrarily, this is sufficient to conclude that $S_n \xrightarrow{\mathcal{D}} N(0,\sigma_f^2)$ as well. Indeed, for any bounded and Lipschitz continuous function $g : S \to \mathbb{R}$ and $\delta > 0$, we obtain

$$\begin{aligned}
&\left|E[g(S_n)] - \int g\, dN(0,\sigma_f^2)\right| \;\leq\; \left|E[g(T_n)] - \int g\, dN(0,\sigma_f^2)\right| \;+\; E\left[|g(S_n) - g(T_n)|\right] \\
&\leq\; \left|E[g(T_n)] - \int g\, dN(0,\sigma_f^2)\right| \;+\; E\left[|g(S_n) - g(T_n)|; |R_n| \leq \delta\right] \;+\; E\left[|g(S_n) - g(T_n)|; |R_n| > \delta\right] \\
&\leq\; \left|E[g(T_n)] - \int g\, dN(0,\sigma_f^2)\right| \;+\; \delta\|g\|_{\mathrm{Lip}} \;+\; 2\|g\|_{\sup} P\left[|R_n| > \delta\right].
\end{aligned}$$

Choosing δ sufficiently small, we see that the limit superior of this expression is smaller than $(1+4\|g\|_{\sup})\varepsilon$, and thus we obtain convergence in distribution by letting ε tend to zero. ■

Exercise (Successful couplings and TV-convergence to equilibrium). Consider a time-homogeneous Markov chain with transition kernel π and invariant probability measure μ.

a) Show that for every initial distribution ν and every fixed integer $t \geq 0$, there exists a coupling (X, Y, P) of the Markov chains with transition kernel π and initial laws ν and μ such that

$$P\left[X_n = Y_n \text{ for all } n \geq t\right] \;=\; \left\|\nu\pi^t - \mu\right\|_{TV}.$$

b) Conclude that $\|\nu\pi^t - \mu\|_{TV} \to 0$ as $t \to \infty$ if and only if for every $\varepsilon > 0$ there exists a coupling of the Markov chains with initial laws ν and μ such that the coupling time

$$T \;=\; \inf\left\{t \geq 0 \;:\; X_n = Y_n \text{ for any } n \geq t\right\}$$

is finite with probability at least $1-\varepsilon$.

***c) A coupling as above is called *successful* if the coupling time is almost surely finite. Show that a successful coupling exists if and only if $\|\nu\pi^t - \mu\|_{TV} \to 0$ as $t \to \infty$.
(This part is optional and very difficult. See Lindvall [37] if you want to read up the proof.)

3.5. Mixing times

Let (p_t) be a Markov semigroup on $(S, \mathcal{B})$ with invariant probability measure μ. For a set $K \in \mathcal{B}$ and $t \geq 0$ let

$$d(t, K) \;=\; \sup_{x\in K} \|p_t(x,\cdot) - \mu\|_{\mathrm{TV}}$$

denote the maximal total variation distance from equilibrium at time t for a corresponding Markov process with initial distribution concentrated on K.

Definition 3.27 (Mixing time). For $\varepsilon > 0$, the *ε-mixing time* of the process with initial value in K is defined by
$$t_{\mathrm{mix}}(\varepsilon, K) \;=\; \inf\{t \geq 0 : d(t,K) \leq \varepsilon\}.$$
Moreover, we denote by $t_{\mathrm{mix}}(\varepsilon)$ the global mixing time $t_{\mathrm{mix}}(\varepsilon, S)$.

Since for every initial distribution $\nu \in \mathcal{P}(S)$, the total variation distance $\|\nu p^t - \mu\|_{\mathrm{TV}}$ is a non-increasing function in t,
$$d(t,K) \leq \varepsilon \qquad \text{for any } t \geq t_{\mathrm{mix}}(\varepsilon, K).$$

The dependence of mixing times on parameters such as the dimension of the underlying state space is an important problem. In particular, the distinction between "**slow mixing**" and "**rapid mixing**", i.e., exponential vs. polynomial increase of the mixing time as a parameter goes to infinity, is related to phase transitions.

Upper bounds in terms of contraction coefficients

To quantify mixing times note that by the triangle inequality for the TV-distance,
$$d_{\mathrm{TV}}(t,S) \;\leq\; \alpha(p^t) \;\leq\; 2 d_{\mathrm{TV}}(t,S),$$
where α denotes the global TV-contraction coefficient. In particular, the bounds derived in Sections 3.1, 3.2 and 3.3 can be applied to control mixing times.

Example (Random colourings). For the random colouring chain with state space T^V, we have shown in the example below Theorem 3.15 that for $|T| > 2\Delta$, the contraction coefficient α_d w.r.t. the Hamming distance $d(\xi,\eta) = |\{x \in V : \xi(x) \neq \eta(x)\}|$ satisfies
$$\alpha_d(p^t) \;\leq\; \alpha_d(p)^t \;\leq\; \exp\left(-\frac{|T|-2\Delta}{|T|-\Delta}\cdot\frac{t}{n}\right). \tag{3.36}$$
Here Δ denotes the degree of the regular graph V and $n = |V|$. Since
$$1_{\xi\neq\eta} \;\leq\; d(\xi,\eta) \;\leq\; n\cdot 1_{\xi\neq\eta} \qquad \text{for all } \xi,\eta \in T^V,$$
we also have
$$\|\nu-\mu\|_{\mathrm{TV}} \;\leq\; \mathcal{W}_d^1(\nu,\mu) \;\leq\; n\|\nu-\mu\|_{\mathrm{TV}} \qquad \text{for all } \nu \in \mathcal{P}(S).$$
Therefore, by (3.36), we obtain
$$\|p^t(\xi,\cdot)-\mu\|_{\mathrm{TV}} \;\leq\; n\alpha_d(p^t) \;\leq\; n\exp\left(-\frac{|T|-2\Delta}{|T|-\Delta}\cdot\frac{t}{n}\right)$$
for any $\xi \in T^V$ and $t \geq 0$. The right-hand side is smaller than ε for $t \geq \frac{|T|-\Delta}{|T|-2\Delta} n \log(n/\varepsilon)$. Thus we have shown that
$$t_{\mathrm{mix}}(\varepsilon) \;=\; O\left(n\log n + n\log\varepsilon^{-1}\right) \qquad \text{for } |T| > 2\Delta.$$
This is a typical example of *rapid mixing* with a *total variation cut-off*: After a time of order $n\log n$, the total variation distance to equilibrium decays to an arbitrary small value $\varepsilon > 0$ in a time window of order $O(n)$.

Exercise (Hard core model). Consider a finite graph (V,E) with n vertices of maximal degree Δ. The corresponding hard core model with fugacity $\lambda > 0$ is the probability measure μ_λ on $\{0,1\}^V$ with mass function
$$\mu_\lambda(\eta) \;=\; Z(\lambda)^{-1}\,\lambda^{\sum_{x\in V}\eta(x)} \quad \text{if } \eta(x)\cdot\eta(y) = 0 \text{ for any } \{x,y\} \in E, \quad \mu_\lambda(\eta) \;=\; 0 \quad \text{otherwise},$$
where $Z(\lambda)$ is a normalization constant.

a) Describe the transition rule for the Gibbs sampler with equilibrium μ_λ, and determine the transition kernel π.

b) Prove that for $\lambda < (\Delta - 1)^{-1}$ and $t \in \mathbb{N}$,

$$\mathcal{W}^1(\nu\pi^t, \mu) \;\le\; \alpha(n,\Delta)^t\, \mathcal{W}^1(\nu,\mu) \;\le\; \exp\left(-\frac{t}{n}\left(\frac{1-\lambda(\Delta-1)}{1+\lambda}\right)\right) \mathcal{W}^1(\nu,\mu),$$

where $\alpha(n,\Delta) = 1 - \frac{1}{n}\left(\frac{1-\lambda(\Delta-1)}{1+\lambda}\right)$, and $\mathcal{W}^1$ is the transportation metric based on the Hamming distance on $\{0,1\}^V$.

c) Show that in this case, the ε-mixing time is of order $O(n\log n)$ for every $\varepsilon \in (0,1)$.

Exercise (Gibbs sampler for the Ising model). Consider a finite graph (V,E) with n vertices of maximal degree Δ. The Ising model with inverse temperature $\beta \ge 0$ is the probability measure μ_β on $\{-1,1\}^V$ with mass function

$$\mu_\beta(\eta) \;=\; \frac{1}{Z(\beta)} \exp\left(\beta \sum_{\{x,y\}\in E} \eta(x)\eta(y)\right),$$

where $Z(\beta)$ is a normalization constant.

a) Show that given $\eta(y)$ for $y \neq x$, $\eta(x) = \pm 1$ with probability $(1 \pm \tanh(\beta m(x,\eta)))/2$, where $m(x,\eta) := \sum_{y\sim x} \eta(y)$ is the local magnetization in the neighbourhood of x. Hence determine the transition kernel π for the Gibbs sampler with equilibrium μ_β.

b) Prove that for every $t \in \mathbb{N}$,

$$\mathcal{W}^1(\nu\pi^t, \mu_\beta) \;\le\; \alpha(n,\beta,\Delta)^t\, \mathcal{W}^1(\nu,\mu_\beta) \;\le\; \exp\left(-\frac{t}{n}\,(1-\Delta\tanh(\beta))\right) \mathcal{W}^1(\nu,\mu_\beta),$$

where $\alpha(n,\beta,\Delta) = 1 - (1-\Delta\tanh(\beta))/n$, and $\mathcal{W}^1$ is the transportation metric based on the Hamming distance on $\{-1,1\}^V$. Conclude that for $\Delta\tanh\beta < 1$, the Gibbs sampler is geometrically ergodic with a rate of order $\Omega(1/n)$. *Hint: You may use the inequality*

$$|\tanh(y+\beta) - \tanh(y-\beta)| \;\le\; 2\tanh(\beta) \qquad \text{for any } \beta \ge 0 \text{ and } y \in \mathbb{R}.$$

c) The *mean-field Ising model* with parameter $\alpha \ge 0$ is the Ising model on the complete graph over $V = \{1,\dots,n\}$ with inverse temperature $\beta = \alpha/n$. Show that for $\alpha < 1$, the ε-mixing time for the Gibbs sampler on the mean field Ising model is of order $O(n\log n)$ for every $\varepsilon \in (0,1)$.

Example (Harris Theorem). In the situation of Theorem 3.22, the global distance $d(t,S)$ to equilibrium does not go to 0 in general. However, on the level sets of the Lyapunov function V,

$$d(t,\{V \le r\}) \le \left(1 + \beta r + \beta \int V d\mu\right) \alpha_\beta(p)^t$$

for any $t, r \ge 0$ where β is chosen as in the theorem, and α_β is the contraction coefficient w.r.t. the corresponding distance d_β. Hence

$$t_{\mathrm{mix}}(\varepsilon,\{V \le r\}) \le \frac{\log\left(1 + \beta r + \beta \int V d\mu\right) + \log(\varepsilon^{-1})}{\log(\alpha_\beta(p)^{-1})}.$$

Upper bounds by coupling

Alternatively, we can also apply the coupling lemma to derive upper bounds for mixing times.

Corollary 3.28 (Coupling times and mixing times). Suppose that for every $x, y \in S$, $\big((X_t,Y_t), P_{x,y}\big)$ is a Markovian coupling of the Markov processes with initial values $x,y \in S$ and transition function (p_t), and let $T = \inf\{t \ge 0 : X_t = Y_t\}$. Then

$$t_{\mathrm{mix}}(\varepsilon) \;\le\; \inf\{t \ge 0 : P_{x,y}[T > t] \le \varepsilon \;\forall\, x,y \in S\}.$$

Proof. The assertion is a direct consequence of the coupling lemma (Theorem 3.24). ■

Example (Lazy Random Walks). A lazy random walk on a graph is a random walk that stays at its current position during each step with probability $1/2$. Lazy random walks are considered to exclude periodicity effects that may occur due to the time discretization. By simple coupling arguments we obtain bounds for total variation distances and mixing times on different graphs:

1) $S = \mathbb{Z}/(m\mathbb{Z})$: The transition probabilities of the lazy simple random walk on a discrete circle with m points are $\pi(x, x+1) = \pi(x, x-1) = 1/4$, $\pi(x,x) = 1/2$, and $\pi(x,y) = 0$ otherwise. A Markovian coupling (X_n, Y_n) is given by moving from (x,y) to $(x+1,y), (x-1,y), (x,y+1), (x,y-1)$ with probability $1/4$ each. Hence only one of the two copies is moving during each step, and thus the two random walks X_n and Y_n can not cross each other without meeting at the same position. The difference process $X_n - Y_n$ is a simple random walk on S, and T is the hitting time of 0. Hence by the Poisson equation,

$$E_{x,y}[T] \;=\; E^{\mathrm{RW}(\mathbb{Z})}_{|x-y|}[T_{\{1,2,\dots,m-1\}^c}] \;=\; |x-y| \cdot (m - |x-y|) \;\leq\; m^2/4.$$

Corollary 3.28 and Markov's inequality now imply that $t_{\mathrm{mix}}(1/4) \leq m^2$, which is a rather sharp upper bound.

2) $S = \{0,1\}^d$: The lazy random walk on the hypercube $\{0,1\}^d$ coincides with the Gibbs sampler for the uniform distribution. A bound for the mixing time can be derived from Theorem 3.15. Alternatively, we can construct a coupling of two lazy random walks on the hypercube by updating the same coordinate to the same value for both copies in each step. Then the coupling time T is bounded from above by the first time where each coordinate has been updated once, i.e., by the number of draws required to collect each of d coupons by sampling with replacement. Therefore, for $c \geq 0$,

$$P[T > d\log d + cd] \;\leq\; \sum_{k=1}^{d} (1 - 1/d)^{\lceil d\log d + cd\rceil} \;\leq\; d\,\exp\left(-\frac{d\log d + cd}{d}\right) \;\leq\; e^{-c},$$

and hence

$$t_{\mathrm{mix}}(\varepsilon) \;\leq\; d\log d + \log(\varepsilon^{-1})d.$$

Conversely, the coupon collecting problem also shows that this upper bound is again almost sharp.

Example (Brownian motion on a flat torus). Let $\mathbb{T}_a = \mathbb{R}/(a\mathbb{Z})$ denote the circle of perimeter a. A Brownian motion on the flat torus $\mathbb{T}_a^d$ can be obtained by projection from a Brownian motion on the covering space $\mathbb{R}^d$. Indeed, suppose that $(B_t)_{t\geq 0}$ is standard Brownian motion on $\mathbb{R}^d$, and let

$$X_t \;=\; x \;+\; \pi(B_t),$$

where $x \in \mathbb{T}_a^d$ is a fixed initial value, $\pi : \mathbb{R}^d \to \mathbb{T}_a^d$ is the canonical projection, and "+" is the addition on the abelian group $\mathbb{T}_a^d$. Then the process $X_t = (X_t^1, \dots, X_t^d)$ is a Markov process with state space $S = \mathbb{T}_a^d$, and its components X_t^i, $i = 1, \dots, d$, are independent Brownian motions on the circle $\mathbb{T}_a$.

A simple antithetic coupling of two Brownian motions X and Y on the torus with initial values $X_0 = x$ and $Y_0 = y$ is given by defining X as above, and setting

$$Y_t \;:=\; y - x - X_t.$$

However, in dimension $d \geq 2$, the corresponding coupling time is almost surely infinite, since two components of X and Y will almost surely not meet at the same time. Therefore, we modify the coupling to a *componentwise coupling* (X, Z) that is defined by

$$Z_t^i \;:=\; \begin{cases} y^i - x^i - X_t^i & \text{for } t \;<\; T_i \;:=\; \inf\{s \geq 0 : Y_s^i = X_s^i\}, \\ X_t^i & \text{for } t \;\geq\; T_i\,. \end{cases}$$

It is left to the reader to verify that (X, Z) is indeed a coupling of two Brownian motions on the torus with initial conditions $X_0 = x$ and $Z_0 = y$. The corresponding coupling time is $T = \max_i T_i$. Moreover,

$$P\left[T_i \geq t\right] \;=\; P\left[T_{\mathbb{R}\setminus(-a/2,a/2)}(B^i) \geq t\right] \;\leq\; \sqrt{2}\,\exp\left(-\frac{\pi^2 t}{8a^2}\right), \quad \text{and thus}$$

$$P\left[T \geq t\right] \;=\; P\left[\exists i : T_i \geq t\right] \;\leq\; \sqrt{2}\,d\,\exp\left(-\frac{\pi^2 t}{8a^2}\right).$$

Here we have used in the first step that for $\lambda \in (0, \pi/a)$, the function $V(x) = \cos(\lambda x)$ is non-negative on $(-a/2, a/2)$ and satisfies $\frac{1}{2}V'' = -\frac{\lambda^2}{2}V$. Applying optional stopping to the corresponding non-negative martingale $\exp(\lambda^2 t/2)V(B_t^i)$ yields the bound for the exit time of B^i from the interval $(-a/2, a/2)$ when choosing $\lambda = \pi/(2a)$. By choosing t such that the upper bound in the last inequality is smaller than ε, we see that

$$t_{\mathrm{mix}}(\varepsilon) \;\leq\; \frac{8a^2}{\pi^2} \log\left(\frac{\sqrt{2}d}{\varepsilon}\right) \;=\; O\left(a^2 \log d + a^2 \log(\varepsilon^{-1})\right).$$

It can be shown that this upper bound has the optimal order in a, d and ε.

Conductance lower bounds

A simple and powerful way to derive lower bounds for mixing times due to constraints by bottlenecks is the conductance. Let π be a Markov kernel on $(S, \mathcal{B})$ with invariant probability measure μ. For sets $A, B \in \mathcal{B}$ with $\mu(A) > 0$, the **equilibrium flow** $Q(A, B)$ from a set A to a set B is defined by

$$Q(A, B) \;=\; (\mu \otimes p)(A \times B) \;=\; \int_A \mu(dx)\, \pi(x, B),$$

and the **conductance** of A is given by

$$\Phi(A) \;=\; \frac{Q(A, A^C)}{\min(\mu(A), \mu(A^c))}.$$

Notice that in particular,

$$\Phi(A) \;\leq\; \frac{\mu(\partial_{\mathrm{int}} A)}{\min(\mu(A), \mu(A^c))},$$

where $\partial_{\mathrm{int}} A \;:=\; \{x \in A : \pi(x, A^c) > 0\}$ is the *interior boundary* of the set A.

Example (Graph with a bottleneck). Consider the uniform distribution on a graph that has two large components A and A^c, consisting each of at least n vertices, which are only connected by a single edge. Then for any Markov chain that only moves along edges of the graph, the conductance of A is bounded from above by $1/n$.

Definition 3.29 (Isoperimetric constant). The *bottleneck ratio* or *isoperimetric constant* Φ_* for a transition kernel π with invariant probability measure μ is defined as the worst-case conductance, i.e.,

$$\Phi_* \;:=\; \inf_{A \in \mathcal{B}} \Phi(A).$$

The inverse of Φ^*, i.e., the maximal inverse conductance, provides a lower bound on mixing times.

Exercise (Conductance and lower bounds for mixing times). Prove the lower bound

$$t_{\mathrm{mix}}\left(\frac{1}{4}\right) \;\geq\; \frac{1}{4\Phi_*}. \tag{3.37}$$

You may proceed in the following way: Let $\mu_A(B) = \mu(B|A)$ denote the conditioned measure given A.

a) Show that for any $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$\|\mu_A \pi - \mu_A\|_{TV} \;=\; (\mu_A \pi)(A^C) \;=\; \Phi(A).$$

Hint: Prove first that

(i) $(\mu_A \pi)(B) - \mu_A(B) \leq 0$ for any measurable $B \subseteq A$, and

(ii) $(\mu_A\pi)(B) - \mu_A(B) = (\mu_A\pi)(B) \geq 0$ for any measurable $B \subseteq A^C$.

b) Conclude that

$$\|\mu_A - \mu\|_{TV} \;\leq\; t\Phi(A) + \|\mu_A\pi^t - \mu\|_{TV} \qquad \text{for all } t \in \mathbb{Z}_+.$$

c) Hence prove (3.37).

In the example above, the conductance lower bound shows that the mixing time is of order $\Omega(n)$. By a similar argument, we can show that for the mean-field Ising model with parameter $\alpha > 1$, the mixing time is exponential in the number of spins:

Example (Mean-field Ising model). The mean-field Ising model with parameters $n \in \mathbb{N}$ and $\alpha \in (0, \infty)$ is the probability measure $\mu_{\alpha,n}$ on $\{-1,+1\}^n$ with mass function

$$\mu_{\alpha,n}(\eta) \;\propto\; \exp\left(\frac{\alpha}{2}\sum_{x=1}^{n}\eta(x)\,\frac{1}{n}\sum_{y=1}^{n}\eta(y)\right) \;=\; \exp\left(\frac{\alpha}{2n}\,m(\eta)^2\right),$$

where $m(\eta) = \sum_{x=1}^n \eta(x)$ is the total magnetization of the configuration η. In the exercise in Section 3.5 above it is shown that for $\alpha < 1$, the Gibbs sampler for the mean-field Ising model has a mixing time of order $O(n \log n)$. Conversely, we will now show by applying the conductance lower bound that for $\alpha > 1$, the mixing time grows exponentially in n. Thus there is a *dynamic phase transition* from rapid mixing to slow mixing at the critical value $\alpha = 1$.

To apply the conductance bound, we consider the sets

$$A_n \;=\; \{\eta \in \{-1,+1\}^n \;:\; m(\eta) > 0\}.$$

Notice that during each step of the Gibbs sampler, the total magnetization changes at most by 2. Therefore, the interior boundary of A_n consists of configurations with $m(\eta) \in \{1, 2\}$. Since by symmetry, $\mu_{\alpha,n}(A_n) \leq 1/2$, we obtain for any $c > 0$,

$$\begin{aligned}\frac{1}{\Phi(A_n)} \;&\geq\; \frac{\mu_{\alpha,n}(A_n)}{\mu_{\alpha,n}(\partial_{\mathrm{int}}A_n)} \;=\; \frac{\mu_{\alpha,n}(m > 0)}{\mu_{\alpha,n}(m \in \{1,2\})} \;\geq\; \frac{\int_{\{m \geq cn\}} \exp(\alpha m^2/(2n))\, d\mu_{0,n}}{\int_{\{m\in\{1,2\}\}} \exp(\alpha m^2/(2n))\, d\mu_{0,n}} \\ &\geq\; \exp\left(-2\alpha n^{-1} + \alpha c^2 n/2\right)\, \mu_{0,n}(m \geq cn) \end{aligned} \tag{3.38}$$

Noting that under $\mu_{0,n}$, the components of η are independent random variables with $\mu_{0,n}(\eta(x) = \pm 1) = 1/2$, we can apply the classical large deviation lower bound (Cramér's Theorem) to conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{0,n}(m \geq cn) \;\geq\; I(c) \;:=\; \frac{1+c}{2}\log(1+c) + \frac{1-c}{2}\log(1-c)\,.$$

Here the exponential rate $I(c)$ is the relative entropy of the Bernoulli distribution with parameter $(1+c)/2$ w.r.t. the uniform distribution. We obtain

$$\frac{1}{\Phi(A_n)} \;\geq\; e^{(\alpha c^2/2 - I(c))(n + o(n))}.$$

A Taylor expansion shows that $\alpha c^2/2 - I(c) = (\alpha-1)c^2/2 + O(c^4)$ as $c \downarrow 0$. Hence if $\alpha > 1$ then we see by choosing $c > 0$ sufficiently small that $1/\Phi(A_n)$ is growing exponentially as $n \to \infty$. Therefore, by (3.37), there exists $\lambda(\alpha) > 0$ such that

$$t_{\mathrm{mix}}(1/4) \;\geq\; 1/(4\Phi(A_n)) \;\geq\; e^{\lambda(\alpha)(n+o(n))}.$$

Part II.

Constructions of Markov processes in continuous time

4. Generators, semigroups and continuous-time Markov processes

This chapter focuses on the connection between Markov processes in continuous time and their generators. Throughout we assume that the state space S is a Polish space with Borel σ-algebra $\mathcal{B}$. Recall that a right-continuous stochastic process $((X_t)_{t\in\mathbb{R}_+}, P)$ that is adapted to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is called a *solution of the martingale problem for a family* $(\mathcal{L}_t, \mathcal{A})$*,* $t \in \mathbb{R}_+$*, of linear operators* with domain $\mathcal{A} \subseteq \mathcal{F}_b(S)$ if and only if

$$M_t^{[f]} = f(X_t) - \int_0^t (\mathcal{L}_s f)(X_s)\, ds \tag{4.1}$$

is an $(\mathcal{F}_t)$ martingale for every function $f \in \mathcal{A}$. Here functions $f : S \to \mathbb{R}$ are extended trivially to $S \cup \{\Delta\}$ by setting $f(\Delta) := 0$. If $((X_t), P)$ solves the martingale problem for $((\mathcal{L}_t), \mathcal{A})$ and the function $(t, x) \mapsto (\mathcal{L}_t f)(x)$ is, for example, continuous and bounded for $f \in \mathcal{A}$, then $(\mathcal{L}_t f)(X_t)$ is the expected rate of change of $f(X_t)$ in the next instant of time given the previous information, i.e.,

$$(\mathcal{L}_t f)(X_t) \;=\; \lim_{h\downarrow 0} E\left[\left.\frac{f(X_{t+h}) - f(X_t)}{h}\right| \mathcal{F}_t\right] \tag{4.2}$$

In general, solutions of a martingale problem are not necessarily Markov processes, but it can be shown under appropriate assumptions, that the strong Markov property follows from uniqueness of solutions of the martingale problem with a given initial law, see Theorem 4.20 below. Now suppose that for any $t \geq 0$ and $x \in S$, $((X_s)_{s\geq t}, P_{(t,x)})$ is an $(\mathcal{F}_t)$ Markov process with initial value $X_t = x$ $P_{(t,x)}$-almost surely and transition function $(p_{s,t})_{0\leq s\leq t}$ that solves the martingale problem above. Then for any $t \geq 0$ and $x \in S$,

$$(\mathcal{L}_t f)(x) \;=\; \lim_{h\downarrow 0} E_x\left[\frac{f(X_{t+h}) - f(X_t)}{h}\right] \;=\; \lim_{h\downarrow 0} \frac{(p_{t,t+h} f)(x) - f(x)}{h},$$

provided $(t, x) \mapsto (\mathcal{L}_t f)(x)$ is continuous and bounded. This indicates that the infinitesimal generator of the Markov process at time t is an extension of the operator $(\mathcal{L}_t, \mathcal{A})$; this fact will be made precise in Section 4.2.

In this chapter we will mostly restrict ourselves to the time-homogeneous case. The time-inhomogeneous case is nevertheless included implicitly, since we may apply most results to the time-space process $\hat{X}_t = (t_0 + t, X_{t_0+t})$ that is always a time-homogeneous Markov process if X is a Markov process. In Section 4.1 we show how to realize transition functions of time-homogeneous Markov processes as strongly continuous contraction semigroups on appropriate Banach spaces of functions, and we establish a one-to-one correspondence between such semigroups and their generators. The connection to martingale problems is made in Section 4.2, and Section 4.4 indicates in a special situation how solutions of martingale problems can be constructed from their generators.

4.1. Semigroups, generators and resolvents

In the discrete time case, there is a one-to-one correspondence between generators $\mathcal{L} = \pi - I$, transition semigroups $p_t = \pi^t$, and time-homogeneous canonical Markov chains $((X_n)_{n\in\mathbb{Z}_+}, (P_x)_{x\in S})$ solving the martingale problem for $\mathcal{L}$ on bounded measurable functions. Our goal in this section is to establish a counterpart to the correspondence between generators and transition semigroups in continuous time. Since the generator will usually be an unbounded operator, this requires the realization of the transition semigroup and the generator on an appropriate Banach space consisting of measurable functions (or equivalence classes of functions) on the state space $(S, \mathcal{B})$. Unfortunately, there is no Banach space that is adequate for all purposes - so the realization on a Banach space also leads to a partially more restrictive setting. Supplementary references for this section are the books on functional analysis and semigroup theory by Yosida [56], Pazy [43], and Davies [9], as well as Ethier and Kurtz [19].

We assume that we are given a time-homogeneous transition function $(p_t)_{t\geq 0}$ on $(S, \mathcal{B})$, i.e.,

(i) $p_t(x, dy)$ is a sub-probability kernel on $(S, \mathcal{B})$ for every $t \geq 0$, and

(ii) $p_0(x, \cdot) = \delta_x$ and $p_t p_s = p_{t+s}$ for all $t, s \geq 0$ and $x \in S$.

Remark (Inclusion of time-inhomogeneous case). Although we restrict ourselves to the time-homogeneous case, the time-inhomogeneous case is included implicitly. Indeed, if $((X_t)_{t\geq s}, P_{(s,x)})$ is a time-inhomogeneous Markov process with transition function $p_{s,t}(x, B) = P_{(s,x)}[X_t \in B]$, then the time-space process $\hat{X}_t = (t + s, X_{t+s})$ is a time-homogeneous Markov process w.r.t. $P_{(s,x)}$ with state space $\mathbb{R}_+ \times S$ and transition function

$$\hat{p}_t\left((s, x), du\, dy\right) = \delta_{t+s}(du)\, p_{s,t+s}(x, dy).$$

Sub-Markovian semigroups and resolvents

The transition kernels p_t act as linear operators $f \mapsto p_t f$ on bounded measurable functions on S. They also act on L^p spaces w.r.t. a measure μ if μ is sub-invariant for the transition kernels:

Definition 4.1 (Sub-invariant measure). A positive measure $\mu \in \mathcal{M}_+(S)$ is called *sub-invariant* w.r.t. the transition semigroup (p_t) iff $\mu p_t \leq \mu$ for any $t \geq 0$ in the sense that

$$\int p_t f\, d\mu \leq \int f\, d\mu \quad \text{for any } f \in \mathcal{F}_+(S) \text{ and } t \geq 0.$$

For processes with finite life-time, non-trivial invariant measures often do not exist, but sub-invariant measures usually do exist.

Lemma 4.2 (Sub-Markov semigroup and contraction properties).

1) *A transition function $(p_t)_{t\geq 0}$ on $(S, \mathcal{B})$ induces a* ***sub-Markovian semigroup*** *of linear operators on $\mathcal{F}_b(S)$ or $\mathcal{F}_+(S)$ respectively, i.e., the following properties hold for any $s, t \geq 0$:*

   (i) Semigroup property: $p_s p_t = p_{s+t}$,

   (ii) Positivity preserving: $f \geq 0 \Rightarrow p_t f \geq 0$,

   (iii) $p_t 1 \leq 1$.

2) ***Contractivity w.r.t. the supremum norm:*** *For any $t \geq 0$,*

$$\|p_t f\|_{sup} \leq \|f\|_{sup} \quad \text{for all } f \in \mathcal{F}_b(S).$$

3) ***Contractivity w.r.t.* $\mathbf{L^p}$ *norms:*** *If $\mu \in \mathcal{M}_+(S)$ is a sub-invariant measure, then for every $p \in [1, \infty]$ and $t \geq 0$,*

$$\int |p_t f|^p d\mu \leq \int |f|^p d\mu \quad \text{for all } f \in \mathcal{L}^p(S, \mu).$$

   In particular, the map $f \mapsto p_t f$ respects μ-classes.

Proof. Most of the statements are straightforward to prove and left as an exercise. We only prove the last statement for $p \in [1,\infty)$. Note first that for $t \geq 0$, the sub-Markov property implies $p_t f \leq p_t|f|$ and $-p_t f \leq p_t|f|$ for any $f \in \mathcal{L}^p(S,\mu)$. Hence by Jensen's inequality,

$$|p_t f|^p \leq (p_t|f|)^p \leq p_t|f|^p.$$

Integration w.r.t. μ yields

$$\int |p_t f|^p \, d\mu \leq \int p_t|f|^p \, d\mu \leq \int |f|^p \, d\mu$$

by the sub-invariance of μ. Hence p_t is a contraction on $\mathcal{L}^p(S,\mu)$. In particular, p_t respects μ-classes since $f = g$ μ-a.e. $\Rightarrow$ $f - g = 0$ μ-a.e. $\Rightarrow$ $p_t(f-g) = 0$ μ-a.e. $\Rightarrow$ $p_t f = p_t g$ μ-a.e. ∎

The theorem shows that (p_t) induces contraction semigroups of linear operators P_t on the following Banach spaces:

- $\mathcal{F}_b(S)$ endowed with the supremum norm,
- $C_b(S)$, provided p_t is *Feller* for every $t \geq 0$,
- $\hat{C}(S) = \{f \in C(S) : \forall \varepsilon > 0 \; \exists K \subset S \text{ compact: } |f| < \varepsilon \text{ on } S \backslash K\}$, provided p_t maps $\hat{C}(S)$ to $\hat{C}(S)$ for every $t \geq 0$ (this is the *classical Feller property*),
- $L^p(S,\mu)$, $p \in [1,\infty]$, provided μ is a sub-invariant measure for (p_t).

We will see below that for obtaining a densely defined generator, an additional property called strong continuity is required for the semigroups. This will possibly exclude some of the Banach spaces above, for example, in most cases, the semigroup is not strongly continuous on $\mathcal{F}_b(S)$.

Before discussing strong continuity, we introduce another fundamental object that is useful to establish the connection between a semigroup and its generator: the resolvent.

Definition 4.3 (Resolvent kernels). The **resolvent kernels** associated to the transition function $(p_t)_{t\geq 0}$ are defined by

$$g_\alpha(x,dy) = \int_0^\infty e^{-\alpha t} p_t(x,dy) \, dt \qquad \text{for } \alpha \in (0,\infty),$$

i.e., for $f \in \mathcal{F}_+(S)$ or $f \in \mathcal{F}_b(S)$,

$$(g_\alpha f)(x) = \int_0^\infty e^{-\alpha t} (p_t f)(x) \, dt.$$

Remark. For any $\alpha \in (0,\infty)$, g_α is a kernel of positive measures on $(S,\mathcal{B})$. Analytically, g_α is the *Laplace transform* of the transition semigroup (p_t). Probabilistically, if (X_t, P_x) is a Markov process with transition function (p_t) then by Fubini's Theorem,

$$(g_\alpha f)(x) = E_x \left[\int_0^\infty e^{-\alpha t} f(X_t) dt \right].$$

In particular, $g_\alpha(x,B)$ is the average occupation time of a set B for the Markov process with start in x and constant *absorption rate* α. Note also that the resolvent resembles the Green's kernel, but in contrast to the latter, it is finite for recurrent Markov processes.

Lemma 4.4 (Sub-Markovian resolvent and contraction properties).

1) *The family $(g_\alpha)_{\alpha>0}$ is a* ***sub-Markovian resolvent*** *acting on $\mathcal{F}_b(S)$ or $\mathcal{F}_+(S)$ respectively, i.e., the following properties hold for any $\alpha, \beta > 0$:*
   - *(i)* Resolvent equation: $g_\alpha - g_\beta = (\beta - \alpha) g_\alpha g_\beta$,
   - *(ii)* Positivity preserving: $f \geq 0 \Rightarrow g_\alpha f \geq 0$,
   - *(iii)* $\alpha g_\alpha 1 \leq 1$.

2) ***Contractivity w.r.t. the supremum norm:*** *For every $\alpha > 0$,*

$$\|\alpha g_\alpha f\|_{sup} \leq \|f\|_{sup} \quad \text{for all } f \in \mathcal{F}_b(S).$$

3) ***Contractivity w.r.t. $\mathbf{L}^\mathbf{p}$ norms:*** *If $\mu \in \mathcal{M}_+(S)$ is sub-invariant w.r.t. (p_t) then*

$$\|\alpha g_\alpha f\|_{L^p(S,\mu)} \leq \|f\|_{L^p(S,\mu)} \quad \text{for any } \alpha > 0,\ p \in [1, \infty], \text{ and } f \in \mathcal{L}^p(S, \mu).$$

Proof. 1) By Fubini's Theorem and the semigroup property,

$$\begin{aligned} g_\alpha g_\beta f &= \int_0^\infty \int_0^\infty e^{-\alpha t} e^{-\beta s} p_{t+s} f \, ds \, dt \\ &= \int_0^\infty \int_0^u e^{(\beta-\alpha)t} dt \, e^{-\beta u} p_u f \, du \\ &= \frac{1}{\beta - \alpha} (g_\alpha f - g_\beta f) \end{aligned}$$

for any $\alpha, \beta > 0$ and $f \in \mathcal{F}_b(S)$. This proves (i). (ii) and (iii) follow easily from the corresponding properties for the semigroup (p_t).

2),3) Let $\|\cdot\|$ be either the supremum norm or an L^p norm. Then contractivity of $(p_t)_{t\geq 0}$ w.r.t. $\|\cdot\|$ implies that also (αg_α) is contractive w.r.t. $\|\cdot\|$:

$$\|\alpha g_\alpha f\| \leq \int_0^\infty \alpha e^{-\alpha t} \|p_t f\| dt \leq \int_0^\infty \alpha e^{-\alpha t} dt \, \|f\| = \|f\| \quad \text{for any } \alpha > 0.$$

■

The lemma shows that $(g_\alpha)_{\alpha>0}$ induces contraction resolvents of linear operators G_α on the Banach spaces $\mathcal{F}_b(S), C_b(S)$ if the semigroup (p_t) is Feller, $\hat{C}(S)$ if (p_t) is Feller in the classical sense, and $L^p(S, \mu)$ if μ is sub-invariant for (p_t). Furthermore, the resolvent equation implies that the range of the operators G_α is independent of α:

$$\text{Range}(G_\alpha) = \text{Range}(G_\beta) \qquad \text{for any } \alpha, \beta \in (0, \infty). \tag{4.3}$$

This property will be important below.

Strong continuity and generator

We now assume that $(P_t)_{t\geq 0}$ is a semigroup of linear contractions on a Banach space E. Our goal is to define the infinitesimal generator L of (P_t) by $Lf = \lim_{t\downarrow 0} \frac{1}{t}(P_t f - f)$ for a class $\mathcal{D}$ of elements $f \in E$ that forms a dense linear subspace of E. Obviously, this can only be possible if $\lim_{t\downarrow 0} \|P_t f - f\| = 0$ for all $f \in \mathcal{D}$, and hence, by contractivity of the operators P_t, for all $f \in E$. A semigroup with this property is called strongly continuous.

Definition 4.5 ($\mathbf{C^0}$ semigroup, Generator).

1) A semigroup $(P_t)_{t\geq 0}$ of linear operators on the Banach space E is called **strongly continuous** ($\mathbf{C^0}$) iff $P_0 = I$ and
$$\|P_t f - f\| \to 0 \quad \text{as } t \downarrow 0 \text{ for any } f \in E.$$

2) The **generator** of $(P_t)_{t\geq 0}$ is the linear operator $(L, \mathrm{Dom}(L))$ given by
$$Lf = \lim_{t\downarrow 0} \frac{P_t f - f}{t}, \quad \mathrm{Dom}(L) = \left\{ f \in E : \lim_{t\downarrow 0} \frac{P_t f - f}{t} \text{ exists} \right\}.$$
Here the limits are taken w.r.t. the norm on the Banach space E.

Remark (Strong continuity and domain of the generator). A contraction semigroup (P_t) is always strongly continuous on the closure of the domain of its generator. Indeed, $P_t f \to f$ as $t \downarrow 0$ for any $f \in \mathrm{Dom}(L)$, and hence for any $f \in \overline{\mathrm{Dom}(L)}$ by an $\varepsilon/3$ - argument. If the domain of the generator is dense in E then (P_t) is strongly continuous on E. Conversely, Theorem 4.9 below shows that the generator of a C^0 contraction semigroup is densely defined.

We have remarked above that a sub-Markov transition function $(p_t)_{t\geq 0}$ on $(S, \mathcal{B})$ induces a contraction semigroup $(P_t)_{t\geq 0}$ on the Banach space $\mathcal{F}_b(S)$. In general, however, this semigroup is not strongly continuous. To obtain strong continuity, we have to restrict to an appropriate closed subspace $E \subseteq \mathcal{F}_b(S)$ that is preserved by the semigroup. The maximal subspace we can choose is
$$E := \left\{ f \in \mathcal{F}_b(S) : \lim_{t\downarrow 0} \|P_t f - f\|_{\sup} = 0 \right\}. \tag{4.4}$$

Theorem 4.6 (Strong continuity of transition functions). The space E defined by (4.4) is a Banach space, and $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on E.

Proof. We first note that E is a closed subspace of $\mathcal{F}_b(S)$, and hence a Banach space. Indeed, let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions in E that converges uniformly to a function $f \in \mathcal{F}_b(S)$. Noting that
$$\|P_t f - f\|_{\sup} \leq \|P_t(f - f_n)\|_{\sup} + \|P_t f_n - f_n\|_{\sup} + \|f_n - f\|_{\sup},$$
and P_t is a contraction for every t, we see by an $\varepsilon/3$ argument that f is in E.

Since (P_t) is strongly continuous on E by definition, it now only remains to verify that $P_t(E) \subseteq E$ for every $t \geq 0$. This is the case, as for every $f \in E$,
$$\|P_s P_t f - P_t f\|_{\sup} = \|P_t(P_s f - f)\|_{\sup} \leq \|P_s f - f\|_{\sup} \longrightarrow 0$$
in the limit as $s \downarrow 0$. ■

The generator $(L, \mathrm{Dom}(L))$ of the semigroup $(P_t)_{t\geq 0}$ on the Banach space E can be seen as a *"full generator"* on bounded functions for a Markov process with transition function (p_t). For practical purposes, however, it is not always feasible to look at the semigroup on E and its generator, since the space E might be too large. If the transition kernels p_t are Feller, then it is often convenient to consider instead the semigroup restricted to the space of continuous functions vanishing at infinity, see Section 4.4 below. Alternatively, if we know a sub-invariant measure, we can also consider the induced semigroup and its generator on L^p spaces.

Exercise (Strong continuity of transition semigroups of Markov processes on L^p spaces). Suppose that $(p_t)_{t\geq 0}$ is the transition function of a *right-continuous*, time homogeneous Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$, and $\mu \in \mathcal{M}_+(S)$ is a sub-invariant measure.

a) Show that for every $f \in C_b(S)$ and $x \in S$,

$$(p_t f)(x) \to f(x) \qquad \text{as } t \downarrow 0.$$

b) Now let f be a non-negative function in $C_b(S) \cap \mathcal{L}^1(S,\mu)$ and $p \in [1,\infty)$. Show that as $t \downarrow 0$,

$$\int |p_t f - f| d\mu \ \to\ 0, \qquad \text{and hence} \qquad p_t f \to f \text{ in } L^p(S,\mu).$$

Hint: You may use that $|x| = x - 2x^-$.

c) Conclude that (p_t) induces a strongly continuous contraction semigroup of linear operators on $L^p(S,\mu)$ for every $p \in [1,\infty)$.

We now return to the general setup of C^0 contraction semigroups on Banach spaces.

Theorem 4.7 (Forward and backward equation). Suppose that $(P_t)_{t\geq 0}$ is a C^0 contraction semigroup with generator L. Then $t \mapsto P_t f$ is continuous for any $f \in E$. Moreover, if $f \in \mathrm{Dom}(L)$ then $P_t f \in \mathrm{Dom}(L)$ for all $t \geq 0$, and

$$\frac{d}{dt} P_t f = P_t L f = L P_t f,$$

where the derivative is a limit of difference quotients on the Banach space E.

The first statement explains why right continuity of $t \mapsto P_t f$ at $t = 0$ for any $f \in E$ is called strong continuity: For contraction semigroups, this property is indeed equivalent to continuity of $t \mapsto P_t f$ for $t \in [0,\infty)$ w.r.t. the norm on E.

Proof. 1) Continuity of $t \mapsto P_t f$ follows from the semigroup property, strong continuity and contractivity: For any $t > 0$,

$$\|P_{t+h} f - P_t f\| = \|P_t(P_h f - f)\| \leq \|P_h f - f\| \to 0 \qquad \text{as } h \downarrow 0,$$

and, similarly, for any $t > 0$,

$$\|P_{t-h} f - P_t f\| = \|P_{t-h}(f - P_h f)\| \leq \|f - P_h f\| \to 0 \qquad \text{as } h \downarrow 0.$$

2) Similarly, the forward equation $\frac{d}{dt} P_t f = P_t L f$ follows from the semigroup property, contractivity, strong continuity and the definition of the generator: For any $f \in \mathrm{Dom}(L)$ and $t \geq 0$,

$$\frac{1}{h}(P_{t+h} f - P_t f) = P_t \frac{P_h f - f}{h} \to P_t L f \qquad \text{as } h \downarrow 0,$$

and, for $t > 0$,

$$\frac{1}{-h}(P_{t-h} f - P_t f) = P_{t-h} \frac{P_h f - f}{h} \to P_t L f \qquad \text{as } h \downarrow 0$$

by strong continuity.

3) Finally, the backward equation $\frac{d}{dt} P_t f = L P_t f$ is a consequence of the forward equation: For $f \in \mathrm{Dom}(L)$ and $t \geq 0$,

$$\frac{P_h P_t f - P_t f}{h} = \frac{1}{h}(P_{t+h} f - P_t f) \to P_t L f \qquad \text{as } h \downarrow 0.$$

Hence $P_t f$ is in the domain of the generator, and $L P_t f = P_t L f = \frac{d}{dt} P_t f$. ■

Corollary 4.8 (From generator to martingale problem). Suppose that $((X_t)_{t\geq 0}, P)$ is a right continuous time homogeneous Markov process with transition function $(p_t)_{t\geq 0}$. Then $((X_t)_{t\geq 0}, P)$ is a solution of the martingale problem for the generator $(L, \mathrm{Dom}(L))$ of the C^0 contraction semigroup $(P_t)_{t\geq 0}$ which is induced by the transition function on the Banach space E defined by (4.4).

Proof. By Fubini's Theorem, the Markov property and the forward equation for (P_t),

$$E\left[f(X_{t+h}) - f(X_t) - \int_0^h (Lf)(X_{t+s})\, ds \middle| \mathcal{F}_t^X \right] \;=\; \left(P_h f - f - \int_0^h P_s L f\, ds \right)(X_t) \;=\; 0$$

holds almost surely for every $t, h \geq 0$ and every $f \in \mathrm{Dom}(L)$. ∎

One-to-one correspondences

Our next goal is to establish a 1-1 correspondence between C^0 contraction semigroups, generators and resolvents. Suppose that $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on a Banach space E with generator $(L, \mathrm{Dom}(L))$. By Theorem 4.7, $t \mapsto P_t f$ is a continuous function. Therefore, a corresponding resolvent can be defined as an E-valued Riemann integral:

$$G_\alpha f = \int_0^\infty e^{-\alpha t} P_t f\, dt \qquad \text{for any } \alpha > 0 \text{ and } f \in E. \tag{4.5}$$

Exercise (Strongly continuous contraction resolvent). Prove that the linear operators $G_\alpha, \alpha \in (0, \infty)$, defined by (4.5) form a *strongly continuous contraction resolvent*, i.e., for any $f \in E$,

(i) $G_\alpha f - G_\beta f = (\beta - \alpha) G_\alpha G_\beta f \quad$ for all $\alpha, \beta > 0$,

(ii) $\|\alpha G_\alpha f\| \leq \|f\| \quad$ for all $\alpha > 0$,

(iii) $\|\alpha G_\alpha f - f\| \to 0 \quad$ as $\alpha \to \infty$.

Theorem 4.9 (Connection between resolvent and generator). For every $\alpha > 0$, $G_\alpha = (\alpha I - L)^{-1}$. In particular, the domain of the generator coincides with the range of G_α, and it is dense in E.

Proof. Let $f \in E$ and $\alpha \in (0, \infty)$. We first show $G_\alpha f \in \mathrm{Dom}(L)$. Indeed, by strong continuity of $(P_t)_{t\geq 0}$,

$$\begin{aligned}
\frac{P_t G_\alpha f - G_\alpha f}{t} &= \frac{1}{t}\left(\int_0^\infty e^{-\alpha s} P_{t+s} f\, ds - \int_0^\infty e^{-\alpha s} P_s f\, ds \right) \\
&= \frac{e^{\alpha t} - 1}{t} \int_0^\infty e^{-\alpha s} P_s f\, ds - e^{\alpha t} \frac{1}{t} \int_0^t e^{-\alpha s} P_0 f\, ds \\
&\to \alpha G_\alpha f - f
\end{aligned}$$

as $t \downarrow 0$. Hence $G_\alpha f$ is contained in the domain of L, and

$$L G_\alpha f \;=\; \alpha G_\alpha f - f,$$

or, equivalently,

$$(\alpha I - L) G_\alpha f \;=\; f.$$

In a similar way it can be shown that for $f \in \mathrm{Dom}(L)$,

$$G_\alpha(\alpha I - L)f \;=\; f.$$

The details are left as an exercise. Hence $G_\alpha = (\alpha I - L)^{-1}$, and, in particular,

$$\mathrm{Dom}(L) \;=\; \mathrm{Dom}(\alpha I - L) \;=\; \mathrm{Range}(G_\alpha) \qquad \text{for any } \alpha > 0.$$

By strong continuity of the resolvent,

$$\alpha G_\alpha f \to f \qquad \text{as } \alpha \to \infty \qquad \text{for any } f \in E,$$

so the domain of L is dense in E. ■

The theorem above establishes a 1-1 correspondence between generators and resolvents. We now want to include the semigroup: We know how to obtain the generator from the semigroup, but to be able to go back, we have to show that a C^0 contraction semigroup is uniquely determined by its generator. This is one of the consequences of the following important theorem:

Theorem 4.10 (Duhamel's perturbation formula). Suppose that $(P_t)_{t\geq 0}$ and $(\widetilde{P}_t)_{t\geq 0}$ are C^0 contraction semigroups on E with generators L and $\widetilde{L}$, and assume that $\mathrm{Dom}(L) \subseteq \mathrm{Dom}(\widetilde{L})$. Then

$$\widetilde{P}_t f - P_t f \;=\; \int_0^t \widetilde{P}_s(\widetilde{L} - L)P_{t-s} f\, ds \qquad \text{for any } t \geq 0 \text{ and } f \in \mathrm{Dom}(L). \tag{4.6}$$

In particular, $(P_t)_{t\geq 0}$ is the only C^0 contraction semigroup with a generator that extends $(L, \mathrm{Dom}(L))$.

Proof. For $0 \leq s \leq t$ and $f \in \mathrm{Dom}(L)$ we have

$$P_{t-s} f \;\in\; \mathrm{Dom}(L) \;\subseteq\; \mathrm{Dom}(\widetilde{L})$$

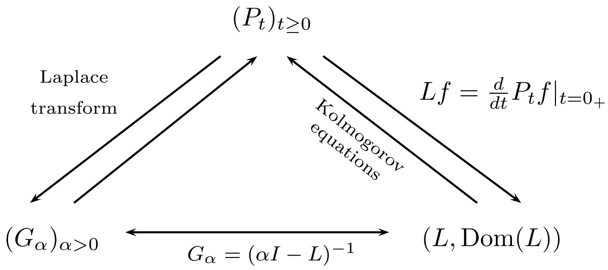
by Theorem 4.7. By combining the forward and backward equation in Theorem 4.7 we can then show that

$$\frac{d}{ds}\widetilde{P}_s P_{t-s} f \;=\; \widetilde{P}_s \widetilde{L} P_{t-s} f - \widetilde{P}_s L P_{t-s} f \;=\; \widetilde{P}_s(\widetilde{L} - L)P_{t-s} f,$$

where the derivative is taken in the Banach space E. The identity (4.6) now follows by the fundamental theorem of calculus for Banach-space valued functions, see for example [33].

In particular, if the generator of $\widetilde{P}_t$ is an extension of L then (4.6) implies that $P_t f = \widetilde{P}_t f$ for any $t \geq 0$ and $f \in \mathrm{Dom}(L)$. Since P_t and $\widetilde{P}_t$ are contractions, and the domain of L is dense in E by Theorem 4.9, this implies that the semigroups (P_t) and $(\widetilde{P}_t)$ coincide. ■

The last theorem shows that a C^0 contraction semigroup is uniquely determined if the generator and the full domain of the generator are known. The semigroup can then be reconstructed from the generator by solving the Kolmogorov equations. We summarize the correspondences in a diagram:



Example (Bounded generators). Suppose that L is a bounded linear operator on E. In particular, this is the case if L is the generator of a jump process with bounded jump intensities. For bounded linear operators, the semigroup can be obtained directly as an operator exponential

$$P_t \;=\; e^{tL} \;=\; \sum_{n=0}^{\infty} \frac{(tL)^n}{n!} \;=\; \lim_{n\to\infty} \left(1 + \frac{tL}{n}\right)^n,$$

where the series and the limit converge w.r.t. the operator norm. Alternatively,

$$P_t \;=\; \lim_{n\to\infty} \left(1 - \frac{tL}{n}\right)^{-n} \;=\; \lim_{n\to\infty} \left(\frac{n}{t} G_{\frac{n}{t}}\right)^n.$$

The last expression makes sense for unbounded generators as well. It tells us how to recover the semigroup from the resolvent.

Hille-Yosida-Theorem

We conclude this section with an important theoretical result showing which linear operators are generators of C^0 contraction semigroups. The proof will be sketched, cf. e.g. Ethier & Kurtz [19] for a detailed proof.

Theorem 4.11 (Hille, Yosida, Lumer, Phillips). A linear operator $(L, \mathrm{Dom}(L))$ on the Banach space E is the generator of a strongly continuous contraction semigroup if and only if the following conditions hold:

(i) $\mathrm{Dom}(L)$ is dense in E,

(ii) $\mathrm{Range}(\alpha I - L) = E$ for some $\alpha > 0$ (or, equivalently, for any $\alpha > 0$),

(iii) L is **dissipative**, i.e.,

$$\|\alpha f - Lf\| \geq \alpha \|f\| \qquad \text{for any } \alpha > 0 \text{ and } f \in \mathrm{Dom}(L).$$

Proof. "$\Rightarrow$": If L generates a C^0 contraction semigroup then by Theorem 4.9, $(\alpha I - L)^{-1} = G_\alpha$, where $(G_\alpha)_{\alpha>0}$ is the corresponding C^0 contraction resolvent. In particular, the domain of L is the range of G_α, and the range of $\alpha I - L$ is the domain of G_α. This shows that properties (i) and (ii) hold. Furthermore, any $f \in \mathrm{Dom}(L)$ can be represented as $f = G_\alpha g$ for some $g \in E$. Hence by contractivity of αG_α,

$$\alpha\|f\| = \|\alpha G_\alpha g\| \leq \|g\| = \|\alpha f - Lf\|.$$

"$\Leftarrow$": We only sketch this part of the proof. The key idea is to "regularize" the possibly unbounded linear operator L via the resolvent. By properties (ii) and (iii), the operator $\alpha I - L$ is invertible for any $\alpha > 0$, and the inverse $G_\alpha = (\alpha I - L)^{-1}$ is one-to-one from E onto the domain of L. Furthermore, it can be shown that $(G_\alpha)_{\alpha>0}$ is a C^0 contraction resolvent. Therefore, for any $f \in \mathrm{Dom}(L)$,

$$Lf = \lim_{\alpha\to\infty} \alpha G_\alpha L f = \lim_{\alpha\to\infty} L^{(\alpha)} f$$

where $L^{(\alpha)}$ is the *bounded* linear operator defined by

$$L^{(\alpha)} = \alpha L G_\alpha = \alpha^2 G_\alpha - \alpha I \qquad \text{for } \alpha \in (0,\infty).$$

Here we have used that L and G_α commute and $(\alpha I - L)G_\alpha = I$. The approximation by the bounded linear operators $L^{(\alpha)}$ is called the **Yosida approximation** of L. One verifies now that for every $\alpha > 0$, the operator exponentials

$$P_t^{(\alpha)} = e^{tL^{(\alpha)}} = \sum_{n=0}^{\infty} \frac{1}{n!} \left(tL^{(\alpha)}\right)^n, \quad t \in [0,\infty),$$

form a C^0 contraction semigroup with generator $L^{(\alpha)}$. Moreover, since for $f \in \mathrm{Dom}(L)$, $\left(L^{(\alpha)} f\right)_{\alpha\in\mathbb{N}}$ is a Cauchy sequence, Duhamel's formula (4.6) shows that also $\left(P_t^{(\alpha)} f\right)_{\alpha\in\mathbb{N}}$ is a Cauchy sequence for every $t \geq 0$. We can hence define

$$P_t f = \lim_{\alpha\to\infty} P_t^{(\alpha)} f \qquad \text{for any } t \geq 0 \text{ and } f \in \mathrm{Dom}(L). \tag{4.7}$$

Since $P_t^{(\alpha)}$ is a contraction for every t and α, P_t is a contraction, too. Since the domain of L is dense in E by Assumption (i), each P_t can be extended to a linear contraction on E, and (4.7) extends to $f \in E$. Now it can be verified that the limiting operators P_t form a C^0 contraction semigroup with generator L. ■

Exercise (Semigroups generated by self-adjoint operators on Hilbert spaces). Show that if E is a Hilbert space (for example an L^2 space) with norm $\|f\| = (f,f)^{1/2}$, and L is a *self-adjoint* linear operator, i.e.,

$$(L, \mathrm{Dom}(L)) = (L^*, \mathrm{Dom}(L^*)),$$

then L is the generator of a C^0 contraction semigroup on E if and only if L is *negative definite*, i.e.,

$$(f, Lf) \leq 0 \qquad \text{for any } f \in \mathrm{Dom}(L).$$

In this case, the C^0 semigroup generated by L is given by

$$P_t = e^{tL},$$

where the exponential is defined by spectral theory, see e.g. Reed & Simon [45, 46].

4.2. From the martingale problem to the generator

In the last section we have seen that there is a one-to-one correspondence between strongly continuous contraction semigroups on Banach spaces and their generators. The connection to Markov processes can be made via the martingale problem. Now suppose that we are given a *right-continuous time-homogeneous Markov process* $((X_t)_{t\in[0,\infty)}, (P_x)_{x\in S})$ with state space $(S,\mathcal{B})$ and transition semigroup $(p_t)_{t\geq 0}$. We assume that either E is a closed linear subspace of $\mathcal{F}_b(S)$ endowed with the supremum norm such that

(A1) $p_t(E) \subseteq E \quad$ for any $t \geq 0, \quad$ and

(A2) $\mu, \nu \in \mathcal{P}(S)$ with $\int f\,d\mu = \int f\,d\nu \;\forall f \in E \;\;\Rightarrow\;\; \mu = \nu$,

or $E = L^p(S,\mu)$ for some $p \in [1,\infty)$ and a (p_t)-sub-invariant measure $\mu \in \mathcal{M}_+(S)$.

From martingale problems to strongly continuous semigroups

In many situations it is known that for every $x \in S$, the process $((X_t)_{t\geq 0}, P_x)$ solves the martingale problem for some linear operator defined on "nice" functions on S. Hence let $\mathcal{A} \subset E$ be a dense linear subspace of the Banach space E, and let

$$\mathcal{L} : \mathcal{A} \subset E \to E$$

be a linear operator.

Theorem 4.12 (From the martingale problem to $\mathbf{C_0}$ semigroups). Suppose that for every $x \in S$ and $f \in \mathcal{A}$, the random variables $f(X_t)$ and $(\mathcal{L}f)(X_t)$ are integrable w.r.t. P_x for all $t \geq 0$, and the process

$$M_t^f = f(X_t) - \int_0^t (\mathcal{L}f)(X_s) ds$$

is an $(\mathcal{F}_t^X)$ martingale w.r.t. P_x. Then the transition function $(p_t)_{t\geq 0}$ induces a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ of linear operators on E, and the generator $(L, \mathrm{Dom}(L))$ of $(P_t)_{t\geq 0}$ is an extension of $(\mathcal{L}, \mathcal{A})$.

Remark (Processes with finite life-time). For Markov processes with finite life-time, the statement of Theorem 4.12 is still valid if functions $f : S \to \mathbb{R}$ are extended trivially to $S \cup \{\Delta\}$ by setting $f(\Delta) := 0$. This convention is always tacitly assumed below.

Proof. The martingale property for M^f w.r.t. P_x implies that the transition function (p_t) satisfies the forward equation. Indeed, for any $t \geq 0$, $x \in S$ and $f \in \mathcal{A}$,

$$\begin{aligned} (p_t f)(x) - f(x) &= E_x[f(X_t) - f(X_0)] = E_x\left[\int_0^t (\mathcal{L}f)(X_s)\, ds\right] \\ &= \int_0^t E_x[(\mathcal{L}f)(X_s)]\, ds = \int_0^t (p_s \mathcal{L}f)(x)\, ds. \end{aligned} \tag{4.8}$$

By the assumptions above and by Lemma 4.2, p_t is contractive w.r.t. the norm on E for any $t \geq 0$. Therefore, by (4.8),

$$\|p_t f - f\|_E \leq \int_0^t \|p_s \mathcal{L}f\|_E\, ds \leq t\|\mathcal{L}f\|_E \to 0 \quad \text{as } t \downarrow 0$$

for any $f \in \mathcal{A}$. Since $\mathcal{A}$ is a dense linear subspace of E, an $\varepsilon/3$ argument shows that the contraction semigroup (P_t) induced by (p_t) on E is strongly continuous. Furthermore, (4.8) implies that for any $f \in \mathcal{A}$,

$$\left\| \frac{p_t f - f}{t} - \mathcal{L}f \right\|_E \leq \frac{1}{t}\int_0^t \|p_s \mathcal{L}f - \mathcal{L}f\|_E\, ds \to 0 \quad \text{as } t \downarrow 0. \tag{4.9}$$

Here we have used that $\lim_{s\downarrow 0} p_s \mathcal{L}f = \mathcal{L}f$ by the strong continuity. By (4.9), the functions in $\mathcal{A}$ are contained in the domain of the generator L of (P_t), and $Lf = \mathcal{L}f$ for any $f \in \mathcal{A}$. ■

Identification of the generator and its domain

We now assume that L is the generator of a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ on E, and that $(L, \mathrm{Dom}(L))$ is an extension of $(\mathcal{L}, \mathcal{A})$. We have seen above that this is what can usually be deduced from knowing that the Markov process solves the corresponding martingale problem for every initial value $x \in S$. The next important question is whether the generator L and (hence) the C^0 semigroup (P_t) are already uniquely determined by the fact that L extends $(\mathcal{L}, \mathcal{A})$. In general the answer is negative - even though $\mathcal{A}$ is a dense subspace of E!

Exercise (Brownian motion with reflection and Brownian motion with killing). Suppose that $S = [0,\infty)$ or $S = (0,\infty)$, and let $E = L^2(S, dx)$. We consider the linear operator $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ with dense domain $\mathcal{A} = C_0^\infty(0,\infty) \subset L^2(S, dx)$. Let $((B_t)_{t\geq 0}, (P_x)_{x\in\mathbb{R}})$ be a canonical Brownian motion on $\mathbb{R}$. Then we can construct several Markov processes on S which induce C^0 contraction semigroups on E with generators that extends $(\mathcal{L}, \mathcal{A})$. In particular:

- *Brownian motion on* $\mathbb{R}_+$ *with reflection at 0* is defined on $S = [0,\infty)$ by
$$X_t^r = |B_t| \qquad \text{for any } t \geq 0. \tag{4.10}$$
- *Brownian motion on* $\mathbb{R}_+$ *with killing at 0* is defined on $S_\Delta = (0,\infty) \cup \{\Delta\}$ by
$$X_t^k = \begin{cases} B_t & \text{for } t < T_0^B, \\ \Delta & \text{for } t \geq T_0^B, \end{cases} \qquad \text{where } T_0^B = \inf\{t \geq 0 : B_t = 0\}. \tag{4.11}$$

Prove that both (X_t^r, P_x) and (X_t^k, P_x) are right-continuous Markov processes that induce C^0 contraction semigroups on $E = L^2(\mathbb{R}_+, dx)$. Moreover, show that both generators extend the operator $(\frac{1}{2}\frac{d^2}{dx^2}, C_0^\infty(0,\infty))$. In which sense do the generators differ from each other?

The exercise shows that it is not always enough to know the generator on a dense subspace of the corresponding Banach space E. Instead, what is really required for identifying the generator L, is to know its values on a subspace that is dense in the domain of L w.r.t. the graph norm

$$\|f\|_L := \|f\|_E + \|Lf\|_E.$$

Definition 4.13 (Closability and closure of linear operators).

1) A linear operator $(L, \mathrm{Dom}(L))$ on E is called **closed** iff $\mathrm{Dom}(L)$ is complete w.r.t. the graph norm $\|\cdot\|_L$.
2) The linear operator $(\mathcal{L}, \mathcal{A})$ is called **closable** iff it has a closed extension.
3) In this case, the smallest closed extension $(\overline{\mathcal{L}}, \mathrm{Dom}(\overline{\mathcal{L}}))$ is called the **closure** of $(\mathcal{L}, \mathcal{A})$. It is given explicitly by

$$\begin{aligned} \mathrm{Dom}(\overline{\mathcal{L}}) &= \text{completion of } \mathcal{A} \text{ w.r.t. the graph norm } \|\cdot\|_{\mathcal{L}}, \\ \overline{\mathcal{L}}f &= \lim_{n\to\infty} \mathcal{L}f_n \quad \text{for any sequence } (f_n)_{n\in\mathbb{N}} \text{ in } \mathcal{A} \text{ such that } f_n \to f \text{ in } E \\ & \qquad\qquad\qquad \text{and } (\mathcal{L}f_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence.} \end{aligned} \tag{4.12}$$

The operator $(\mathcal{L}, \mathcal{A})$ is closable if and only if $\mathcal{L}f_n \to 0$ for every sequence $f_n \in \mathcal{A}$ such that $f_n \to 0$ and $\mathcal{L}f_n$ is Cauchy. In this case, the closure is *well-defined*, i.e., the definition of $\overline{\mathcal{L}}f$ does not depend on the choice of the approximating sequence (f_n). Moreover, it is easy to verify that the extension defined by (4.12) is indeed the smallest closed extension of $(\mathcal{L}, \mathcal{A})$. Since the graph norm is stronger than the norm on E, the domain of the closure is a linear subspace of E. The graph of the closure is exactly the closure of the graph of the original operator in $E \times E$.

Example (Generator of heat semigroup on $L^p(\mathbb{R}, dx)$). For $p \in [1,\infty)$, the domain of the closure of the operator $\mathcal{L}f = f''/2$ with domain $\mathcal{A} = C_0^\infty(\mathbb{R})$ on the Banach space $L^p(\mathbb{R}, dx)$ is the Sobolev space $H^{2,p}(\mathbb{R}, dx)$.

Generators $(L, \mathrm{Dom}(L))$ of C^0 contraction semigroups are always closed. Indeed, by Theorem 4.9, $\alpha I - L$ is the inverse of the resolvent G_α. Since the resolvent is a bounded linear operator that is defined on the whole Banach space, it is closed. Thus the graph of the resolvent is a closed subset of $E \times E$, and hence the same holds for the graph of $\alpha I - L$. This shows that $(\alpha I - L, \mathrm{Dom}(L))$ and $(L, \mathrm{Dom}(L))$ are closed linear operators. There are operators that are not closable, but in the setup considered above we already know that there is a closed extension of $(\mathcal{L}, \mathcal{A})$ given by the generator $(L, \mathrm{Dom}(L))$.

Definition 4.14 (Operator core). Suppose that L is a linear operator on E with $\mathcal{A} \subseteq \mathrm{Dom}(L)$. Then $\mathcal{A}$ is called a **core** for L iff $\mathcal{A}$ is dense in $\mathrm{Dom}(L)$ w.r.t. the graph norm $\|\cdot\|_L$.

The subspace $\mathcal{A} \subseteq \mathrm{Dom}(L)$ is a core for L if and only if $(L, \mathrm{Dom}(L))$ is the closure of $(\mathcal{L}, \mathcal{A})$. In this case, for all $f \in E$ and $t, \alpha \in (0, \infty)$, the functions $P_t f$ and $G_\alpha f$ are contained in the completion $\overline{\mathcal{A}}^L$ of $\mathcal{A}$ w.r.t. the graph norm $\|\cdot\|_L$. The next theorem contains a converse statement that provides practical conditions to verify that a given subspace $\mathcal{A}$ of the domain is a core, and it shows that in this case, the semigroup is uniquely determined by the values of the generator on this subspace.

Theorem 4.15 (Strong uniqueness). Suppose that $\mathcal{A}$ is a dense linear subspace of the domain of the generator L with respect to the norm $\|\cdot\|_E$. Then the following statements are equivalent:

(i) $\mathcal{A}$ is a core for L.

(ii) There exists a dense linear subspace $\mathcal{A}_0$ of E such that for every $t \in (0, \infty)$ and every $f \in \mathcal{A}_0$, $P_t f$ is contained in the completion $\overline{\mathcal{A}}^L$ of $\mathcal{A}$ w.r.t. the graph norm $\|\cdot\|_L$.

(iii) There exist $\alpha > 0$ and a dense linear subspace $\mathcal{A}_0$ of E such that for every $f \in \mathcal{A}_0$, $G_\alpha f$ is contained in $\overline{\mathcal{A}}^L$.

(iv) There exists $\alpha > 0$ such that $(\alpha I - L)\left(\overline{\mathcal{A}}^L\right) = E$.

If the equivalent assertions (i)-(iv) are satisfied, then

(v) $(P_t)_{t \geq 0}$ is the only strongly continuous contraction semigroup on E with a generator that extends $(\mathcal{L}, \mathcal{A})$.

Proof. (i) $\Rightarrow$ (ii) holds since by Theorem 4.7, for any $t > 0$ and $f \in \mathrm{Dom}(L)$, $P_t f$ is contained in the domain of L.
(ii) $\Rightarrow$ (iii): Let $f \in \mathcal{A}_0$ and choose an arbitrary $\alpha \in (0, \infty)$. Approximating the Bochner integrals by Riemann sums shows that

$$\begin{aligned} G_\alpha f &= \int_0^\infty e^{-\alpha t} P_t f \, dt = \lim_{n \to \infty} g_n, \qquad \text{and} \\ L G_\alpha f &= \int_0^\infty e^{-\alpha t} P_t L f \, dt = \lim_{n \to \infty} L g_n, \qquad \text{where} \\ g_n &:= \sum_{k=0}^{n^2} \frac{1}{n} e^{-\alpha k/n} P_{k/n} f. \end{aligned}$$

Hence g_n converges to $G_\alpha f$ w.r.t. the graph norm of L. If (ii) holds then g_n is contained in $\overline{\mathcal{A}}^L$ for every $n \in \mathbb{N}$, and hence $G_\alpha f$ is in $\overline{\mathcal{A}}^L$ as well.

(iii) $\Rightarrow$ (iv): If (iii) holds then for all $f \in \mathcal{A}_0$, we have

$$f \;=\; (\alpha I - L)G_\alpha f \;\in\; (\alpha I - L)(\overline{\mathcal{A}}^L).$$

Since $(L, \overline{\mathcal{A}}^L)$ is a closed linear operator, we can conclude that $(\alpha I - L)(\overline{\mathcal{A}}^L)$ is a closed linear subspace of E that contains $\mathcal{A}_0$, and thus $(\alpha I - L)(\overline{\mathcal{A}}^L) = E$.
(iv) $\Rightarrow$ (i) and (v): Since $(L, \mathrm{Dom}(L))$ is the generator of a C^0 contraction semigroup (P_t), it a dissipative linear operator that extends $(L, \overline{\mathcal{A}}^L)$. Hence this operator is dissipative as well, and the Hille-Yosida theorem 4.9 implies that it also generates a C^0 contraction semigroup if (iv) is satisfied. By Theorem 4.10, both semigroups and their generators coincide, because the generator $(L, \mathrm{Dom}(L))$ is an extension of the other generator $(L, \overline{\mathcal{A}}^L)$. Thus $\overline{\mathcal{A}}^L = \mathrm{Dom}(L)$. More generally, suppose that $(\widetilde{P})_{t\geq 0}$ is an arbitrary C^0 contraction semigroup with a generator $\widetilde{L}$ extending $(\mathcal{L}, \mathcal{A})$. Then $\widetilde{L}$ is also an extension of L, because it is a closed linear operator by Theorem 4.9. Hence, by the same argument as above, we see that $(\widetilde{P}_t)$ and (P_t) agree. ■

Generators and boundary conditions

We now consider some examples of generators and their domains. At first, we apply Theorem 4.15 to identify exactly the domain of the generator of Brownian motion on $\mathbb{R}^n$. The transition semigroup of Brownian motion is the heat semigroup given by

$$(p_t f)(x) \;=\; (f * \varphi_t)(x) \;=\; \int_{\mathbb{R}^n} f(y)\,\varphi_t(x-y)\,dy \quad \text{for any } t \geq 0,$$

where $\varphi_t(x) = (2\pi t)^{-n/2} \exp\left(-|x|^2/(2t)\right)$.

Corollary 4.16 (Generator of Brownian motion). The transition function $(p_t)_{t\geq 0}$ of Brownian motion induces strongly continuous contraction semigroups on $\hat{C}(\mathbb{R}^n)$ and on $L^p(\mathbb{R}^n, dx)$ for every $p \in [1, \infty)$. The generators of these semigroups are given by

$$L = \frac{1}{2}\Delta, \qquad \mathrm{Dom}(L) = \overline{C_0^\infty(\mathbb{R}^n)}^\Delta,$$

where $\overline{C_0^\infty(\mathbb{R}^n)}^\Delta$ stands for the completion of $C_0^\infty(\mathbb{R}^n)$ w.r.t. the graph norm of the Laplacian on the underlying Banach space $\hat{C}(\mathbb{R}^n), L^p(\mathbb{R}^n, dx)$ respectively. In particular, the domain of L contains all C^2 functions with derivatives up to second order in $\hat{C}(\mathbb{R}^n), L^p(\mathbb{R}^n, dx)$, respectively.

Proof. By Itô's formula or by direct computation, Brownian motion (B_t, P_x) solves the martingale problem for the operator $\frac{1}{2}\Delta$ with domain $C_0^\infty(\mathbb{R}^n)$. Moreover, $p_t(\hat{C}(S)) \subseteq \hat{C}(S)$ holds for any $t \geq 0$, and Lebesgue measure is invariant for the transition kernels p_t, since by Fubini's theorem,

$$\int_{\mathbb{R}^n} p_t f\,dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \varphi_t(x-y) f(y)\,dy\,dx = \int_{\mathbb{R}^n} f(y)\,dy \quad \text{for any } f \in \mathcal{F}_+(\mathbb{R}^n).$$

Hence by Theorem 4.12, $(p_t)_{t\geq 0}$ induces C^0 contraction semigroups on $\hat{C}(S)$ and on $L^p(\mathbb{R}^n, dx)$ for $p \in [1, \infty)$, and the generators are extensions of $\left(\frac{1}{2}\Delta, C_0^\infty(\mathbb{R}^n)\right)$.

Moreover, a standard approximation argument shows that the completions $\overline{C_0^\infty(\mathbb{R}^n)}^\Delta$ w.r.t. the graph norms contain all functions in $C^2(\mathbb{R}^n)$ with derivatives up to second order in $\hat{C}(\mathbb{R}^n), L^p(\mathbb{R}^n, dx)$, respectively. Therefore, $p_t f = f * \varphi_t$ is contained in $\overline{C_0^\infty(\mathbb{R}^n)}^\Delta$ for any $f \in C_0^\infty(\mathbb{R}^n)$ and $t \geq 0$. Hence, by Theorem 4.15, the generators on $\hat{C}(S)$ and $L^p(\mathbb{R}^n, dx)$ coincide with the closures of $\left(\frac{1}{2}\Delta, C_0^\infty(\mathbb{R}^n)\right)$. ■

Example (Generator of Brownian motion on $\mathbb{R}$). In the one-dimensional case, the generators on $L^p(\mathbb{R}, dx)$ and on $\hat{C}(\mathbb{R})$ are given explicitly by

$$Lf = \frac{1}{2}f'', \ \mathrm{Dom}(L) = \left\{f \in L^p(\mathbb{R}, dx) \cap C^1(\mathbb{R}) : f' \text{ absolutely continuous}, \ f'' \in L^p(\mathbb{R}, dx)\right\}, \quad (4.13)$$

$$Lf = \frac{1}{2}f'', \ \mathrm{Dom}(L) = \left\{f \in \hat{C}(\mathbb{R}) \cap C^2(\mathbb{R}) : f'' \in \hat{C}(\mathbb{R})\right\}, \quad \text{respectively.} \quad (4.14)$$

Example (Domain of generator in multi-dimensional case, Sobolev spaces). In dimensions $n \geq 2$, the domains of the generators contain functions that are not twice differentiable in the classical sense. The domain of the L^p generator is the Sobolev space $H^{2,p}(\mathbb{R}^n, dx)$ consisting of *weakly* twice differentiable functions with derivatives up to second order in $L^p(\mathbb{R}^n, dx)$, see e.g. [20].

Next, we consider Brownian motion on $\mathbb{R}_+$ with different boundary conditions at 0. For a function $f : [0, \infty) \to \mathbb{R}$, we denote by $\widetilde{f}(x) := f(|x|)$ and $\hat{f}(x) := f(|x|)\,\mathrm{sgn}(x)$ its symmetric and antisymmetric extension to $\mathbb{R}$. Note that the transition semigroup (p_t^r) of Brownian motion with reflection at 0 is given by

$$(p_t^r f)(x) \;=\; E_x\left[f(|B_t|)\right] \;=\; E_x\left[\widetilde{f}(B_t)\right] \;=\; (p_t\widetilde{f})(x) \quad (4.15)$$

for any $x \in [0, \infty)$ and $f \in \mathcal{F}_b([0, \infty))$. Here, (p_t) denotes the transition semigroup of Brownian motion on $\mathbb{R}$. Similarly, the strong Markov property shows that the transition semigroup of Brownian motion with killing at 0 is given by

$$(p_t^k f)(x) \;=\; E_x\left[f(B_t); t < T_0\right] \;=\; E_x\left[\hat{f}(B_t); t < T_0\right] \;=\; E_x\left[\hat{f}(B_t)\right] \;=\; (p_t\hat{f})(x). \quad (4.16)$$

In particular, the functions in the range of p_t^r are restrictions of smooth symmetric functions on $\mathbb{R}$ to $[0, \infty)$, and thus their first derivative at 0 vanishes. Correspondingly, the functions in the range of p_t^k are restrictions of smooth antisymmetric functions on $\mathbb{R}$, and thus at 0, both the functions and their second derivatives vanish. Since $p_t^r f$ and $p_t^k f$ take values in the domains of the corresponding generators, we can expect that these domains also consist of functions with corresponding properties.

Corollary 4.17 (L^p generators of Brownian motions with reflection and killing at 0). For every $p \in [1, \infty)$, the following domains are cores for the generators L^r and L^k of Brownian motion with reflection and killing at 0 on the Banach space $L^p(\mathbb{R}_+, dx)$:

$$\begin{aligned} \mathcal{A}^r &= \left\{f \in C_0^2([0, \infty)) \;:\; f'(0) = 0\right\}, \\ \mathcal{A}^k &= \left\{f \in C_0^2([0, \infty)) \;:\; f(0) = 0\right\}. \end{aligned}$$

The generators are determined by $L^r f = f''/2$ for all $f \in \mathcal{A}^r$, and $L^k f = f''/2$ for all $f \in \mathcal{A}^k$.

Proof. We first show $\mathcal{A}^r \subseteq \mathrm{Dom}(L^r)$. Let $f \in \mathcal{A}^r$. Since $f'(0) = 0$, the symmetric extension $\widetilde{f}$ is contained in $C_0^1(\mathbb{R})$. Moreover, the derivative $\widetilde{f}'$ is continuously differentiable except possibly at 0, where it is continuous. Therefore, $\widetilde{f}$ is contained in the Sobolev space $H^{2,p}(\mathbb{R}, dx)$, which is the domain of the generator L of Brownian motion on $L^p(\mathbb{R}, dx)$. Now, (4.15) implies that f is contained in the domain of L^r, and

$$L^r f \;=\; (L\widetilde{f})\Big|_{\mathbb{R}_+} \;=\; \frac{1}{2}f''.$$

To verify that $\mathcal{A}^r$ is a core for L^r, we observe that by (4.15), for every $f \in L^p(\mathbb{R}, dx)$, $p_t^r f$ is C^2 on $[0, \infty)$ with $(p_t^r f)'(0) = 0$, and its derivatives up to order 2 are in $L^p(\mathbb{R}_+, dx)$. Therefore, it can be approximated by functions in $\mathcal{A}^r$ w.r.t. the graph norm of L^r. Therefore, by Theorem 4.15, $\mathcal{A}^r$ is indeed a core for L^r. The corresponding statement for L^k follows similarly from (4.16). ■

One might expect a similar description for the generators of Brownian motion on spaces of continuous functions. This turns out to be true for the generator of reflected Brownian motion, but in the case of Brownian motion with killing a subtle difference occurs.

Remark (Feller generators of Brownian motions with reflection and killing at 0). By similar arguments as above, Brownian motion with reflection at 0 induces a C^0 contraction semigroup on the Banach space $E = \hat{C}([0,\infty))$ with generator given by

$$L^r f \;=\; f''/2, \qquad \mathrm{Dom}(L^r) \;=\; \left\{ f \in C^2([0,\infty)) \cap E : f'(0) = 0, f'' \in E \right\}.$$

On the contrary, Brownian motion with killing at 0 does *not* induce a C^0 contraction semigroup on E, because it can not be started at 0 without loosing right continuity. Instead, it does induce a C^0 contraction semigroup on the Banach space $E_0 = \hat{C}((0,\infty))$ consisting of continuous functions vanishing both at 0 and at ∞. The generator on this Banach space is given by

$$L^k f \;=\; f''/2, \qquad \mathrm{Dom}(L^k) \;=\; \left\{ f \in C^2((0,\infty)) \cap E_0 : f'' \in E_0 \right\}.$$

Note that the functions in the domain of L^k satisfy the additional boundary condition $f''(0) = 0$, whereas the boundary condition $f(0) = 0$ is automatically satisfied for all functions in the Banach space E_0.

Exercise (Brownian motion with absorption at 0). Brownian motion with absorption at 0 is the Markov process with state space $S = [0,\infty)$ defined by $X_t = B_{t\wedge T_0}$ where (B_t, P_x) is a Brownian motion on $\mathbb{R}$. On which Banach spaces does this process induce C^0 contraction semigroups? Identify the corresponding generators!

4.3. From infinitesimal information to finite time properties

The infinitesimal behaviour of a time homogeneous Markov process is encoded in its generator. Often, we want to deduce properties of the Markov process from this infinitesimal information. In general, this requires knowing the generator on a core. In this section we prove an infinitesimal characterization of invariant probability measures, and we show that the solution of a martingale problem is uniquely determined by the values of the generator on an operator core. Infinitesimal characterizations of long-time stability properties of Markov processes will be studied later in Chapter 9 below.

Throughout this section, we assume that E is a closed linear subspace of $\mathcal{F}_b(S)$ satisfying (A2). Let L be the generator of a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ on E, and let $\mathcal{A}$ be a linear subspace of the domain of L.

Characterization of invariant probability measures

A probability measure μ on $(S,\mathcal{B})$ is called **invariant** for the semigroup (P_t) if and only if

$$\int P_t f \, d\mu \;=\; \int f \, d\mu \qquad \text{for all } t \in [0,\infty) \text{ and } f \in E. \tag{4.17}$$

Note that by (A2), this definition is consistent with our usual definition of invariant probability measures for Markov processes. Indeed, if (P_t) is induced by the transition semigroup (p_t) of a time homogeneous Markov process then (4.17) holds if and only if $\mu = \mu p_t$ for every $t \geq 0$.

Theorem 4.18 (Infinitesimal characterization of invariant probility measures). A probability measure μ is invariant for (P_t) if and only if there exists a core $\mathcal{A}$ of the generator L such that

$$\int Lf \, d\mu \;=\; 0 \qquad \text{for all } f \in \mathcal{A}. \tag{4.18}$$

Proof. If μ is invariant for (P_t) then for every $f \in E$ and $t > 0$,

$$\int \frac{P_t f - f}{t}\, d\mu \;=\; 0.$$

For $f \in \mathrm{Dom}(L)$, the difference quotient converges uniformly to Lf as $t \downarrow 0$, and hence (4.18) holds.

Conversely, if $\mathcal{A}$ is a core for L then for every $f \in \mathrm{Dom}(L)$, there exists a sequence $f_n \in \mathcal{A}$ such that $Lf_n \to Lf$ uniformly. Thus if (4.18) holds then $\int Lf\, d\mu = 0$ for all $f \in \mathrm{Dom}(L)$. The backward equation thus implies

$$\frac{d}{dt}\int P_t f\, d\mu \;=\; \int LP_t f\, d\mu \;=\; 0,$$

and thus $\int P_t f\, d\mu = \int f\, d\mu$ for all $f \in \mathrm{Dom}(L)$. This implies invariance of μ since the domain of the generator is dense in E w.r.t. the supremum norm. ■

Example (Brownian motion with absorption at the boundary). A Brownian motion on the interval $[0,1]$ that is absorbed when reaching the boundary is defined by $X_t = B_{t\wedge T}$ where (B_t, P_x) is a Brownian motion on $\mathbb{R}$ starting at $x \in [0,1]$, and

$$T \;=\; \inf\{t \geq 0 : B_t \in \{0,1\}\}.$$

The process (X_t, P_x) is a Markov process with continuous paths that solves the martingale problem for the operator $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ with domain $\mathcal{A} = C_0^\infty((0,1))$. The uniform distribution μ on $[0,1]$ is infinitesimally invariant in the sense that

$$\int \mathcal{L}f\, d\mu \;=\; \frac{1}{2}\int_0^1 f''(x)\, dx \;=\; 0 \qquad \text{for any } f \in C_0^\infty((0,1)).$$

Nevertheless, μ is not invariant for the corresponding transition function (p_t) as P_μ-almost surely, $X_t \in \{0,1\}$ eventually.

The example shows that the assumption that $\mathcal{A}$ is a core for the generator can not be relaxed easily. Similar phenomena can arise by boundaries at infinity, singularities, and in other ways.

Exercise (Infinitesimal characterization of invariant measures - Another counterexample). Consider the minimal time-homogeneous Markov jump process (X_t, P_x) with state space $\mathbb{Z}$ and generator $\mathcal{L} = \lambda\,(\pi - I)$, where

$$\lambda(x) \;=\; 1 + x^2 \qquad \text{and} \qquad \pi(x,\cdot) \;=\; \delta_{x+1} \qquad \text{for all } x \in \mathbb{Z}.$$

a) Show that the probability measure μ with weights $\mu(x) \propto 1/(1+x^2)$ is infinitesimally invariant in the sense that

$$(\mu\mathcal{L})(y) \;=\; 0 \qquad \text{for all } y \in \mathbb{Z}.$$

b) Show that nevertheless, μ is not an invariant measure for the transition semigroup (p_t) of the process.

Uniqueness of martingale problems

The next theorem shows that a solution to the martingale problem for $(L, \mathcal{A})$ with given initial distribution is unique if $\mathcal{A}$ is a core for L.

Theorem 4.19 (Markov property and uniqueness for solutions of martingale problem). Suppose that $\mathcal{A}$ is a core for L. Then any solution $((X_t)_{t\geq 0}, P)$ of the martingale problem for $(L, \mathcal{A})$ is a Markov process with transition function determined uniquely by

$$p_t f = P_t f \qquad \text{for any } t \geq 0 \text{ and } f \in E. \tag{4.19}$$

In particular, all right-continuous solutions of the martingale problem for $(L, \mathcal{A})$ with given initial distribution $\mu \in \mathcal{P}(S)$ coincide in law.

Proof. We only sketch the main steps in the proof. For a detailed proof see Ethier and Kurtz [19, Chapter 4, Theorem 4.1].

Step 1: If the process (X_t, P) solves the martingale problem for $(L, \mathcal{A})$ then an approximation based on the assumption that $\mathcal{A}$ is dense in $\mathrm{Dom}(L)$ w.r.t. the graph norm shows that (X_t, P) also solves the martingale problem for $(L, \mathrm{Dom}(L))$. Therefore, we may assume w.l.o.g. that $\mathcal{A} = \mathrm{Dom}(L)$.

Step 2: **Extended martingale problem.** The fact that (X_t, P) solves the martingale problem for $(L, \mathcal{A})$ implies that the process

$$M_t^{[f,\alpha]} := e^{-\alpha t} f(X_t) + \int_0^t e^{-\alpha s}(\alpha f - Lf)(X_s)ds$$

is a martingale for any $\alpha \geq 0$ and $f \in \mathcal{A}$. The proof can be carried out directly by Fubini's Theorem or via the product rule from Stieltjes calculus. The latter shows that

$$e^{-\alpha t} f(X_t) - f(X_0) = \int_0^t e^{-\alpha s}(Lf - \alpha f)(X_s)ds + \int_0^t e^{-\alpha s} dM_s^{[f]}$$

where $\int_0^t e^{-\alpha s} dM_s^{[f]}$ is an Itô integral w.r.t. the martingale $M_t^f = f(X_t) - \int_0^t (Lf)(X_s)ds$, and hence a martingale, cf. [17].

Step 3: **Markov property in resolvent form.** Applying the martingale property to the martingales $M^{[f,\alpha]}$ shows that for any $s \geq 0$ and $g \in E$,

$$E\left[\int_0^\infty e^{-\alpha t} g(X_{s+t}) \middle| \mathcal{F}_s^X\right] = (G_\alpha g)(X_s) \quad P\text{-a.s.} \tag{4.20}$$

Indeed, let $f = G_\alpha g$. Then f is contained in the domain of L, and $g = \alpha f - Lf$. Therefore, for $s, t \geq 0$,

$$\begin{aligned} 0 &= E\left[M_{s+t}^{[f,\alpha]} - M_s^{[f,\alpha]} \middle| \mathcal{F}_s^X\right] \\ &= e^{-\alpha(s+t)} E\left[f(X_{s+t}) | \mathcal{F}_s^X\right] - e^{-\alpha s} f(X_s) + E\left[\int_0^t e^{-\alpha(s+r)} g(X_{s+r}) dr \middle| \mathcal{F}_s^X\right] \end{aligned}$$

holds almost surely. The identity (4.20) follows as $t \to \infty$.

Step 4: **Markov property in semigroup form.** One can now conclude that

$$E[g(X_{s+t}) | \mathcal{F}_s^X] = (P_t g)(X_s) \quad P\text{-a.s.} \tag{4.21}$$

holds for any $s, t \geq 0$ and $g \in E$. The proof is based on the approximation

$$P_t g \;=\; \lim_{n \to \infty} \left(\frac{n}{t} G_{\frac{n}{t}}\right)^n g$$

of the semigroup by the resolvent, see the exercise below.

Step 5: **Conclusion.** By Step 4 and Assumption (A2), the process $((X_t), P)$ is a Markov process with transition semigroup $(p_t)_{t \geq 0}$ satisfying (4.19). In particular, the transition semigroup and (hence) the law of the process with given initial distribution are uniquely determined. ■

Exercise (Approximation of semigroups by resolvents). Suppose that $(P_t)_{t\geq 0}$ is a Feller semigroup with resolvent $(G_\alpha)_{\alpha>0}$. Prove that for any $t > 0, n \in \mathbb{N}$ and $x \in S$,

$$\left(\frac{n}{t} G_{\frac{n}{t}}\right)^n g(x) = E\left[P_{\frac{E_1+\cdots+E_n}{n}t}\, g(x)\right]$$

where $(E_k)_{k\in\mathbb{N}}$ is a sequence of independent exponentially distributed random variables with parameter 1. Hence conclude that

$$\left(\frac{n}{t} G_{\frac{n}{t}}\right)^n g \to P_t g \quad \text{uniformly as } n \to \infty. \tag{4.22}$$

How could you derive (4.22) more directly when the state space is finite?

Remark (Other uniqueness results for martingale problems). It is often not easy to verify the assumption that $\mathcal{A}$ is a core for L in Theorem 4.19. Further uniqueness results for martingale problems with assumptions that may be easier to verify in applications can be found in Stroock/Varadhan [51] and Ethier/Kurtz [19].

Strong Markov property

In Theorem 4.19 we have used the Markov property to establish uniqueness. The next theorem shows conversely that under modest additional conditions, the strong Markov property for solutions is a consequence of uniqueness of martingale problems.

Let $\mathcal{D}(\mathbb{R}_+, S)$ denote the space of all càdlàg (right continuous with left limits) functions $\omega : [0,\infty) \to S$. If S is a polish space then $\mathcal{D}(\mathbb{R}_+, S)$ is again a polish space w.r.t. the **Skorokhod topology**, see e.g. Billingsley [3]. Furthermore, the Borel σ-algebra on $\mathcal{D}(\mathbb{R}_+, S)$ is generated by the evaluation maps $X_t(\omega) = \omega(t),\ t \in [0,\infty)$.

Theorem 4.20 (Uniqueness of martingale problem ⇒ Strong Markov property). Suppose that the following conditions are satisfied:

(i) $\mathcal{A}$ is a linear subspace of $C_b(S)$, and $\mathcal{L} : \mathcal{A} \to \mathcal{F}_b(S)$ is a linear operator such that $\mathcal{A}$ is separable w.r.t. $\|\cdot\|_{\mathcal{L}}$.

(ii) For every $x \in S$ there is a unique probability measure P_x on $\mathcal{D}(\mathbb{R}_+, S)$ such that the canonical process $((X_t)_{t\geq 0}, P_x)$ solves the martingale problem for $(\mathcal{L}, \mathcal{A})$ with initial value $X_0 = x$ P_x-a.s.

(iii) The map $x \mapsto P_x[A]$ is measurable for any Borel set $A \subseteq \mathcal{D}(\mathbb{R}_+, S)$.

Then $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ is a **strong Markov process**, i.e.,

$$E_x\left[F(X_{T+\cdot})|\mathcal{F}_T^X\right] = E_{X_T}[F] \quad P_x\text{-a.s.}$$

for any $x \in S, F \in \mathcal{F}_b(\mathcal{D}(\mathbb{R}_+, S))$, and any finite $(\mathcal{F}_t^X)$ stopping time T.

Remark (Non-uniqueness). If uniqueness does not hold then one can not expect that any solution of a martingale problem is a Markov process, because different solutions can be combined in a non-Markovian way (e.g. by switching from one to the other when a certain state is reached).

Proof (Sketch of proof of Theorem 4.20). Fix $x \in S$. Since $\mathcal{D}(\mathbb{R}_+, S)$ is again a polish space there is a regular version $(\omega, A) \mapsto Q_\omega(A)$ of the conditional distribution $P_x[\,\cdot\,|\mathcal{F}_T]$. Suppose we can prove the following statement:
Claim: For P_x-almost every ω, the process $(X_{T+\cdot}, Q_\omega)$ solves the martingale problem for $(\mathcal{L}, \mathcal{A})$ w.r.t. the

filtration $(\mathcal{F}^X_{T+t})_{t\geq 0}$.
Then we are done, because of the martingale problem with initial condition $X_T(\omega)$ now implies

$$(X_{T+\cdot}, Q_\omega) \sim (X, P_{X_T(\omega)}) \quad \text{for } P_x\text{-a.e. } \omega,$$

which is the strong Markov property.

The reason why we can expect the claim to be true is that for any given $0 \leq s < t$, $f \in \mathcal{A}$ and $A \in \mathcal{F}^X_{T+s}$,

$$\begin{aligned}
&E_{Q_\omega}\left[f(X_{T+t}) - f(X_{T+s}) - \int_{T+s}^{T+t} (\mathcal{L}f)(X_r)dr; A \right] \\
&= E_x\left[\left(M^{[f]}_{T+t} - M^{[f]}_{T+s} \right) 1_A \middle| \mathcal{F}^X_T \right](\omega) \\
&= E_x\left[E_x\left[M^{[f]}_{T+t} - M^{[f]}_{T+s} \middle| \mathcal{F}^X_{T+s} \right] 1_A \middle| \mathcal{F}^X_T \right](\omega) = 0
\end{aligned}$$

holds for P_x-a.e. ω by the optional sampling theorem and the tower property of conditional expectations. However, this is not yet a proof since the exceptional set depends on s, t, f and A. To turn the sketch into a proof one has to use the separability assumptions to show that the exceptional set can be chosen independently of these objects, cf. Stroock and Varadhan [51], Rogers and Williams [48, 49], or Ethier and Kurtz [19]. ■

4.4. Feller processes and their generators

In this section we restrict ourselves to **Feller processes**. These are càdlàg Markov processes with a locally compact separable state space S whose transition semigroup preserves $\hat{C}(S)$. We will establish a one-to-one correspondence between sub-Markovian C^0 semigroups on $\hat{C}(S)$, their generators, and Feller processes. Moreover, we will show that the generator L of a Feller process with continuous paths on $\mathbb{R}^n$ acts as a second order differential operator on functions in $C_0^\infty(\mathbb{R}^n)$ if this is a subspace of the domain of L.

Feller semigroups

We start with a definition:

Definition 4.21 (Feller semigroup). A **Feller semigroup** is a sub-Markovian C^0 semigroup $(P_t)_{t\geq 0}$ of linear operators on $\hat{C}(S)$, i.e., a Feller semigroup has the following properties that hold for any $f \in \hat{C}(S)$:

(i) **Strong continuity:** $\|P_t f - f\|_{sup} \to 0$ as $t \downarrow 0$,

(ii) **Sub-Markov:** $f \geq 0 \Rightarrow P_t f \geq 0$, $f \leq 1 \Rightarrow P_t f \leq 1$,

(iii) **Semigroup:** $P_0 f = f$, $P_t P_s f = P_{t+s} f$ for any $s, t \geq 0$.

Remark. Property (ii) implies that P_t is a contraction w.r.t. the supremum norm for any $t \geq 0$.

Let us now assume again that $(p_t)_{t\geq 0}$ is the transition function of a *right-continuous* time homogeneous Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ defined for any initial value $x \in S$. We have shown above that (p_t) induces contraction semigroups on different Banach spaces consisting of functions (or equivalence classes of functions) from S to $\mathbb{R}$. The following example shows, however, that these semigroups are not necessarily strongly continuous:

Example (Strong continuity of the heat semigroup). Let $S = \mathbb{R}^1$. The heat semigroup (p_t) is the transition semigroup of Brownian motion on S. It is given explicitly by

$$(p_t f)(x) = (f * \varphi_t)(x) = \int_{\mathbb{R}} f(y)\varphi_t(x-y)\,dy,$$

where $\varphi_t(z) = (2\pi t)^{-1/2}\exp\left(-z^2/(2t)\right)$ is the density of the normal distribution $N(0,t)$. The heat semigroup induces contraction semigroups on the Banach spaces $\mathcal{F}_b(\mathbb{R}), C_b(\mathbb{R}), \hat{C}(\mathbb{R})$ and $L^p(\mathbb{R}, dx)$ for $p \in [1,\infty]$. However, the semigroups on $\mathcal{F}_b(\mathbb{R}), C_b(\mathbb{R})$ and $L^\infty(\mathbb{R}, dx)$ are not strongly continuous. Indeed, since $p_t f$ is a continuous function for any $f \in \mathcal{F}_b(\mathbb{R})$,

$$\|p_t 1_{(0,1)} - 1_{(0,1)}\|_\infty \geq \frac{1}{2} \qquad \text{for any } t > 0.$$

This shows that strong continuity fails on $\mathcal{F}_b(\mathbb{R})$ and on $L^\infty(\mathbb{R}, dx)$. To see that (p_t) is not strongly continuous on $C_b(\mathbb{R})$ either, we may consider the function $f(x) = \sum_{n=1}^{\infty} \exp\left(-2^n(x-n)^2\right)$. It can be verified that $\limsup_{x\to\infty} f(x) = 1$ whereas for any $t > 0$, $\lim_{x\to\infty}(p_t f)(x) = 0$. Hence $\|p_t f - f\|_{\sup} \geq 1$ for any $t > 0$. Theorem 4.22 below shows that the semigroup induced by (p_t) on the Banach space $\hat{C}(\mathbb{R})$ is strongly continuous.

The example explains why we consider the transition semigroup of a Feller process on the Banach space $\hat{C}(S)$ and not, for instance, on the space of all bounded continuous functions. The next theorem shows that the transition semigroup of a Feller process on S is always a Feller semigroup:

Theorem 4.22 (Strong continuity of transition functions of Feller processes). Suppose that $(p_t)_{t\geq 0}$ is the transition function of a right-continuous time-homogeneous Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ such that $p_t\left(\hat{C}(S)\right) \subseteq \hat{C}(S)$ for any $t \geq 0$. Then $(p_t)_{t\geq 0}$ induces a Feller semigroup $(P_t)_{t\geq 0}$ on $\hat{C}(S)$. If L denotes the generator then the process $((X_t), P_x)$ solves the martingale problem for $(L, \mathrm{Dom}(L))$ for any $x \in S$.

Proof. We prove strong continuity. Filling in the other missing details is left as an exercise. To prove strong continuity, we proceed in several steps:

1) For $f \in \hat{C}(S)$, the function $t \mapsto f(X_t)$ is almost surely right continuous and bounded. Therefore, by dominated convergence, for any $x \in S$,

$$(p_t f)(x) \;=\; E_x\left[f(X_t)\right] \;\to\; E_x\left[f(X_0)\right] \;=\; f(x) \qquad \text{as } t \downarrow 0. \tag{4.23}$$

2) Now suppose that $f = g_\alpha h = \int_0^\infty e^{-\alpha s} p_s h\,ds$ for some $\alpha > 0$ and a function h in $\hat{C}(S)$. Then for $t \geq 0$,

$$p_t f \;=\; \int_0^\infty e^{-\alpha s} p_{s+t} h\,ds \;=\; e^{\alpha t}\int_t^\infty e^{-\alpha u} p_u h\,du \;=\; e^{\alpha t} f - e^{\alpha t}\int_0^t e^{-\alpha u} p_u h\,du,$$

and hence

$$\|p_t f - f\|_{\sup} \;\leq\; (e^{\alpha t} - 1)\|f\|_{\sup} + e^{\alpha t}\int_0^t \|p_u h\|_{\sup}\,du.$$

Since $\|p_u h\|_{\sup} \leq \|h\|_{\sup}$, the right-hand side converges to 0 as $t \downarrow 0$. Hence strong continuity holds for functions f in the range of g_α.

3) To complete the proof we show by contradiction that $g_\alpha\left(\hat{C}(S)\right)$ is dense in $\hat{C}(S)$ for any fixed $\alpha > 0$. The claim then follows once more by an $\varepsilon/3$-argument. Hence suppose that the closure of $g_\alpha\left(\hat{C}(S)\right)$ does not agree with $\hat{C}(S)$. Then there exists a non-trivial finite signed measure μ on $(S, \mathcal{B})$ such that

$$\mu(g_\alpha h) = 0 \qquad \text{for any } h \in \hat{C}(S),$$

cf. e.g. [**XXX**]. By the resolvent equation, $g_\alpha\left(\hat{C}(S)\right) = g_\beta\left(\hat{C}(S)\right)$ for any $\beta \in (0,\infty)$. Hence we even have

$$\mu\left(g_\beta h\right) = 0 \quad \text{for any } \beta > 0 \text{ and } h \in \hat{C}(S).$$

Moreover, (4.23) implies that $\beta g_\beta h \longrightarrow h$ pointwise as $\beta \longrightarrow \infty$. Therefore, by dominated convergence,

$$\mu(h) \;=\; \mu\left(\lim_{\beta\to\infty} \beta g_\beta h\right) \;=\; \lim_{\beta\to\infty} \beta\mu\left(g_\beta h\right) \;=\; 0 \quad \text{for any } h \in \hat{C}(S).$$

This contradicts the fact that μ is a non-trivial measure. ■

Existence of Feller processes

In the framework of Feller semigroups, the one-to-one correspondence between generators and semigroups can be extended to a correspondence between generators, semigroups and canonical Markov processes. Let $\Omega = \mathcal{D}(\mathbb{R}_+, S\cup\{\Delta\})$, $X_t(\omega) = \omega(t)$, and $\mathfrak{A} = \sigma(X_t : t \geq 0)$.

Theorem 4.23 (Existence and uniqueness of canonical Feller processes). Suppose that $(P_t)_{t\geq 0}$ is a Feller semigroup on $\hat{C}(S)$ with generator L. Then there exist unique probability measures P_x $(x \in S)$ on $(\Omega, \mathfrak{A})$ such that the canonical process $((X_t)_{t\geq 0}, P_x)$ is a Markov process satisfying $P_x[X_0 = x] = 1$ and

$$E_x[f(X_t)|\mathcal{F}_s^X] = (P_{t-s}f)(X_s) \quad P_x\text{-almost surely} \tag{4.24}$$

for any $x \in S, 0 \leq s \leq t$ and $f \in \hat{C}(S)$, where we set $f(\Delta) := 0$. Moreover, $((X_t)_{t\geq 0}, P_x)$ is a solution of the martingale problem for $(L, \mathrm{Dom}(L))$ for any $x \in S$.

Below, we outline two different proofs for the theorem. Additionally, it can also be shown that $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ is a strong Markov process:

Exercise (Strong Markov property for Feller processes). Let $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ be a canonical right-continuous time-homogeneous Markov process on S, and suppose that the transition semigroup satisfies $p_t\left(\hat{C}(S)\right) \subseteq \hat{C}(S)$ for any $t \geq 0$. Show that the strong Markov property holds. *Hint: Try to mimic the proof of the strong Markov property for Brownian motion.*

Proof (Sketch of proof of Theorem 4.23.). We only mention the main steps in the proof, details can be found for instance in Rogers and Williams [49].

1) One can show that the sub-Markov property implies that for any $t \geq 0$ there exists a sub-probability kernel $p_t(x, dy)$ on $(S, \mathcal{B})$ such that

$$(P_t f)(x) = \int p_t(x, dy) f(y) \quad \text{for any } f \in \hat{C}(S) \text{ and } x \in S.$$

By the semigroup property of $(P_t)_{t\geq 0}$, the kernels $(p_t)_{t\geq 0}$ form a transition function on $(S, \mathcal{B})$.

2) Now the **Kolmogorov extension theorem** shows that for any $x \in S$ there is a unique probability measure P_x^0 on the product space $S_\Delta^{[0,\infty)}$ with marginals

$$P_x \circ \left(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\right)^{-1} = p_{t_1}(x, dy_1)p_{t_2-t_1}(y_1, dy_2)\ldots p_{t_n-t_{n-1}}(y_{n-1}, dy_n)$$

for any $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \cdots < t_n$. Note that consistency of the given marginal laws follows from the semigroup property.

3) **Path regularisation:** To obtain a modification of the process with càdlàg sample paths, martingale theory can be applied. Suppose that $f = G_1 g$ for some non-negative function $g \in \hat{C}(S)$. Then

$$f - Lf \;=\; g \;\geq\; 0,$$

and hence the process $e^{-t} f(X_t)$ is a supermartingale w.r.t. P_x^0 for any x. The supermartingale convergence theorems now imply that P_x^0-almost surely, the limits

$$\lim_{\substack{s \downarrow t \\ s \in \mathbb{Q}}} e^{-s} f(X_s)$$

exist and define a càdlàg function in t. Applying this simultaneously for all functions g in a countable dense subset of the non-negative functions in $\hat{C}(S)$, one can prove that the process

$$\widetilde{X}_t = \lim_{\substack{s \downarrow t \\ s \in \mathbb{Q}}} X_s \quad (t \in \mathbb{R}_+)$$

exists P_x^0-almost surely and defines a càdlàg modification of $((X_t), P_x^0)$ for any $x \in S$. We can then choose P_x as the law of $(\widetilde{X}_t)$ under P_x^0.

4) **Uniqueness:** Finally, the measures P_x $(x \in S)$ are uniquely determined since the finite-dimensional marginals are determined by (4.24) and the initial condition. ■

We remark that alternatively, it is possible to construct a Feller process as a limit of jump processes, cf. Ethier and Kurtz [19, Chapter 4, Theorem 5.4]. Indeed, the Yosida approximation

$$\begin{aligned}
Lf &= \lim_{\alpha \to \infty} \alpha G_\alpha L f = \lim_{\alpha \to \infty} L^{(\alpha)} f, \quad L^{(\alpha)} f := \alpha(\alpha G_\alpha f - f), \\
P_t f &= \lim_{\alpha \to \infty} e^{t L^{(\alpha)}} f,
\end{aligned}$$

is an approximation of the generator by bounded linear operators $L^{(\alpha)}$ that can be represented in the form

$$L^{(\alpha)} f \;=\; \alpha \int (f(y) - f(x))\, \alpha g_\alpha(x, dy)$$

with sub-Markov kernels αg_α. For any $\alpha \in (0, \infty)$, $L^{(\alpha)}$ is the generator of a canonical jump process $((X_t)_{t \geq 0}, (P_x^{(\alpha)})_{x \in S})$ with bounded jump intensities. By using that for any $f \in \mathrm{Dom}(L)$,

$$L^{(\alpha)} f \to Lf \quad \text{uniformly as } \alpha \to \infty,$$

one can prove by similar techniques as in Section 6.2 that the family $\{P_x^{(\alpha)} : \alpha \in \mathbb{N}\}$ of probability measures on $\mathcal{D}(\mathbb{R}_+, S \cup \{\Delta\})$ is tight, i.e., there exists a weakly convergent subsequence. Denoting the limit by P_x, the canonical process $((X_t), P_x)$ is a Markov process that solves the martingale problem for the generator $(L, \mathrm{Dom}(L))$.

Generators of Feller semigroups

It is possible to classify all generators of Feller processes in $\mathbb{R}^d$ that contain $C_0^\infty(\mathbb{R}^d)$ in the domain of their generator. The key observation is that the sub-Markov property of the semigroup implies a maximum principle for the generator. Indeed, the following variant of the Hille-Yosida theorem holds:

Theorem 4.24 (Characterization of Feller generators). A linear operator $(L, \mathrm{Dom}(L))$ on $\hat{C}(S)$ is the generator of a Feller semigroup $(P_t)_{t\geq 0}$ if and only if the following conditions hold:

(i) $\mathrm{Dom}(L)$ is a dense subspace of $\hat{C}(S)$.

(ii) $\mathrm{Range}(\alpha I - L) = \hat{C}(S)$ for some $\alpha > 0$.

(iii) L satisfies the **positive maximum principle:** If f is a function in the domain of L and $f(x_0) = \max f$ for some $x_0 \in S$ then $(Lf)(x_0) \leq 0$.

Proof. "$\Rightarrow$" If L is the generator of a Feller semigroup then (i) and (ii) hold by the Hille-Yosida Theorem 4.22. Furthermore, suppose that $f \leq f(x_0)$ for some $f \in \mathrm{Dom}(L)$ and $x_0 \in S$. Then $0 \leq \frac{f^+}{f(x_0)} \leq 1$, and hence by the sub-Markov property, $0 \leq P_t \frac{f^+}{f(x_0)} \leq 1$ for any $t \geq 0$. Thus $P_t f \leq P_t f^+ \leq f(x_0)$, and

$$(Lf)(x_0) = \lim_{t\downarrow 0} \frac{(P_t f)(x_0) - f(x_0)}{t} \leq 0.$$

"$\Leftarrow$" Conversely, if (iii) holds then L is dissipative. Indeed, for any function $f \in \hat{C}(S)$ there exists $x_0 \in S$ such that $\|f\|_{\sup} = |f(x_0)|$. Assuming w.l.o.g. $f(x_0) \geq 0$, we conclude by (iii) that

$$\alpha\|f\|_{\sup} \leq \alpha f(x_0) - (Lf)(x_0) \leq \|\alpha f - Lf\|_{\sup} \quad \text{for any } \alpha > 0.$$

The Hille-Yosida Theorem 4.22 now shows that L generates a C^0 contraction semigroup $(P_t)_{t\geq 0}$ on $\hat{C}(S)$ provided (i),(ii) and (iii) are satisfied. It only remains to verify the sub-Markov property. This is done in two steps:

a) αG_α is sub-Markov for any $\alpha > 0$: $0 \leq f \leq 1 \Rightarrow 0 \leq \alpha G_\alpha f \leq 1$. This follows from the maximum principle by contradiction. Suppose for instance that $g := \alpha G_\alpha f \leq 1$, and let $x_0 \in S$ such that $g(x_0) = \max g > 1$. Then by (iii), $(Lg)(x_0) \leq 0$, and hence $f(x_0) = \frac{1}{\alpha}(\alpha g(x_0) - (Lg)(x_0)) > 1$.

b) P_t is sub-Markov for any $t \geq 0 : 0 \leq f \leq 1 \Rightarrow 0 \leq P_t f \leq 1$. This follows from a) by Yosida approximation: Let $L^{(\alpha)} := L\alpha G_\alpha = \alpha^2 G_\alpha - \alpha I$. If $0 \leq f \leq 1$ then the sub-Markov property for αG_α implies

$$e^{tL^{(\alpha)}} f = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} (\alpha G_\alpha)^n f \in [0,1] \quad \text{for any } t \geq 0.$$

Hence also $P_t f = \lim_{\alpha\to\infty} e^{tL^{(\alpha)}} f \in [0,1]$ for any $t \geq 0$. ■

Example (Sticky Brownian motions). Let $S = [0,\infty)$ and fix $\lambda \in [0,1]$. We verify that the linear operator $Lf = f''/2$ with domain

$$\mathrm{Dom}(L) = \left\{ f \in \hat{C}(S) : f', f'' \in \hat{C}(S), (1-\lambda)f'(0) = \lambda f''(0) \right\}$$

generates a Feller semigroup. Condition (i) in Theorem 4.24 is satisfied. To check (ii), we have to show that for $\alpha > 0$, the boundary value problem

$$\alpha f - \frac{1}{2} f'' = g, \quad (1-\lambda)f'(0) = \lambda f''(0), \quad \lim_{x\to\infty} f(x) = 0, \tag{4.25}$$

has a solution $f \in \mathrm{Dom}(L)$ for any $g \in \hat{C}(S)$. The general solution of the corresponding homogeneous o.d.e. (i.e., for $g = 0$) is $f(x) = c_1 e^{\sqrt{2\alpha}x} + c_2 e^{\sqrt{-2\alpha}x}$. For $\lambda \in [0,1]$, only the trivial solution satisfies the boundary conditions. Therefore, the inhomogeneous boundary value problem has a unique solution that can be computed explicitly by variation of constants. The positive maximum principle (iii) is satisfied as well. Indeed, suppose that f is a function in $\mathrm{Dom}(L)$ that has a maximum at $x_0 \in [0,\infty)$. If $x_0 > 0$ then $(Lf)(x_0) = f''(x_0)/2 \leq 0$. Furthermore, if $x_0 = 0$ then $f'(0) \leq 0$. For $\lambda \in (0,1]$, this implies $(Lf)(x_0) = f''(0)/2 \leq 0$ by the boundary condition in (4.25). Moreover, for $\lambda = 0$ we have $f'(0) = 0$ by (4.25), and hence again $(Lf)(x_0) = f''(0)/2 \leq 0$, because $x_0 = 0$ is a maximum.

Notice that the condition $\lambda \in [0,1]$ is essential to ensure that L satisfies the positive maximum principle and hence generates a Feller semigroup (P_t). For each such λ, Theorems 4.24 and 4.23 show that there is a Feller process with generator L.

For $\lambda = 0$ and $\lambda = 1$ these processes can be identified as a *Brownian motion with reflection at* 0 and a *Brownian motion trapped at* 0, respectively. Indeed, let

$$X_t^0 \;:=\; |B_t| \qquad \text{and} \qquad X_t^1 \;:=\; B_{t\wedge T_0},$$

where $(B,(P_x)_{x\in\mathbb{R}})$ is a Brownian motion on $\mathbb{R}$. Then both $(X^0,(P_x)_{x\in[0,\infty)})$ and $(X^1,(P_x)_{x\in[0,\infty)})$ are Feller processes, and, by a similar argument as in Section 4.2, their generators extend the operator $(L,\mathrm{Dom}(L))$ with boundary condition $f'(0) = 0$, $f''(0) = 0$, respectively.

The corresponding Feller process for $\lambda \in (0,1)$ is called a *Brownian motion with a sticky boundary at* 0. To obtain some information on the surprising properties of this process, we consider a sequence (g_n) of functions in $\hat{C}(S)$ such that $g_n \geq 0$ for all n and $g_n \searrow 1_{\{0\}}$ as $n \to \infty$. An explicit computation of the resolvents $G_\alpha g_n$ (by solving the boundary value problem (4.25)) shows that the canonical Feller process $((X_t),(P_x))$ generated by L satisfies

$$E_0\left[\int_0^\infty \alpha e^{-\alpha t} 1_{\{X_t=0\}}\,dt\right] \;=\; \alpha \lim_{n\to\infty} (G_\alpha g_n)(0) \;=\; \frac{\sqrt{2\alpha}\,\lambda}{\sqrt{2\alpha}\,\lambda + 1 - \lambda}$$

for any $\alpha > 0$. Hence the process starting at 0 spends in total a positive amount of time at 0, i.e.,

$$\mathrm{Leb}\left(\{t \geq 0 : X_t = 0\}\right) \;>\; 0 \qquad P_0\text{-almost surely.}$$

Nevertheless, by the strong Markov property, almost surely, the process leaves the state 0 immediately, and there is no non-empty time interval $(a,b) \subset \mathbb{R}_+$ such that $X_t = 0$ for all $t \in (a,b)$. In other words: with probability one, $\{t \geq 0 : X_t = 0\}$ is a set with positive Lebesgue measure that does contain any non-empty open interval !

XXX Include picture

Exercise (Sticky boundaries). Fill in the missing details in the arguments in the example above. You may consult Liggett [34, Example 3.59 and Exercise 3.61] for further hints.

For diffusion processes on $\mathbb{R}^d$, the maximum principle combined with a Taylor expansion shows that the generator L is a second order differential operator provided $C_0^\infty(\mathbb{R}^d)$ is contained in the domain of L. Recall that a Markov process $((X_t),(P_x))$ is called *conservative* iff for any $x \in S$, the life-time is P_x almost surely infinite.

Theorem 4.25 (Dynkin). Suppose that $(P_t)_{t\geq 0}$ is a Feller semigroup on $\mathbb{R}^d$ such that $C_0^\infty(\mathbb{R}^d)$ is a subspace of the domain of the generator L. If $(P_t)_{t\geq 0}$ is the transition semigroup of a conservative Markov process $((X_t)_{t\geq 0},(P_x)_{x\in\mathbb{R}^d})$ with continuous paths then there exist functions $a_{ij}, b_i \in C(\mathbb{R}^d)$ $(i,j = 1,\ldots,d)$ such that for any x, $a_{ij}(x)$ is non-negative definite, and

$$(Lf)(x) \;=\; \sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x)\frac{\partial f}{\partial x_i}(x) \quad \forall\, f \in C_0^\infty(\mathbb{R}^d). \tag{4.26}$$

Proof. 1) **L is a local operator:** We show that

$$f, g \in \mathrm{Dom}(L), f = g \text{ in a neighbourhood of } x \Rightarrow (Lf)(x) = (Lg)(x).$$

For the proof we apply optional stopping to the martingale $M_t^f = f(X_t) - \int_0^t (Lf)(X_s)ds$. For an arbitrary bounded stopping time T and $x \in \mathbb{R}^d$, we obtain **Dynkin's formula**

$$E_x[f(X_T)] = f(x) + E_x\left[\int_0^T (Lf)(X_s)ds\right].$$

By applying the formula to the stopping times

$$T_\varepsilon = \min\{t \geq 0 : X_t \notin B(x, \varepsilon)\} \wedge 1, \quad \varepsilon > 0,$$

we can conclude that

$$(Lf)(x) = \lim_{\varepsilon \downarrow 0} \frac{E_x\left[\int_0^{T_\varepsilon} (Lf)(X_s)ds\right]}{E_x[T_\varepsilon]} = \lim_{\varepsilon \downarrow 0} \frac{E_x[f(X_{T_\varepsilon})] - f(x)}{E_x[T_\varepsilon]}. \tag{4.27}$$

Here we have used that Lf is bounded and $\lim\limits_{s \downarrow 0} (Lf)(X_s) = (Lf)(x)$ P_x-almost surely by right-continuity. The expression on the right-hand side of (4.27) is known as "**Dynkin's characteristic operator**". Assuming continuity of the paths, we obtain $X_{T_\varepsilon} \in \overline{B(x, \varepsilon)}$. Hence if $f, g \in \mathrm{Dom}(L)$ coincide in a neighbourhood of x then $f(X_{T_\varepsilon}) \equiv g(X_{T_\varepsilon})$ for $\varepsilon > 0$ sufficiently small, and thus $(Lf)(x) = (Lg)(x)$ by (4.27).

2) **Local maximum principle:** Locality of L implies the following extension of the positive maximum principle: If f is a function in $C_0^\infty(\mathbb{R}^d)$ that has a **local** maximum at x and $f(x) \geq 0$, then $(Lf)(x) \leq 0$. Indeed, if $f(x) > 0$ then we can find a function $\widetilde{f} \in C_0^\infty(\mathbb{R}^d)$ that has a global maximum at x such that $\widetilde{f} = f$ in a neighbourhood of x. Since L is a local operator by Step 1, we can conclude that

$$(Lf)(x) \;=\; (L\widetilde{f})(x) \;\leq\; 0.$$

If $f(x) = 0$ then we can apply the same argument to $f + \varepsilon g$ where ε is a positive constant, and g is a function in $C_0^\infty(\mathbb{R}^d)$ such that $g \equiv 1$ in a neighbourhood of x. In this case,

$$(Lf)(x) \;=\; (L(f + \varepsilon g))(x) + \varepsilon(Lg)(x) \;\leq\; \varepsilon(Lg)(x) \qquad \text{for any } \varepsilon > 0,$$

and hence we can conclude again that $(Lf)(x) \leq 0$.

3) **Taylor expansion:** For proving that L is a differential operator of the form (4.26) we fix $x \in \mathbb{R}^d$ and functions $\varphi, \psi_1, \ldots, \psi_d \in C_0^\infty(\mathbb{R}^d)$ such that $\varphi(y) = 1$, $\psi_i(y) = y_i - x_i$ in a neighbourhood U of x. Let $f \in C_0^\infty(\mathbb{R}^d)$. Then by Taylor's formula there exists a function $R \in C_0^\infty(\mathbb{R}^d)$ such that $R(y) = \mathrm{o}(|y - x|^2)$ and

$$f(y) = f(x)\phi(y) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)\psi_i(y) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\psi_i(y)\psi_j(y) + R(y) \tag{4.28}$$

in a neighbourhood of x. Since L is a local linear operator, we obtain

$$(Lf)(x) = c(x)f(x) + \sum_{i=1}^d b_i(x)\frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + (LR)(x) \tag{4.29}$$

with $c(x) := (L\phi)(x), b_i(x) := (L\psi_i)(x)$, and $a_{ij}(x) := L(\psi_i\psi_j)(x)$. Since ϕ has a local maximum at x, $c(x) \leq 0$. Similarly, for any $\xi \in \mathbb{R}^d$, the function

$$\sum_{i,j=1}^{d} \xi_i\xi_j\psi_i(y)\psi_j(y) = \left|\sum_{i=1}^{d} \xi_i\psi_i(y)\right|^2$$

equals $|\xi \cdot (y-x)|^2$ in a neighbourhood of x, so it has a local minimum at x. Hence

$$\sum_{i,j=1}^{d} \xi_i\xi_j a_{ij}(x) = L\left(\sum_{i,j} \xi_i\xi_j\psi_i\psi_j\right) \geq 0,$$

i.e., the matrix $(a_{ij}(x))$ is non-negative definite. By (4.29), it only remains to show $(LR)(x) = 0$. To this end consider

$$R_\varepsilon(y) := R(y) - \varepsilon\sum_{i=1}^{d} \psi_i(y)^2.$$

Since $R(y) = \mathrm{o}(|y-x|^2)$, the function R_ε has a local maximum at x for $\varepsilon > 0$. Hence

$$0 \geq (LR_\varepsilon)(x) = (LR)(x) - \varepsilon\sum_{i=1}^{d} a_{ii}(x) \quad \forall\varepsilon > 0.$$

Letting ε tend to 0, we obtain $(LR)(x) \leq 0$. On the other hand, R_ε has a local minimum at x for $\varepsilon < 0$, and in this case the local maximum principle implies

$$0 \leq (LR_\varepsilon)(x) = (LR)(x) - \varepsilon\sum_{i=1}^{d} a_{ii}(x) \quad \forall\varepsilon < 0,$$

and hence $(LR)(x) \geq 0$. Thus $(LR)(x) = 0$.

4) **Vanishing of c:** If the process is conservative then $p_t 1 = 1$ for any $t \geq 0$. Informally this should imply $c = L1 = \frac{d}{dt}p_t 1|_{t=0_+} = 0$. However, the constant function 1 is not contained in the Banach space $\hat{C}(\mathbb{R}^d)$. To make the argument rigorous, one can approximate 1 by C_0^∞ functions that are equal to 1 on balls of increasing radius. The details are left as an exercise. ■

Theorem 4.25 has an extension to generators of general Feller semigroups including those corresponding to processes with discontinuous paths. We state the result without proof:

Theorem 4.26 (Courrège).
Suppose that L is the generator of a Feller semigroup on $\mathbb{R}^d$, and $C_0^\infty(\mathbb{R}^d) \subseteq \mathrm{Dom}(L)$. Then there exist functions $a_{ij}, b_i, c \in C(\mathbb{R}^d)$ and a kernel ν of positive Radon measures such that

$$\begin{aligned}(Lf)(x) = &\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 f}{\partial x_i\partial x_j}(x) + \sum_{i=1}^{d} b_i(x)\frac{\partial f}{\partial x_i}(x) + c(x)f(x)\\ &+ \int_{\mathbb{R}^d\setminus\{x\}} \left(f(y) - f(x) - 1_{\{|y-x|<1\}}(y-x)\cdot\nabla f(x)\right)\nu(x,dy)\end{aligned}$$

holds for any $x \in \mathbb{R}^d$ and $f \in C_0^\infty(\mathbb{R}^d)$. The associated Markov process has continuous paths if and only if $\nu \equiv 0$.

For transition semigroups of Lévy processes (i.e., processes with independent and stationary increments), the coefficients a_{ij}, b_i, c, and the measure ν do not depend on x. In this case, the theorem is a consequence of the Lévy-Khinchin representation, cf. e.g. [17].

5. Jump processes and interacting particle systems

Section 5.1 is devoted to an explicit construction of piecewise deterministic Markov processes (PDMPs), a class of Markov processes that combines jumps and deterministic motion. We construct the processes from their characteristics, and identify the generators. A special case are jump processes with (possibly time dependent) jump rates of finite intensity. In Sections 5.2 and 5.3, we construct interacting particle systems on finite and infinite graphs, and Section 5.4 discusses ergodicity and phase transitions for attractive particle systems.

5.1. Piecewise deterministic Markov processes

An additional reference for this section is the book [10] by Davis.

We assume that S is a locally compact and separable state space with Borel σ-algebra $\mathcal{B}$. We will construct a Markov process on S that combines jumps and deterministic motion between the jumps. The jumps are characterized by a jump rate given by a measurable function $\lambda : S \to [0, \infty)$, and a probability kernel π on $(S, \mathcal{B})$ that determines the law of the jump transitions. The deterministic motion is described by a continuous deterministic flow, i.e., a continuous function $\xi : [0, \infty) \times S \to S$ such that for any $s, t \geq 0$ and $x \in S$;

$$\xi_t(\xi_s(x)) = \xi_{t+s}(x). \tag{5.1}$$

For example, $S = \mathbb{R}^d$, and ξ is the flow of a vector field $b : \mathbb{R}^d \to \mathbb{R}^d$, i.e., the solution to the ordinary differential equation

$$\frac{d}{dt}\xi_t(x) = b(\xi_t(x)), \qquad \xi_0(x) = x.$$

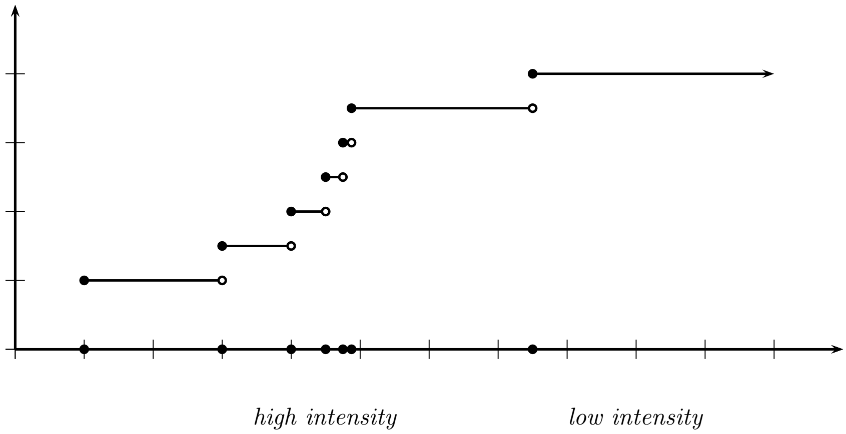
Examples of PDMPs

Before constructing piecewise deterministic Markov processes (PDMPs), we consider some basic examples that show that such processes occur in numerous applications.

Example. 1) **Time inhomogeneous jump processes.** Suppose we want to construct a pure jump process $(X_t)_{t\geq 0}$ with time dependent jump rate $q_t(x, dy)$. To avoid time inhomogeneity, it is natural to consider the time-space process $(t, X_t)_{t\geq 0}$. However, this process is no longer a pure jump process. Indeed, it is moving deterministically between the jumps, and therefore, it is a PDMP. For $(s, x) \in [0, \infty) \times S$, the characteristics of the process are given by

$$\begin{aligned} \xi_t(s,x) &= (s+t, x), \\ \lambda(s,x) &= q_s(x, S), \\ \pi\left((s,x), dt\, dy\right) &= \delta_s(dt)\, \lambda(s,x)^{-1} q_s(x, dy). \end{aligned}$$

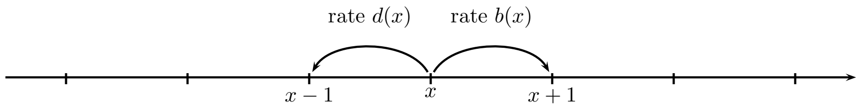
For example, an *inhomogeneous Poisson process* with intensities $\lambda(t)$ is a time dependent jump process with state space $\mathbb{Z}_+$ and jump rates $q_t(x, \cdot) = \lambda(t)\delta_{x+1}$.



More generally, a *time-dependent birth-death process* is a Markov jump process on $\mathbb{Z}_+$ with transition rates given by

$$q_t(x, \cdot) \;=\; b(t,x)\delta_{x+1} + d(t,x)\delta_{x-1}$$

for functions $b, d : [0,\infty) \times \mathbb{Z}_+ \longrightarrow [0,\infty)$.



For example, in a *time-dependent branching process*, the particles in a population are independently giving birth to a child with rate $b(t)$ and dying with rate $d(t)$. The total population size at time t can then be described as a time-dependent birth-death process with transition rates $b(t,x) = b(t)x$ and $d(t,x) = d(t)x$.

2) **Compensated Poisson processes.** A very simple example of a PDMP is a compensated Poisson process $X_t = N_t - \lambda t$ where (N_t) is a Poisson process with intensity λ. In this case, the state space is $\mathbb{R}$, and the characteristics are given by $\xi_t(x) = x - \lambda t$, $\lambda(x) = \lambda$ and $\pi(x, \cdot) = \delta_{x+1}$ for all x.

3) **Random kinetic models, Andersen dynamics.** Consider a particle that is moving deterministically but is accelerated by random kicks that occur after independent exponential waiting times. We can model such a process as a PDMP on the phase space

$$S \;=\; \mathbb{R}^d \times \mathbb{R}^d \;=\; \left\{(x,v) : x, v \in \mathbb{R}^d\right\},$$

where the first and second component stand for the position and the velocity of the particle. For example, a model for *physical Brownian motion* is given by a PDMP with characteristics

$$\xi_t(x,v) = (x + tv, v), \quad \lambda(x) = \lambda, \quad \pi\left((x,v), d(x',v')\right) = \delta_x(dx')\,\mathcal{N}\left(\sqrt{1-\gamma^2}v, \gamma^2 I_d\right)(dv'),$$

where $\lambda \in [0,\infty)$ and $\gamma \in [0,1]$ are fixed constants. For $\gamma = 1$, the velocity is completely refreshed at each kick, whereas $\gamma < 1$ corresponds to a partial refreshment combined with a damping.

A more general model that is frequently considered in molecular dynamics is *Andersen dynamics*. Here one considers m particles of unit mass taking values in $\mathbb{R}^n$. The positions and velocities of the m-particle configuration form then an element $(x,v) = (x^i, v^i)_{i=1,\dots,m}$ in the state space

$S = \mathbb{R}^d \times \mathbb{R}^d$ where $d = mn$. We now assume that the particles are subject to independent random kicks, and in between kicks, they are moving independently according to Hamiltonian dynamics with respect to a potential energy function $U(x_1, \ldots, x_m)$ defined on the configuration space. In that case, the dynamics is described by a PDMP, where $\xi_t(x, v) = (x_t, v_t)$ is the solution of Hamilton's equation of motion

$$\frac{dx}{dt} = v, \quad \frac{dv}{dt} = -\nabla U(x), \quad x_0 = x, \quad v_0 = v,$$

the total intensity $\lambda(x, v) = \lambda$ of random kicks is constant, and

$$\pi\left((x,v), d(x',v')\right) \;=\; \delta_x(dx')\, \frac{1}{m}\sum_{i=1}^{m} \mathcal{N}\left(\sqrt{1-\gamma^2}v_i, \gamma^2 I_n\right)(dv_i') \prod_{j\neq i} \delta_{v_j}(dv_j'). \tag{5.2}$$

This means that at each random kick, only the velocity of one (randomly chosen) particle is changed.

Construction

We will now give an explicit construction of a piecewise deterministic Markov process with given characteristics. This construction also forms the basis for algorithms for the simulation of PDMPs.

We will denote the increasing sequence of jump times of the process by $(\mathcal{J}_n)_{n\in\mathbb{N}}$ and the corresponding sequence of jump targets by $(Y_n)_{n\in\mathbb{N}}$. We allow explosions in finite time; the explosion time is $\mathcal{J}_\infty = \sup \mathcal{J}_n$. The PDMP with initial condition Y_0 at initial time $\mathcal{J}_0$ is then given by

$$X_t \;=\; \begin{cases} \xi_{t-J_n}(Y_n) & \text{for } t \in [\mathcal{J}_n, \mathcal{J}_{n+1}), \quad n \in \mathbb{Z}_+, \\ \Delta & \text{for } t \in [J_\infty, \infty). \end{cases}$$

We fix a deterministic initial time $t_0 \in \mathbb{R}_+$ and an initial position $x_0 \in S$. In our construction, the random variables $\mathcal{J}_n$ and Y_n and, consequently, the process (X_t) are defined by the following algorithm:

Algorithm 1: Construction of PDMP

Input: $t_0 \in \mathbb{R}_+$, $x_0 \in S$, $n \in \mathbb{N}$.
Output: Jump times $(\mathcal{J}_i)_{i=0,\ldots,n}$ and jump targets $(Y_i)_{i=0,\ldots,n}$ of PDMP up to nth jump.

$\mathcal{J}_0 \leftarrow t_0$, $Y_0 \leftarrow x_0$;
for $i \leftarrow 1$ **to** n **do**
  Sample $E_i \sim \mathrm{Exp}(1)$, independently of $Y_0, \ldots, Y_{i-1}, E_1, \ldots, E_{i-1}$;
  $\mathcal{J}_i \leftarrow \inf\left\{t \geq 0 : \int_{\mathcal{J}_{i-1}}^{t} \lambda(\xi_{s-\mathcal{J}_{i-1}}(Y_{i-1}))\, ds \geq E_i\right\}$;
  Sample $Y_i | (Y_0, \ldots, Y_{i-1}, E_0, \ldots, E_i) \sim \pi\left(\xi_{\mathcal{J}_i - \mathcal{J}_{i-1}}(Y_{i-1}), \cdot\right)$;
return $(\mathcal{J}_i)_{i=0}^n$ and $(Y_i)_{i=0}^n$;

Here we set $\inf \varnothing := \infty$. It may happen (for instance if $\lambda \equiv 0$) that $\mathcal{J}_n = \infty$ with positive probability, and in that case we set $\mathcal{J}_{n+1} - \mathcal{J}_n := 0$. Note that in this case, the process gets stuck at Y_n.

Let $P_{(t_0,x_0)}$ denote the underlying probability measure for the random variables defined by the algorithm. Under $P_{(t_0,x_0)}$, $\mathcal{J}_0 = t_0$ and $Y_0 = x_0$ almost surely, and the conditional laws of $\mathcal{J}_{n+1}$ and Y_{n+1} are given by

$$P_{(t_0,x_0)}\left[\mathcal{J}_{n+1} > t \,|\, \mathcal{J}_0, Y_0, \ldots, \mathcal{J}_n, Y_n\right] \;=\; \exp\left(-\int_{J_n}^{t} \lambda(\xi_{s-\mathcal{J}_n}(Y_n))\, ds\right) 1_{t > \mathcal{J}_n}, \tag{5.3}$$

$$P_{(t_0,x_0)}\left[Y_{n+1} \in B \,|\, \mathcal{J}_0, Y_0, \ldots, \mathcal{J}_n, Y_n, \mathcal{J}_{n+1}\right] \;=\; \pi\left(\xi_{\mathcal{J}_{n+1}-\mathcal{J}_n}(Y_n), B\right). \tag{5.4}$$

Although in general, $(\mathcal{J}_n)_{n\in\mathbb{Z}_+}$ and $(Y_n)_{n\in\mathbb{Z}_+}$ are not Markov chains on their own, the process $(\mathcal{J}_n, Y_n)_{n\in\mathbb{Z}_+}$ is a time-homogeneous Markov chain w.r.t. the filtration

$$\mathcal{G}_n \;=\; \sigma(Y_0, \ldots, Y_n, E_1, \ldots, E_n),$$

and with transition kernel

$$\widetilde{\pi}((s,x),d(t,y)) \;=\; \lambda(\xi_{t-s}(x))\,\exp\left(-\int_s^t \lambda(\xi_{r-s}(x))dr\right) 1_{(s,\infty)}(t)\,\pi(\xi_{t-s}(x),dy)\,dt.$$

In particular, conditionally given $\mathcal{G}_n$, $\mathcal{J}_{n+1}$ has a **survival distribution with hazard rate** $\lambda(X_t)$, i.e., as $h \downarrow 0$,

$$\frac{P_{(t_0,x_0)}\left[\mathcal{J}_{n+1} > t+h\,|\mathcal{G}_n\right]}{P_{(t_0,x_0)}\left[\mathcal{J}_{n+1} > t\,|\mathcal{G}_n\right]} \;=\; e^{-\int_t^{t+h}\lambda(\xi_{r-\mathcal{J}_n}(Y_n))\,dr} \;=\; 1 \,-\, h\,\lambda(X_t) \,+\, o(h)$$

for all $t \geq J_n$. Here we have used that

$$\xi_{r-\mathcal{J}_n}(Y_n) = X_r \qquad \text{for all } r < \mathcal{J}_{n+1}.$$

Notice that an exponential distribution is a special survival distribution with time-independent hazard rate λ. Furthermore, similarly to the exponential distributions, survival distributions have a generalized *memoryless property* that will be crucial in verifying that the process (X_t) is indeed a Markov process.

Example (Pure jump process). Suppose that $\xi_t(x) = x$ for all $t \geq 0$ and $x \in S$, i.e., the process does not move between jumps. In this case, the sequence $(Y_n)_{n\in\mathbb{Z}_+}$ of positions of the process is a Markov chain with transition kernel π. Given $\sigma(Y_n : n \in \mathbb{Z}_+)$, the waiting times for the next jumps are independent with law

$$\mathcal{J}_n - \mathcal{J}_{n-1} \;=\; \frac{E_n}{\lambda(Y_{n-1})} \;\sim\; \mathrm{Exp}(\lambda(Y_{n-1})).$$

Things simplify further if we consider a pure jump process with bounded total jump intensities $q(x,dy)$, i.e., $\sup_{x\in S} q(x, S\setminus\{x\}) \leq \lambda$ for a finite constant λ. In this case we can write the jump intensities in the form

$$q(x,B) = \lambda\pi(x,B) \qquad \text{for all measurable } B \subseteq S\setminus\{x\},$$

where π is a transition kernel satisfying $\pi(x,\{x\}) = 1 - q(x,S\setminus\{x\})/\lambda$. The process constructed correspondingly according to the algorithm above has i.i.d. waiting times

$$\mathcal{J}_n - \mathcal{J}_{n-1} \sim \mathrm{Exp}(\lambda)$$

between jumps. Hence the number of jumps up to time t is a Poisson process with parameter λ. The process (Y_n) of positions is an independent Markov chain with transition kernel π, and the continuous-time process is given by

$$X_t = Y_{N_t}. \tag{5.5}$$

By (5.5), it is easy to compute the transition functions of the process in continuous time. Indeed, by independence of (Y_n) and (N_t),

$$\begin{aligned}(P_tf)(x) \;&=\; E_{0,x}\left[f(X_t)\right] \;=\; \sum_{k=0}^{\infty} E_{0,x}\left[f(Y_k); N_t = k\right] \;=\; \sum_{k=0}^{\infty}(\pi^k f)(x)\frac{(\lambda t)^k}{k!}e^{-\lambda t}\\ &=\; \left(e^{\lambda t\pi}f\right)(x)e^{-\lambda t} \;=\; \left(e^{t\mathcal{L}}f\right)(x)\end{aligned}$$

for any $t \geq 0, x \in S$ and $f \in \mathcal{F}_b(S)$ where $e^{t\mathcal{L}} = \sum_{k=0}^{\infty}\frac{1}{k!}(t\mathcal{L})^k$ denotes the exponential of the bounded linear operator

$$(\mathcal{L}f)(x) \;=\; (\lambda(\pi - I)f)(x) \;=\; \int q(x,dy)(f(y) - f(x)).$$

Standard properties of the operator exponential now show that p_tf satisfies the Kolmogorov forward and backward equation

$$\frac{d}{dt}p_tf \;=\; p_t\mathcal{L}f \;=\; \mathcal{L}p_tf$$

where the derivative can be taken w.r.t. the supremum norm. For unbounded jump intensities, the derivation of Kolmogorov's equations will be technically much more demanding.

Markov property

We now want to show that the process constructed above is indeed a Markov process. We recall that under $P_{(t_0,x_0)}$, the process $(\mathcal{J}_n, Y_n)_{n\in\mathbb{Z}_+}$ is a time-homogeneous Markov chain with transition kernel $\widetilde{\pi}$ w.r.t. the filtration $\mathcal{G}_n = \sigma(Y_0,\dots,Y_n,E_1,\dots,E_n)$, and the continuous-time process $(X_t)_{t\geq t_0}$ is obtained as a deterministic function

$$X_t = \Phi_t(\mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1, \dots)$$

of the Markov chain, where

$$\Phi_t(t_0,x_0,t_1,x_1,\dots) := \begin{cases} X_n & \text{if } t\in[t_n,t_{n+1}) \text{ for some } n\in\mathbb{Z}_+, \\ \Delta & \text{if } t\geq \sup t_n. \end{cases} \tag{5.6}$$

Let $\mathcal{F}_t^X = \sigma(X_s : s\in[t_0,t])$ denote the filtration generated by the continuous time process.

Theorem 5.1 (Markov property). The process $((X_t)_{t\geq t_0}, (P_{t_0,x_0}))$ is an $(\mathcal{F}_t^X)$ Markov process, i.e.,

$$E_{(t_0,x_0)}\left[F(X_{s:\infty})1_{\{s<\zeta\}}\middle|\mathcal{F}_s^X\right] = E_{(s,X_s)}\left[F(X_{s:\infty})\right] \quad P_{(t_0,x_0)}\text{-a.s. on } \{s<\zeta\}$$

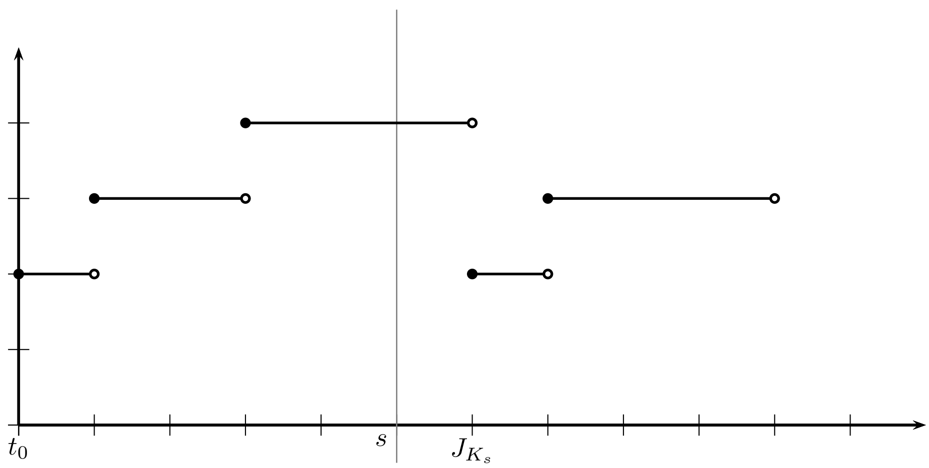
for any $0\leq t_0\leq s$, $x_0\in S$, and any bounded measurable function $F:\mathcal{D}(\mathbb{R}_+, S\cup\{\Delta\})\to\mathbb{R}$.
Moreover, the Markov process is time-homogeneous.

For proving the theorem we will use the Markov property in discrete time. However, the problem is that the relevant filtration is $(\mathcal{G}_n)_{n\in\mathbb{N}}$, whereas we are interested in the conditional expectation given $\mathcal{F}_s^X$. To overcome this difficulty let

$$K_s = \min\{n\in\mathbb{Z}_+ : \mathcal{J}_n > s\}$$

denote the index of the first jump after time s. Note that K_s is a stopping time w.r.t. $(\mathcal{G}_n)$, and

$$\{K_s<\infty\} = \{s<\zeta\}.$$



Besides the Markov property of the chain $(\mathcal{J}_n, Y_n)$, the following fact will be crucial for the proof of Theorem 5.1.

Lemma 5.2 (Memoryless property). *The conditional law of $(\mathcal{J}_{K_s}, Y_{K_s})$ given $\mathcal{F}_s^X$ coincides with the law of $(\mathcal{J}_1, Y_1)$ under $P_{(s,X_s)}$. More precisely, for any $s \geq t_0$, $t \geq s$ and $B \in \mathcal{B}$,*

$$P_{(t_0,x_0)}\left[\mathcal{J}_{K_s} > t \text{ and } Y_s \in B \mid \mathcal{F}_s^X\right] \;=\; P_{(s,X_s)}\left[\mathcal{J}_1 > t \text{ and } Y_1 \in B\right] \qquad \textit{almost surely on } \{s < \zeta\}.$$

Proof. Let $A \in \mathcal{F}_s^X$. Then it can be verified that for any $n \in \mathbb{N}$, there exists an event $A_n \in \mathcal{G}_{n-1}$ such that $A \cap \{s < \mathcal{J}_n\} = A_n \cap \{s < \mathcal{J}_n\}$. This is because for $s < \mathcal{J}_n$, X_s is a deterministic function of $\mathcal{J}_0, Y_0, \ldots, \mathcal{J}_{n-1}, Y_{n-1}$. Therefore, by the Markov property for the discrete time chain, we obtain

$$\begin{aligned}
&P_{(t_0,x_0)}\left[\mathcal{J}_n > t, Y_n \in B \mid \mathcal{G}_{n-1}\right] \;=\; \widetilde{\pi}\left((\mathcal{J}_{n-1}, Y_{n-1}), (t,\infty)\times B\right)\\
&\quad = \int_t^\infty \lambda(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}))\,\exp\left(-\int_{\mathcal{J}_{n-1}}^v \lambda(\xi_{r-\mathcal{J}_{n-1}}(Y_{n-1}))\,dr\right)\pi\left(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}), B\right)\,dv\\
&\quad = P_{(t_0,x_0)}\left[\mathcal{J}_n > s | \mathcal{G}_{n-1}\right]\int_t^\infty \lambda(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}))\,\exp\left(-\int_s^v \lambda(\xi_{r-\mathcal{J}_{n-1}}(Y_{n-1}))\,dr\right)\pi\left(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}), B\right)\,dv\\
&\quad = E_{(t_0,x_0)}\left[1_{\{s<\mathcal{J}_n\}}\int_t^\infty \lambda(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}))\,\exp\left(-\int_s^v \lambda(\xi_{r-\mathcal{J}_{n-1}}(Y_{n-1}))\,dr\right)\pi\left(\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}), B\right)\,dv\,\middle|\,\mathcal{G}_{n-1}\right].
\end{aligned}$$

almost surely on $\{s \geq \mathcal{J}_{n-1}\}$. Moreover, by the flow property,

$$\xi_{v-\mathcal{J}_{n-1}}(Y_{n-1}) \;=\; \xi_{v-s}(\xi_{s-\mathcal{J}_{n-1}}(Y_{n-1})) \;=\; \xi_{v-s}(X_s) \qquad \text{on } \{s < \mathcal{J}_n\}.$$

Therefore, we obtain

$$\begin{aligned}
&P_{(t_0,x_0)}\left[\{\mathcal{J}_{K_s} > t, Y_{K_s} \in B\} \cap A \cap \{K_s = n\}\right]\\
&\quad = P_{(t_0,x_0)}\left[\{\mathcal{J}_n > t, Y_n \in B\} \cap A_n \cap \{s \in [\mathcal{J}_{n-1}, \mathcal{J}_n)\}\right]\\
&\quad = E_{(t_0,x_0)}\left[P\left[\mathcal{J}_n > t, Y_n \in B \mid \mathcal{G}_{n-1}\right]; A_n \cap \{s \geq \mathcal{J}_{n-1}\}\right]\\
&\quad = E_{(t_0,x_0)}\left[\int_t^\infty \lambda(\xi_{v-s}(X_s))\,\exp\left(-\int_s^v \lambda(\xi_{r-s}(X_s))\,dr\right)\pi\left(\xi_{v-s}(X_s), B\right)\,dv\,;\, A_n \cap \{s \in [\mathcal{J}_{n-1}, \mathcal{J}_n)\}\right]\\
&\quad = E_{(t_0,x_0)}\left[P_{(s,X_s)}\left[\mathcal{J}_1 > t,\, Y_1 \in B\right]\,;\, A \cap \{K_s = n\}\right].
\end{aligned}$$

Summing over n gives the assertion, because $\{s < \zeta\} = \dot{\bigcup}_{n\in\mathbb{N}}\{K_s = n\}$. ■

We are now ready to prove the Markov property for the continuous time process.

Proof (of Theorem 5.1). Let $s \geq t_0$. Then for any $t \geq 0$,

$$X_{s+t} \;=\; \Phi_{s+t}(s, X_s, \mathcal{J}_{K_s}, Y_{K_s}, \mathcal{J}_{K_s+1}, \ldots) \qquad \text{on} \qquad \{K_s < \infty\} = \{s < \zeta\},$$

where Φ_{s+t} is defined by (5.6). In other words, the process $X_{s:\infty} = (X_{s+t})_{t\geq 0}$ from time s onwards is constructed in the same way from $s, X_s, \mathcal{J}_{K_s}, \ldots$ as the original process is constructed from $t_0, Y_0, J_1, \ldots$. Let $F : \mathcal{D}(\mathbb{R}_+, S \cup \{\Delta\}) \to \mathbb{R}$ be bounded and measurable.

Note that $\mathcal{F}_s^X \subseteq \mathcal{G}_{K_s}$. Therefore, by the *strong Markov property* for the Markov chain $(\mathcal{J}_n, Y_n)$, we see that for almost every $\omega \in \{K_s < \infty\} = \{s < \zeta\}$,

$$\begin{aligned}
E_{(t_0,x_0)}\left[F(X_{s:\infty})\,1_{\{s<\zeta\}} \,\middle|\, \mathcal{G}_{K_s}\right](\omega) &= E_{(t_0,x_0)}\left[F \circ \Phi_{s:\infty}(s, X_s, \mathcal{J}_{K_s}, Y_{K_s}, \ldots)\,1_{\{K_s<\infty\}} \mid \mathcal{G}_{K_s}\right](\omega)\\
&= E_{(\mathcal{J}_{K_s}(\omega), Y_{K_s}(\omega))}\left[F \circ \Phi_{s:\infty}(s, X_s(\omega), \mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1, \ldots)\,1_{\{s<\zeta\}}\right]
\end{aligned}$$

and thus, by the tower property of the conditional expectation and by Lemma 5.2,

$$\begin{aligned}
&E_{(t_0,x_0)}\left[F(X_{s:\infty})\,1_{\{s<\zeta\}} \mid \mathcal{F}_s^X\right](\omega)\\
&\quad = E_{(s,X_s(\omega))}\left[E_{(\mathcal{J}_1,Y_1)}\left[F \circ \Phi_{s:\infty}(s, X_s(\omega), \mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1, \ldots)\,1_{\{s<\zeta\}}\right]\right]\\
&\quad = E_{(s,X_s(\omega))}\left[F \circ \Phi_{s:\infty}(s, X_s(\omega), \mathcal{J}_1, Y_1, \mathcal{J}_2, Y_2, \ldots)\right] \;=\; E_{(s,X_s(\omega))}\left[F(X_{s:\infty})\right]
\end{aligned}$$

almost surely on $\{s < \zeta\}$. Here we have used the Markov property in discrete time in the last step.

This completes the proof of the Markov property for the continuous time process $((X_t)_{t\geq t_0}, P_{t_0,x_0})$. Moreover, this Markov process is time-homogeneous, because by construction,

$$P_{(t_0,x_0)} \circ (\mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1, \dots)^{-1} \;=\; P_{(0,x_0)} \circ (t_0 + \mathcal{J}_0, Y_0, t_0 + \mathcal{J}_1, Y_1, \dots)^{-1},$$

and therefore the laws of the processes $((X_{t_0+t})_{t\geq 0}, P_{t_0,x_0})$ and $((X_t)_{t\geq 0}, P_{0,x_0})$ coincide. ■

Generator

Theorem 5.1 shows that the process $((X_t)_{t\geq t_0}, P_{(t_0,x_0)})$ constructed as above is a time-homogeneous Markov process w.r.t. the filtration $(\mathcal{F}_t^X)_{t\geq t_0}$. The transition function is given by

$$p_t(x, B) \;=\; P_{(0,x)}\,[X_t \in B] \qquad \text{for any } t \geq 0,\ x \in S \text{ and } B \in \mathcal{B}.$$

We will now identify the generator of the process. We start with a lemma that follows by conditioning on the first jump of the process.

Lemma 5.3. *For every function $f \in \mathcal{F}_b(S \cup \{\Delta\})$ with $f(\Delta) = 0$, and for all $t \geq 0$ and $x \in S$,*

$$(p_t f)(x) \;=\; e^{-\int_0^t \lambda(\xi_r(x))\,dr} f(\xi_t(x)) \;+\; \int_0^t \lambda(\xi_r(x)) e^{-\int_0^r \lambda(\xi_u(x))\,du} (\pi p_{t-r} f)(\xi_r(x))\,dr\,. \tag{5.7}$$

Proof. Recall that $\mathcal{G}_1 = \sigma(\mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1)$. Since $X_t = \Phi_t(\mathcal{J}_0, Y_0, \mathcal{J}_1, Y_1, \mathcal{J}_2, Y_2, \dots)$, the Markov property for the chain $(\mathcal{J}_n, Y_n)$ implies

$$\begin{aligned}
E_{(0,x)}\,[f(X_t)\,|\,\mathcal{G}_1]\,(\omega) \;&=\; E_{(\mathcal{J}_1(\omega),Y_1(\omega))}\,[f(\Phi_t(0,x,\mathcal{J}_0,Y_0,\mathcal{J}_1,Y_1,\dots))] \\
&=\; f(\xi_t(x))\,1_{t<\mathcal{J}_1(\omega)} \;+\; E_{(\mathcal{J}_1(\omega),Y_1(\omega))}\,[f(X_t)]\;1_{t\geq\mathcal{J}_1(\omega)} \\
&=\; f(\xi_t(x))\,1_{t<\mathcal{J}_1(\omega)} \;+\; \left(p_{t-\mathcal{J}_1(\omega)} f\right)(Y_1(\omega))\,1_{t\geq\mathcal{J}_1(\omega)}.
\end{aligned}$$

for almost every ω. Here we have used that

$$\Phi_t(0,x,\mathcal{J}_0,Y_0,\mathcal{J}_1,Y_1,\dots) \;=\; \begin{cases} \xi_t(x) & \text{for } t < \mathcal{J}_0, \\ \Phi_t(\mathcal{J}_0,Y_0,\mathcal{J}_1,Y_1,\dots) \;=\; X_t & \text{for } t \geq \mathcal{J}_0. \end{cases}$$

Taking expectations w.r.t. $P_{(0,x)}$, we obtain

$$(p_t f)(x) \;=\; f(\xi_t(x))\,P_{(0,x)}[\mathcal{J}_1 > t] \;+\; E_{(0,x)}\left[(p_{t-\mathcal{J}_1} f)(Y_1);\, t \geq \mathcal{J}_1\right].$$

The assertion follows by (5.3) and (5.4). ■

The domain of the generator of a piecewise deterministic Markov process consists of functions that are differentiable along the flow.

Definition 5.4 (Uniform continuity and differentiability along the flow).

1) A bounded measurable function $f : S \to \mathbb{R}$ is called **uniformly continuous along the flow** ξ iff

$$\lim_{t\downarrow 0} \sup_{x\in S} |f(\xi_t(x)) - f(x)| \;=\; 0.$$

We denote by $UC_b(\xi)$ the space of all such functions.

2) A function $f \in UC_b(\xi)$ is called **uniformly continuously differentiable along the flow** ξ iff there exists a function $\mathcal{X} f \in UC_b(\xi)$ such that

$$\frac{d}{dt} f(\xi_t(x))\Big|_{t=0} = (\mathcal{X} f)(x) \qquad \text{for all } x \in S. \tag{5.8}$$

In this case, the function $\mathcal{X} f$ is called the **directional derivative of f along the flow**. The space consisting of all uniformly continuously differentiable functions along the flow ξ is denoted by $UC_b^1(\xi)$.

For a function $f \in UC_b^1(\xi)$, the flow property implies

$$\frac{d}{dt} f(\xi_t(x)) = \frac{d}{dh} f(\xi_h(\xi_t(x)))\Big|_{h=0} = (\mathcal{X} f)(\xi_t(x)) \qquad \text{for all } t \geq 0 \text{ and } x \in S, \qquad \text{and thus}$$

$$f \circ \xi_t = f + \int_0^t (\mathcal{X} f) \circ \xi_s \, ds = f + t \mathcal{X} f + o(t), \tag{5.9}$$

where $o(t)$ stands for a function that is of order $o(t)$ w.r.t. the supremum norm.

Example. Suppose that $S = \mathbb{R}^d$ and ξ is the flow of a vector field $b \in C(\mathbb{R}^d, \mathbb{R}^d)$. Then (5.8) is satisfied for every function $f \in C_b^1(\mathbb{R}^d)$ with $\mathcal{X} f = b^T \nabla f$.

Recall that by Theorem 4.6, the transition function $(p_t)_{t\geq 0}$ induces a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ of linear operators on the Banach space

$$E = \left\{ f \in \mathcal{F}_b(S) : \lim_{t \downarrow 0} \|p_t f - f\|_{\sup} = 0 \right\}.$$

We denote the corresponding generator and its domain by $(L, \mathrm{Dom}(L))$, and we define

$$\mathcal{A} := \left\{ f \in UC_b^1(\xi) : \pi f \in UC_b(\xi) \right\}.$$

Theorem 5.5 (Generator of PDMP).

1) If the intensity function $\lambda : S \to [0, \infty)$ is bounded then $E = UC_b(\xi)$.

2) If, moreover, λ is continuous, then $\mathcal{A} \subseteq \mathrm{Dom}(L)$, and

$$Lf = \mathcal{X} f + \lambda (\pi f - f) \qquad \text{for all } f \in \mathcal{A}. \tag{5.10}$$

In particular, the process $((X_t)_{t\geq 0}, P_{(0,x)})$ solves the martingale problem for $(L, \mathcal{A})$.

The expression (5.10) for the generator is not surprising. The first summand corresponds to the generator of the deterministic motion, and the second part is the generator of a pure jump process with jump intensity λ and jump transition kernel π. The theorem thus shows that we have indeed constructed a Markov process whose generator takes the expected form on functions in $\mathcal{A}$.

Proof. 1) Lemma 5.3 shows that as $t \downarrow 0$, $\sup |p_t f - f \circ \xi_t| \to 0$. Therefore, $p_t f$ converges uniformly to f if and only if f is in $UC_b(\xi)$.

2) Now suppose that λ is in $UC_b(\xi)$. Then as $t \downarrow 0$, $\lambda \circ \xi_t$ converges uniformly to λ, and so

$$e^{-\int_0^t \lambda \circ \xi_r \, dr} \;=\; 1 \;-\; t\lambda \;+\; o(t),$$

where $o(t)$ stands for a function with supremum norm of order $o(t)$. Now let $f \in \mathcal{A}$. Then, by 1), and since $f \mapsto \pi f$ is a contraction w.r.t. the supremum norm,

$$(\pi p_{t-r} f) \circ \xi_r \;=\; (\pi f) \circ \xi_r \;+\; o(1) \;=\; \pi f \;+\; o(1),$$

where the $o(1)$ terms stand for different functions with supremum norm of order $o(1)$, uniformly for $r \in [0, t]$. Therefore, by Lemma 5.3 and (5.9),

$$\begin{aligned} p_t f &= (1 - t\lambda)(f + t\mathcal{X} f) \;+\; t\lambda\pi f \;+\; o(t) \\ &= f \;+\; t\,(\mathcal{X} f + \lambda(\pi f - f)) \;+\; o(t). \end{aligned}$$

Thus f is in the domain of the generator, and (5.10) holds.

■

In general, it is not trivial to identify the precise domain of the generator of a PDMP. In the case $S = \mathbb{R}^d$, one can show under additional regularity assumptions, that $(P_t)_{t \ge 0}$ is a Feller semigroup and $C_0^1(\mathbb{R}^d)$ is a core for the generator. We only state a simple result that holds under quite restrictive assumptions. Let

$$\hat{C}^1(\mathbb{R}^d) \;:=\; \left\{ f \in C^1(\mathbb{R}^d) : f \in \hat{C}(\mathbb{R}^d),\ \partial_i f \in \hat{C}(\mathbb{R}^d) \text{ for all } i = 1, \dots d \right\}.$$

Theorem 5.6 (Identification of Feller generator). Suppose that $S = \mathbb{R}^d$, and ξ is the flow generated by a vector field $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ with bounded derivative ∂b. Moreover, assume that $\lambda \in C_b^1(\mathbb{R}^d)$, and that there exists a finite constant C such that for all $g \in \mathcal{F}_b(\mathbb{R}^d)$,

$$\pi g \in \hat{C}^1(\mathbb{R}^d) \quad \text{with} \quad \sup |\partial(\pi g)| \le C \sup |g|. \tag{5.11}$$

Then the following assertions hold.

1) $P_t(C_0^1(\mathbb{R}^d)) \;\subseteq\; \hat{C}^1(\mathbb{R}^d)$ for all $t \ge 0$.

2) $(P_t)_{t \ge 0}$ is a Feller semigroup.

3) The Feller generator is the closure of the operator $(L, C_0^1(\mathbb{R}^d))$ where

$$Lf \;=\; b^T \nabla f \;+\; \lambda\,(\pi f - f)\,. \tag{5.12}$$

In particular, the corresponding piecewise deterministic Markov process $((X_t)_{t \ge 0}, P_{(0,x_0)})$ is the unique solution of the martingale problem for $(L, C_0^1(\mathbb{R}^d))$ with initial condition x_0.

The assumption (5.11) means that the transition kernel is smoothing the function. For example, it is satisfied if π is the convolution with a non-degenerate Gaussian measure.

Proof (Sketch). 1) Since ∂b is bounded, the flow map ξ_t is a C^1 diffeomorphism for every $t \ge 0$, and the derivative $\partial \xi_t$ is bounded uniformly on finite time intervals. Now let $f \in C_0^1(\mathbb{R}^d)$. By the assumptions, it can be verified that for $0 \le r \le t$, the functions $f \circ \xi_t$ and $(\pi p_{t-r} f) \circ \xi_r$ are in $\hat{C}^1(\mathbb{R}^d)$, and the functions $\lambda \circ \xi_t$ and $\int_0^t \lambda \circ \xi_r \, dr$ are in $C_b^1(\mathbb{R}^d)$. Then by Lemma 5.3, one can conclude that $p_t f$ is contained in $\hat{C}^1(\mathbb{R}^d)$ as well.

2) By 1) and since the subspace $C_0^1(\mathbb{R}^d)$ is dense in $\hat{C}(\mathbb{R}^d)$ w.r.t. the supremum norm, we can conclude that
$$P_t(\hat{C}(\mathbb{R}^d)) \subseteq \hat{C}(\mathbb{R}^d) \qquad \text{for all } t \geq 0.$$
This shows that (P_t) is a Feller semigroup.

3) By Theorem 5.5, the generator of (P_t) is an extension of the operator $(L, C_0^1(\mathbb{R}^d))$ defined by (5.12). To show that $C_0^1(\mathbb{R}^d)$ is a core for the Feller generator, we observe that by 1),
$$P_t(C_0^1(\mathbb{R}^d)) \subseteq \hat{C}^1(\mathbb{R}^d) \subseteq \overline{C_0^1(\mathbb{R}^d)}^L \qquad \text{for all } t \geq 0. \tag{5.13}$$
Here we have used that every function in $\hat{C}^1(\mathbb{R}^d)$ can be approximated in the graph norm by functions in $C_0^1(\mathbb{R}^d)$, because for every compact set K, there exists a finite constant C_K such that
$$\sup_K |Lf| \leq C_K \sup_K (|f| + |\nabla f|).$$
By (5.13) and Theorem 4.15, $C_0^1(\mathbb{R}^d)$ is a core for the generator, and hence by Theorem 4.19, the PDMP is the unique solution of the corresponding martingale problem.

■

We finally return to the examples that we have considered at the beginning of this section.

Example. 1) **Time inhomogeneous jump processes.** The time-space process $(t, X_t)_{t\geq 0}$ of a time inhomogeneous jump process on $\mathbb{R}^d$ is a PDMP with state space $[0,\infty)\times\mathbb{R}^d$ and characteristics
$$\begin{aligned} \xi_t(s,x) &= (s+t, x), \\ \lambda(s,x) &= q_s(x, \mathbb{R}^d), \\ \pi((s,x), dt\,dy) &= \delta_s(dt)\,\lambda(s,x)^{-1} q_s(x,dy). \end{aligned}$$
The flow is generated by the constant vector field $b(t,x) \equiv (1,0,\ldots,0)^T$. It can be verified that if the jump intensity function λ is bounded and measurable, then the full generator of this process is given by
$$(Lf)(t,x) = \frac{\partial f}{\partial t}(t,x) + \int (f(t,y) - f(t,x))\, q_t(x,dy), \qquad f \in \mathrm{Dom}(L),$$
where the domain consists of all functions $f \in \mathcal{F}_b([0,\infty)\times\mathbb{R}^d)$ such that $t \mapsto f(t,x)$ is absolutely continuous for every x with derivative $\frac{\partial f}{\partial t} \in \mathcal{F}_b([0,\infty)\times\mathbb{R}^d)$.

In particular, the time-space process is the unique solution of the martingale problem for this generator. Under additional regularity conditions, this also follows from Theorem 5.6.

Since the first component of the PDMP (t, X_t) is deterministic, the process (X_t) is also a Markov process on its own. This process is time inhomogeneous, and its transition function is given by
$$p_{s,t}(x,B) = p_{t-s}((s,x), \mathbb{R}^d \times B) \qquad \text{for all } 0 \leq s \leq t \text{ and } B \in \mathcal{B}(\mathbb{R}^d).$$
The process solves the time dependent martingale problem for the operators
$$(\mathcal{L}_t g)(x) = \int (g(y) - g(x))\, q_t(x,dy),$$
i.e., for every function $f \in \mathrm{Dom}(L)$ and for every $s \geq 0$ and $x \in \mathbb{R}^d$, the process
$$M_t^f = f(t,X_t) - \int_s^t \left(\frac{\partial f}{\partial r} + \mathcal{L}_r f\right)(r, X_r)\, dr, \qquad t \geq s,$$
is an $(\mathcal{F}_t^X)$ martingale under $P_{(s,x)}$. Furthermore, the forward and backward equations for the semigroup of the time-space process imply the corresponding equations
$$\frac{d}{dt} p_{s,t} f(t,\cdot) = p_{s,t}\left(\frac{\partial f}{\partial t}(t,\cdot) + \mathcal{L}_t f(t,\cdot)\right) \qquad \text{for } t \geq s \text{ and } f \in \mathrm{Dom}(L), \tag{5.14}$$
$$-\frac{d}{ds} p_{s,t} f(t,\cdot) = \mathcal{L}_s\, p_{s,t} f \qquad \text{for } s \leq t \text{ and } f \in \mathcal{F}_b(\mathbb{R}^d). \tag{5.15}$$

2) **Compensated Poisson process.** A compensated Poisson process with constant intensity λ is a PDMP with state space $\mathbb{R}$ and full generator

$$(Lf)(x) \;=\; \lambda\,(f(x+1)-f(x)-f'(x)), \qquad f \in \mathrm{Dom}(L),$$

where the domain consists of all bounded absolutely continuous functions f on $\mathbb{R}$ with bounded measurable derivative f'. The proof is left as an exercise.

3) **Andersen dynamics.** Here the state space is $\mathbb{R}^d \times \mathbb{R}^d$ where $d = mn$, and the vector field generating the flow is $b(x,v) = (v, -\nabla U(x))^T$, i.e., b is the gradient of the Hamiltonian

$$H(x,v) \;=\; U(x) + |v|^2/2.$$

The corresponding PDMP is a Feller process with generator

$$Lf \;=\; v^T\nabla_x f \;-\; \nabla U(x)^T\nabla_v f \;+\; \lambda\,(\pi f - f), \qquad f \in \mathrm{Dom}(L),$$

where the intensity λ is constant, and π is given by (5.2). Note that Theorem 5.6 can not be applied, because the kernel π is smoothing only in the v-component but not on in the x-component. Nevertheless, it can be shown that $C_0^1(\mathbb{R}^d \times \mathbb{R}^d)$ is a core for the generator, see e.g. [23]. Consequently, by Theorem 4.18, a probability measure μ on $\mathbb{R}^d \times \mathbb{R}^d$ is invariant for Andersen dynamics if and only if $\int Lf\,d\mu = 0$ for all $f \in C_0^1(\mathbb{R}^d \times \mathbb{R}^d)$. It can be checked by integration by parts, that this condition is indeed satisfied for the Boltzmann-Gibbs distribution

$$\mu_{BG}(d(x,v)) \;\propto\; \exp(-H(x,v))\,dx\,dv.$$

Localization

To identify the generator, we have assumed above that the jump intensity function λ is bounded. Moreover, we have assumed in our construction that the flow is non-explosive and defined on all of $\mathbb{R}^d$. These assumptions can be relaxed substantially by a localization procedure.

For example, suppose that the state space S is an open subset of $\mathbb{R}^d$, the flow ξ is generated by a smooth vector field $b : S \to \mathbb{R}^d$, and λ is a continuous non-negative function on S. Then for every compact subset $K \subseteq S$, there exist a smooth vector field $b_K : \mathbb{R}^d \to \mathbb{R}^d$ and a bounded continuous function $\lambda_K : \mathbb{R}^d \to [0,\infty)$ such that $b_K = b$ and $\lambda_K = \lambda$ on K. For a given transition kernel π on $\mathbb{R}^d$, we can then construct a PDMP $((X_t^K)_{t\geq 0}, P_{(0,x)})$ corresponding to b_K, λ_K and π in the same way as above for every compact set $K \subset S$ and every initial value $x \in \mathbb{R}^d$. Moreover, the construction ensures that for any two compact sets $K, \widetilde{K} \subset S$, the processes X_t^K and $X_t^{\widetilde{K}}$ coincide up to the minimum of the exit times from the sets K and $\widetilde{K}$, respectively. Therefore, for $x \in S$, we obtain a well-defined process $((X_t)_{t\geq 0}, P_{(0,x)})$ on S by choosing an increasing sequence $(K_n)_{n\in\mathbb{N}}$ of compact subsets of S such that $S = \bigcup K_n$, and setting

$$X_t \;:=\; X_t^{K_n} \qquad \text{for } t < T_n, \qquad \zeta \;:=\; \sup T_n,$$

where $T_n := \inf\{s \geq 0 : X_s^{K_n} \notin K_n\}$.

Exercise (Martingale problem and generator for PDMP with unbounded coefficients). 1) Show that the process $((X_t)_{t\geq 0}, P_{(0,x)})$ solves a local martingale problem for the operator $(L, \mathcal{A})$ where L is given by (5.10) with $\mathcal{X}f = b^T\nabla f$, and $\mathcal{A}$ is an appropriate class of test functions.

2) Conclude that all functions in $C_0^1(S)$ are contained in the domain of the full generator of the process.

3) Show that if the process is non-explosive then it solves a global martingale problem, and for an appropriate class of functions f, the transition function $p_t f$ satisfies the forward equation.

In particular, the construction above can be applied to construct jump processes with unbounded and time-dependent jump rates.

Example (Time-dependent branching). Suppose a population consists initially ($t = 0$) of one particle, and particles die with time-dependent rates $d_t > 0$ and divide into two with rates $b_t > 0$ where $d, b\colon \mathbb{R}^+ \to \mathbb{R}^+$ are continuous functions, and b is bounded. Then the total number X_t of particles at time t is a birth-death process with rates

$$q_t(n,m) = \begin{cases} n \cdot b_t & \text{if} \quad m = n+1, \\ n \cdot d_t & \text{if} \quad m = n-1, \\ 0 & \text{else,} \end{cases} \qquad\qquad \lambda_t(n) = n \cdot (b_t + d_t).$$

The generator is

$$\mathcal{L}_t \;=\; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ d_t & -(d_t + b_t) & b_t & 0 & 0 & 0 & \cdots \\ 0 & 2d_t & -2(d_t + b_t) & 2b_t & 0 & 0 & \cdots \\ 0 & 0 & 3d_t & -3(d_t + b_t) & 3b_t & 0 & \cdots \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Since the rates are unbounded, we have to test for explosion. To this end we consider the Lyapunov function $\psi(n) = n$. We have

$$(\mathcal{L}_t \psi)(n) \;=\; n \cdot b_t \cdot (n + 1 - n) + n \cdot d_t \cdot (n - 1 - n) \;=\; n \cdot (b_t - d_t) \;\leq\; n \sup_{t \geq 0} b_t$$

Since the individual birth rates b_t, $t \geq 0$, are bounded, we can conclude that the process is non-explosive. To study long-time survival of the population, we consider the generating functions

$$G_t(s) = E\left[s^{X_t}\right] = \sum_{n=0}^{\infty} s^n P[X_t = n], \quad 0 < s \leq 1,$$

of the population size. For $f_s(n) = s^n$ we have

$$\begin{aligned} (\mathcal{L}_t f_s)(n) &= n b_t s^{n+1} - n(b_t + d_t)s^n + n d_t s^{n-1} \\ &= \left(b_t s^2 - (b_t + d_t)s + d_t\right) \cdot \frac{\partial}{\partial s} f_s(n) \end{aligned}$$

Since the process is non-explosive and f_s and $\mathcal{L}_t f_s$ are bounded on finite time-intervals, the forward equation holds. We obtain

$$\begin{aligned} \frac{\partial}{\partial t} G_t(s) &= \frac{\partial}{\partial t} E\left[f_s(X_t)\right] = E\left[(\mathcal{L}_t f_s)(X_t)\right] \\ &= (b_t s^2 - (b_t + d_t)s + d_t) \cdot E\left[\frac{\partial}{\partial s} s^{X_t}\right] \\ &= (b_t s - d_t)(s - 1) \cdot \frac{\partial}{\partial s} G_t(s), \\ G_0(s) &= E\left[s^{X_0}\right] = s \end{aligned}$$

The solution of this first order partial differential equation for $s < 1$ is

$$G_t(s) = 1 - \left(\frac{e^{\varrho_t}}{1 - s} + \int_0^t b_u e^{\varrho_u}\, du\right)^{-1}$$

where $\varrho_t := \int_0^t (d_u - b_u)\, du$ is the accumulated rate of decay. In particular, we obtain an explicit formula for the extinction probability: Since $b = d - \varrho'$,

$$P[X_t = 0] \;=\; \lim_{s \downarrow 0} G_t(s) = \left(e^{\varrho_t} + \int_0^t b_n e^{\varrho_u}\, du\right)^{-1} \;=\; 1 - \left(1 + \int_0^t d_u e^{\varrho_u}\, du\right)^{-1}.$$

Thus we have shown:

$$P[X_t = 0 \text{ eventually}] = 1 \iff \int_0^{\infty} d_u e^{\varrho_u}\, du = \infty.$$

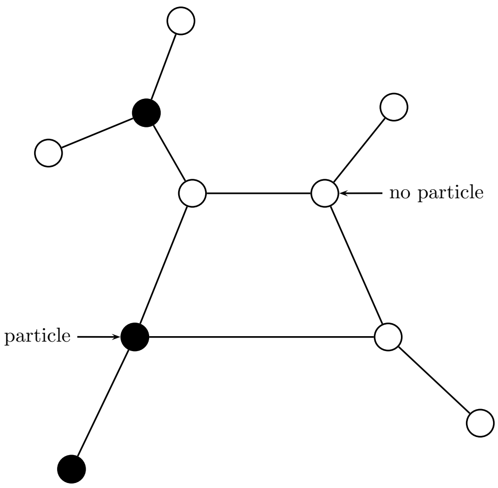
5.2. Interacting particle systems on finite graphs

Let $G = (V, E)$ be an (undirected) graph with V the set of vertices and E the set of edges. We write $x \sim y$ if and only if $\{x, y\} \in E$. We call

$$S = T^V = \{\eta \colon V \to T\}$$

the **configuration space**. T can be the space of types, states, spins etc.
E.g.

$$T = \{0, 1\}, \qquad \eta(x) = \begin{cases} 1 & \text{particle at } x \\ 0 & \text{no particle at } x \end{cases}$$



Markovian dynamics: $\eta(x)$ changes to state i with rate

$$c_i(x, \eta) = g_i\left((\eta(x), (\eta(y))_{y \sim x}\right)$$

i.e.

$$q(\eta, \xi) = \begin{cases} c_i(x, \eta) & \text{if } \xi = \eta^{x,i} \\ 0 & \text{otherwise} \end{cases}$$

where

$$\eta^{x,i}(y) = \begin{cases} \eta(y) & \text{for } y \neq x \\ i & \text{for } y = x \end{cases}$$

Example. (i) **Contact process:** (Spread of plant species, infection,...) $T = \{0, 1\}$. Each particle dies with rate $d > 0$, produces descendent at any neighbor site with rate $b > 0$ (if not occupied)

$$c_0(x, \eta) = d$$
$$c_1(x, \eta) = b \cdot N_1(x, \eta); \qquad N_1(x, \eta) := |\{y \sim x \ : \ \eta(y) = 1\}|$$

Spatial branching process with exclusion rule (only one particle per site).

(ii) **Voter model:** $\eta(x)$ opinion of voter at x,

$$c_i(x, \eta) \ = \ N_i(x, \eta) \ := \ |\{y \sim x \ : \ \eta(y) = i\}|$$

changes opinion to i with rate equal to number of neighbors with opinion i.

(iii) **Ising model with Glauber (spin flip) dynamics:** $T = \{-1, 1\}$, $\beta > 0$ inverse temperature.

a) *Metropolis dynamics:*

$$\Delta(x, \eta) := \sum_{y \sim x} \eta(y) = N_1(x, \eta) - N_{-1}(x, \eta) \quad \text{total magnetization}$$

$$c_1(x, \eta) := \min\left(e^{2\beta \cdot \Delta(x,\eta)}, 1\right), \qquad c_{-1}(x, \eta) := \min\left(e^{-2\beta \cdot \Delta(x,\eta)}, 1\right).$$

b) *Heath bath dynamics:*

$$c_1(x, \eta) = \frac{e^{\beta\Delta(x,\eta)}}{e^{\beta\Delta(x,\eta)} + e^{-\beta\Delta(x,\eta)}} \qquad c_{-1}(x, \eta) = \frac{e^{-\beta\Delta(x,\eta)}}{e^{\beta\Delta(x,\eta)} + e^{-\beta\Delta(x,\eta)}}.$$

$\beta = 0$ (infinite temperature): $c_1 \equiv c_{-1} \equiv \frac{1}{2}$. Random walk on hypercube $\{0, 1\}^V$.

$\beta \to \infty$ (zero temperature):

$$c_1(x, \eta) = \begin{cases} 1 & \text{if } \Delta(x, \eta) > 0 \\ \frac{1}{2} & \text{if } \Delta(x, \eta) = 0 \\ 0 & \text{if } \Delta(x, \eta) < 0 \end{cases}, \qquad c_0(x, \eta) = \begin{cases} 1 & \text{if } \Delta(x, \eta) < 0 \\ \frac{1}{2} & \text{if } \Delta(x, \eta) = 0 \\ 0 & \text{if } \Delta(x, \eta) > 0 \end{cases}.$$

Voter model with majority vote.

We now assume at first that the vertex set V is finite. In this case, the configuration space $S = T^V$ is finite-dimensional. If, moreover, the type space T is also finite then S itself is a finite graph w.r.t. the **Hamming distance**

$$d(\eta, \xi) = |\{x \in V \,;\, \eta(x) \neq \xi(x)\}|$$

Hence a continuous time Markov chain (η_t, P_x) on the configuration space can be constructed as above from the jump rates $q_t(\xi, \eta)$. The process is non-explosive, and the asymptotic results for Markov chains with finite state space apply. In particular, if irreducibility holds then there exists a unique invariant probability measure, and the ergodic theorem applies.

Example. (i) **Ising Model:** The Boltzmann distribution

$$\mu_\beta(\eta) = Z_\beta^{-1}\, e^{-\beta H(\eta)}, \qquad Z_\beta = \sum_\eta e^{-\beta H(\eta)},$$

with Hamiltonian

$$H(\eta) = \frac{1}{2} \sum_{\{x,y\} \in E} (\eta(x) - \eta(y))^2 = \sum_{\{x,y\} \in E} \eta(x)\eta(y) + |E|$$

is invariant, since it satisfies the detailed balance condition

$$\mu_\beta(\eta) q(\eta, \xi) = \mu_\beta(\xi) q(\xi, \eta) \quad \forall\, \xi, \eta \in S.$$

Moreover, irreducibility holds - so the invariant probability measure is unique, and the ergodic theorem applies.

(ii) **Voter model:** The constant configurations $\underline{i}(x) \equiv i$, $i \in T$, are *absorbing states*, i.e. $c_j(x, \underline{i}) = 0$ for all $x \in V$ and $j \neq i$. Any other state is transient, so

$$P\left[\bigcup_{i \in T} \{\eta_t = \underline{i} \text{ eventually}\}\right] = 1.$$

Moreover, the integer valued process

$$N_i(\eta_t) := |\{x \in V : \eta_t(x) = i\}|$$

is a martingale (Exercise), so

$$\begin{aligned} N_i(\xi) &= E_\xi[N_i(\eta_t)] \stackrel{t\to\infty}{\longrightarrow} E_\xi[N_i(\eta_\infty)] = |V| \cdot P_\xi[\eta_t = \underline{i} \text{ eventually}], \quad \text{i.e.,} \\ & P[\eta_t = \underline{i} \text{ eventually}] = N_i(\eta)/|V|. \end{aligned}$$

The invariant probability measures are the Dirac measures $\delta_{\underline{i}}$, $i \in T$, and their convex combinations.

(iii) **Contact process:** The configuration $\underline{z}$ is absorbing, all other states are transient. Hence $\delta_{\underline{z}}$ is the unique invariant measure and ergodicity holds.

We see that on finite graphs the situation is rather simple as long as we are only interested in existence and uniqueness of invariant measures, and ergodicity. Below, we will show that on infinite graphs the situation is completely different, and phase transitions occur. On finite subgraphs of an infinite graph these phase transitions effect the rate of convergence to the stationary distribution and the variances of ergodic averages but not the ergodicity properties themselves.

Mean field models

Suppose that G is the complete graph with n vertices, i.e.

$$V = \{1, \dots, n\} \quad \text{and} \quad E = \{\{x, y\} \ : \ x, y \in V\}$$

Let

$$L_n(\eta) = \frac{1}{n} \sum_{x=1}^{n} \delta_{\eta(x)}$$

denote the **empirical distribution** of a configuration $\eta \colon \{1, \dots, n\} \to T$, the *mean field*. In a **mean-field model** the rates

$$c_i(x, \eta) = f_i(L_n(\eta))$$

are *independent of* x, and *depend on* η *only through the* mean field $L_n(\eta)$.

Example. Multinomial resampling (e.g. population genetics), mean field voter model.
With rate 1 replace each type $\eta(x)$, $x \in V$, by a type that is randomly selected from $L_n(\eta)$:

$$c_i(x, \eta) = L_n(\eta)(i) = \frac{1}{n} |\{x \in V \ : \ \eta(x) = i\}|$$

As a special case we now consider mean-field models with type space $T = \{0, 1\}$ or $T = \{-1, 1\}$. In this case the empirical distribution is completely determined by the frequence of type 1 in a configuration:

$$\begin{aligned} L_n(\eta) \longleftrightarrow N_1(\eta) &= |\{x \ : \ \eta(x) = 1\}| \\ c_i(x, y) &= \widetilde{f}_i(N_1(\eta)) \end{aligned}$$

If (η_t, P_x) is the corresponding mean field particle system, then (Exercise) $X_t = N_1(\eta)$ is a birth-death process on $\{0, 1, \dots, n\}$ with birth/death rates

$$b(k) = (n - k) \cdot \widetilde{f}_1(k), \qquad\qquad d(k) = k \cdot \widetilde{f}_0(k)$$

where $(n - k)$ is the number of particles with state 0 and $\widetilde{f}_1(k)$ is the birth rate per particle.
$\rightsquigarrow$ Explicit computation of hitting times, stationary distributions etc.!

Example. (i) **Binomial resampling:** For multinomial resampling with $T = \{0, 1\}$ we obtain

$$b(k) = d(k) = \frac{k \cdot (n-k)}{n}$$

(ii) **Mean-field Ising model:** For the Ising model on the complete graph with inverse temperature β and interaction strength $\frac{1}{n}$ the stationary distribution is

$$\mu_\beta(\eta) \propto e^{-\frac{\beta}{4n}\sum_{x,y}(\eta(x)-\eta(y))^2} \propto e^{\frac{\beta}{2n}\sum_x \eta(x)\cdot\sum_y \eta(y)} = e^{\frac{\beta}{2n}m(\eta)^2}$$

where

$$m(\eta) = \sum_{x=1}^{n} \eta(x) = N_1(\eta) - N_{-1}(\eta) = 2N_1(\eta) - n$$

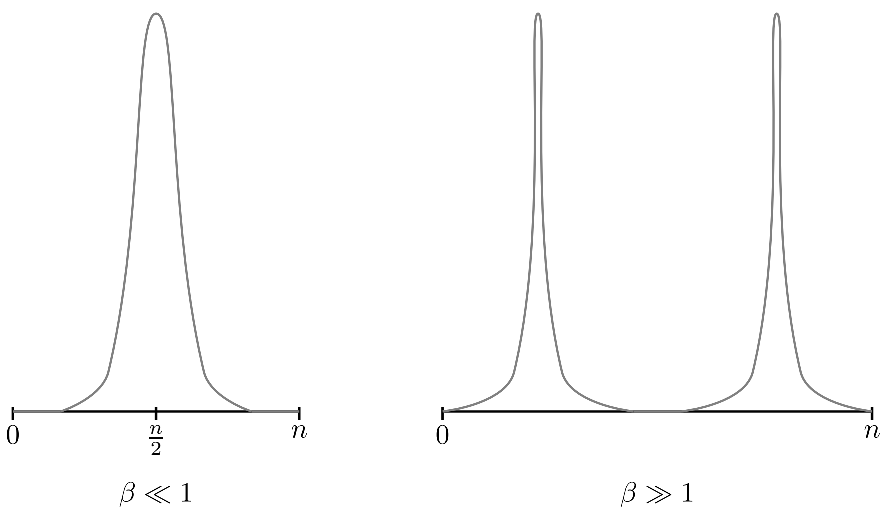
is the **total magnetization**. Note that each $\eta(x)$ is interacting with the mean field $\frac{1}{n}\sum \eta(y)$, which explains the choice of interacting strength of order $\frac{1}{n}$. The birth-death chain $N_1(\eta_t)$ corresponding to the heat bath dynamics has birth and death rates

$$b(k) = (n-k)\cdot\frac{e^{\beta\frac{k}{n}}}{e^{\beta\frac{k}{n}} + e^{\beta\frac{n-k}{n}}}, \qquad d(k) = k\cdot\frac{e^{\beta\frac{n-k}{n}}}{e^{\beta\frac{k}{n}} + e^{\beta\frac{n-k}{n}}}$$

and stationary distribution

$$\bar{\mu}_\beta(k) = \sum_{\eta\,:\,N_1(\eta)=k} \mu_\beta(\eta) \propto \binom{n}{k} 2^{-n} e^{\frac{2\beta}{n}\left(k-\frac{n}{2}\right)^2}, \quad 0 \le k \le n$$

The binomial distribution $\mathrm{Bin}(n, \frac{1}{2})$ has a maximum at its mean value $\frac{n}{2}$, and standard deviation $\frac{\sqrt{n}}{2}$. Hence for large n, the measure $\bar{\mu}_\beta$ has one sharp mode of standard deviation $O(\sqrt{n})$ if β is small, and two modes if β is large:



The transition from uni- to multimodality occurs at an inverse temperature β_n with

$$\lim_{n\to\infty} \beta_n = 1 \qquad \text{(Exercise)}$$

The asymptotics of the stationary distribution as $n \to \infty$ can be described more accurately using large deviation results, cf. below.

Now consider the heat bath dynamics with an initial configuration η_0 with $N_1(\eta_0) \leq \frac{n}{2}$, n even, and let

$$T := \inf\left\{t \geq 0 \ : \ N_1(\eta_t) > \frac{n}{2}\right\}.$$

By the formula for mean hitting times for a birth-and-death process,

$$E[T] \geq \frac{\bar{\mu}_\beta\left(\left\{0, 1, \dots, \frac{n}{2}\right\}\right)}{\bar{\mu}_\beta\left(\frac{n}{2}\right) \cdot b\left(\frac{n}{2}\right)} \geq \frac{\frac{1}{2}}{\bar{\mu}_\beta\left(\frac{n}{2}\right) \cdot \frac{n}{2}} \geq \frac{e^{\beta \frac{n}{2}}}{n2^n}$$

since

$$\bar{\mu}_\beta\left(\frac{n}{2}\right) = \binom{n}{\frac{n}{2}} \cdot e^{-\frac{\beta n}{2}} \bar{\mu}_\beta(0) \leq 2^n e^{-\frac{\beta n}{2}}.$$

Hence the average time needed to go from configurations with negative magnetization to states with positive magnetization is increasing exponentially in n for $\beta > 2 \log 2$. Thus although ergodicity holds, for large n the process gets stuck for a very large time in configurations with negative resp. positive magnetization.
$\rightsquigarrow$ *Metastable behaviour*.

More precisely, one can show using large deviation techniques that metastability occurs for any inverse temperature $\beta > 1$, cf. below.

5.3. Interacting particle systems on $\mathbb{Z}^d$

In this section we consider an explicit construction for a class of continuous time Markov processes with unbounded jump intensities. Since these processes may have infinitely many jumps in a finite time interval, the construction can not be carried out as easily as for processes with finite jump intensity. This section is partially based on the lecture notes [14] by R. Durrett and the monograph [35] by T. Liggett.

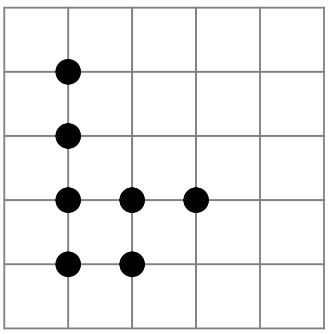
Graphical construction of interacting particle systems

In the setup introduced above, we now consider the case where $V = \mathbb{Z}^d$ and $E = \{(x, y) \ : \ |x - y| = 1\}$ is the usual graph structure on $\mathbb{Z}^d$. The state space T for each particle is a finite set. We endow the configuration space

$$S = T^{\mathbb{Z}^d} = \left\{\eta : \mathbb{Z}^d \to T\right\}$$

with the product topology, i.e., a sequence $(\eta_k)_{k \in \mathbb{N}}$ of configurations in S converges to a limit configuration η if and only if $\eta_k(x) \to \eta(x)$ for any $x \in \mathbb{Z}^d$. Since T is finite, this means that locally (on finite subsets of $\mathbb{Z}^d$), convergent sequences of configurations are eventually constant:

$$\eta_k \to \eta \ \Leftrightarrow \ \text{For every } x \in \mathbb{Z}^d, \quad \eta_k(x) = \eta(x) \text{ eventually.}$$



Throughout this section, we impose the following assumptions on the jump rates:

(A1) *Uniformly bounded jump rates:* There exists $\lambda \in (0,\infty)$ such that

$$c_i(x,\eta) \;\leq\; \lambda \qquad \text{for all } i \in T, x \in \mathbb{Z}^d \text{ and } \eta \in S.$$

(A2) *Translation invariance and nearest neighbour interactions:* There exist functions $g_i : T^{2d+1} \to [0,\infty)$ such that

$$c_i(x,\eta) \;=\; g_i\left(\eta(x),(\eta(y))_{y\sim x}\right) \qquad \text{for all } i \in T,\ x \in \mathbb{Z}^d \text{ and } \eta \in S.$$

Note that since there are infinitely many particles, the total jump intensity is usually infinite. We now give an explicit construction of an interacting particle system corresponding to the given jump rates. To this end, we introduce a family $(N_t^{x,i})_{t\geq 0}$ $(x \in \mathbb{Z}^d,\ i \in T)$ of independent Poisson processes with rate λ. At each of the arrival times

$$T_n^{x,i} \;=\; \inf\left\{t \geq 0 : N_t^{x,i} = n\right\} \qquad (n \in \mathbb{N},\ x \in \mathbb{Z}^d,\ i \in T),$$

a jump of the particle at x to state i is proposed. Note that almost surely, all the arrival times are different, and thus there is no ambiguity about which of two proposed jumps should be taken into account first. To obtain the correct transition rates, proposed jumps are carried out with probability $c_i(x,\eta)/\lambda$. To decide whether the jumps are carried out, we introduce an independent family of independent random variables $U_n^{x,i}$ $(x \in \mathbb{Z}^d, i \in T)$ that are uniformly distributed on $(0,1)$. Now, we would like to carry out transitions according to the following rule:

Recipe: At time $T_n^{x,i}$, change $\eta(x)$ to i provided $U_n^{x,i} \leq c_i(x,\eta)/\lambda$.

The problem with this approach is however, that since there are infinitely many Poisson processes, there will be infinitely many proposed transitions in each finite time interval. In particular, there is *no first transition*, and so it is not possible to construct the transitions sequentially!

So how can we nevertheless consistently define a process from the jump times? The idea is to use that during a small time interval, the state $\eta_t(x)$ at position x is only influenced by the particles in a finite neighbourhood of x. More precisely, we will see that if $t - s \geq 0$ is sufficiently small, then $\mathbb{Z}^d$ splits into almost surely finite random components that do not influence each other between times s and t. Then the construction of η_t from η_s can be carried out independently on each of these components.

For a finite subset $A \subset \mathbb{Z}^d$ and $\xi \in S$, the restricted configuration space

$$S_{\xi,A} \;:=\; \{\eta \in S : \eta = \xi \text{ on } A^c\}$$

is finite. Hence for all $s \geq 0$, there exists a unique Markov jump process $(\eta_t^{(s,\xi,A)})_{t\geq s}$ on $S_{\xi,A}$ with initial condition $\eta_s^{(s,\xi,A)} = \xi$ and transitions $\eta \to \eta^{x,i}$ at times $T_n^{x,i}$ whenever $x \in A$ and $U_n^{x,i} \leq c_i(x,\eta)/\lambda$. Our goal is now to define a Markov process $\eta_t^{(s,\xi)}$ on S for $t - s$ small by setting

$$\eta_t^{(s,\xi)}(x) \;:=\; \eta_t^{(s,\xi,A)}(x)$$

where A is an appropriately chosen finite neighborhood of x. The neighborhood should be chosen in such a way that during the considered time interval, the configuration $\eta_t^{(s,\xi)}$ restricted to A has only been effected by previous values on A. That this is possible is guaranteed by the crucial observation in the next lemma. For $0 \leq s \leq t$ we define a random subgraph $(\mathbb{Z}^d, E_{s,t}(\omega))$ of (V,E) by setting

$$E_{s,t} \;=\; \left\{\{x,y\} \in E \;:\; \text{For some } n \in \mathbb{N} \text{ and } i \in T,\ T_n^{x,i} \in (s,t] \text{ or } T_n^{y,i} \in (s,t]\right\}$$

If the state of the particle at x effects the state of the particle at y during the time interval $(s,t]$ or vice versa then $\{x,y\} \in E_{s,t}$.

Lemma 5.7. *If*

$$t - s \;\le\; \frac{1}{8 \cdot d^2 \cdot |T| \cdot \lambda} \;=:\; \delta$$

then

$$P\left[\textit{all connected components of } (\mathbb{Z}^d, E_{s,t}) \textit{ are finite}\right] \;=\; 1.$$

Proof. By translation invariance it suffices to show

$$P[|C_0^{s,t}| < \infty] \;=\; 1$$

where $C_0^{s,t}$ is the component of $(\mathbb{Z}^d, E_{s,t})$ containing 0. If x is in $C_0^{s,t}$ then there exists a self-avoiding path in $(\mathbb{Z}^d, E_{s,t})$ starting at 0 with length $\ge \|x\|_{l^1}$. Hence

$$\begin{aligned} &P[\exists\, x \in C_0^{s,t} \;:\; \|x\|_{l^1} \ge 2n-1] \\ &\le\;\; P[\exists \text{ self-avoiding path } z_0 = 0, z_1, \dots, z_{2n-1} \text{ s.t. } (z_k, z_{k+1}) \in E_{s,t} \;\forall\, i] \\ &\le\;\; 2d(2d-1)^{2n-2} \prod_{i=0}^{n-1} P[(z_{2k}, z_{2k+1}) \in E_{s,t}] \end{aligned}$$

Here we have used that the events $\{(z_{2k}, z_{2k+1}) \in E_{s,t}\}$, $k = 0, 1, \dots, n-1$, are independent, and $2d(2d-1)^{2n-2}$ is an upper bound for the number of self-avoiding paths $z_0, z_1, \dots z_{2n-1}$ in $\mathbb{Z}^d$ starting at 0. Hence,

$$P\left[\exists\, x \in C_0^{s,t} : \|x\|_{l^1} \ge 2n-1\right] \;\le\; \frac{2d}{(2d-1)^2}\left((2d-1)^2(1 - e^{-2|T|\lambda(t-s)})\right)^n$$

where $e^{-2|T|\lambda(t-s)}$ is the probability that no arrival occurs during $[s,t]$ in the $2|T|$ independent Poisson processes $N^{z_{2k},i}$ and $N^{z_{2k+1},i}$, $i \in T$. Using the upper bound $1 - e^{-x} \le x$, we see that for $t - s \le \delta$, the right hand side converges to 0 as $n \to \infty$. Hence in this case, $C_0^{s,t}$ is almost surely finite. ■

Remark (Connection to percolation). The random graphs $(\mathbb{Z}^d, E_{s,t})$ actually form a dynamic percolation model. One can show that there is a phase transition: If $t - s$ is large then the random graph almost surely contains an infinite component.

As a consequence of the lemma, for time intervals $[t, t+h]$ of length $h \le \delta$, we can construct the configuration up to time $t+h$ from the configuration at time t in each component of the random graph by the standard construction for jump processes with finite state space. Iterating this procedure, we obtain the dynamics for all times.

Theorem 5.8 (Construction of IPS on $\mathbb{Z}^d$). Let $s \in [0, \infty)$ and $\xi \in T^{\mathbb{Z}^d}$. Then there exists a unique (up to equivalence) stochastic process $(\eta_t^{(s,\xi)})_{t \ge s}$ with state space $T^{\mathbb{Z}^d}$ such that almost surely,

(i) $\eta_s^{(s,\xi)} = \xi$

(ii) For every $t \in [s, \infty)$ and $h \in [0, \delta]$, and for each connected component C of $(\mathbb{Z}^d, E_{t,t+h})$, the configuration $\eta_{t+h}^{(s,\xi)}\big|_C$ is obtained from $\eta_t^{(s,\xi)}\big|_C$ by subsequently taking into account the finite number of transitions in C during $[t, t+h]$.

We do not carry out the relatively straightforward details of the proof. Note that the theorem allows not only to construct the process for a given initial configuration, but indeed we obtain the processes starting from arbitrary initial configurations on a joint probability space. In other words, we have really constructed the stochastic flow of the interacting particle system. More generally, we can even realize the dynamics for different jump rates that are bounded by λ on a joint probability space, i.e., the construction automatically provides a coupling between interacting particle systems with different transition rates. Let

$$\eta_t^\xi \;:=\; \eta_t^{(0,\xi)}$$

denote the process started at time 0. The cocycle property

$$\eta_t^\xi \;=\; \eta_t^{(s,\eta_s^\xi)} \qquad \text{for all } 0 \le s \le t \tag{5.16}$$

follows directly from of our construction of the process. As a consequence, one can verify that the interacting particle system is a Markov process.

Corollary 5.9. The stochastic dynamics constructed in Theorem 5.8 has the following properties.

(i) **Time-homogeneity:** $\left(\eta_{s+t}^{(s,\xi)}\right)_{t\ge 0} \sim \left(\eta_t^\xi\right)_{t\ge 0}$

(ii) **Markov property:** (η_t^ξ, P) is a time homogeneous Markov process with transition semigroup

$$(p_t f)(\xi) \;=\; E[f(\eta_t^\xi)]$$

(iii) **Feller property:** $f \in C_b(S) \implies p_t f \in C_b(S)\ \forall\, t \ge 0$

(iv) **Spatial homogeneity:** Let $\xi\colon \Omega \to S$ be a random variable that is independent of the Poisson processes $(N_t^{x,i})$ and the random variables $U_n^{x,i}$ for all x, i and n. If the law of ξ is *translation invariant*, i.e., $\xi(x+\bullet) \sim \xi$ for all $x \in \mathbb{Z}^d$, then the law of η_t^ξ is *translation invariant* for all $t \ge 0$.

We remark that by compactness, functions on S that are continuous w.r.t. the product topology are automatically bounded. Thus the Feller property means that p_t maps continuous functions to continuous functions.

Proof (Sketch). (i) follows from the time homogeneity of the Poisson processes $(N_t^{x,i})$ and the fact that the random variables $U_n^{x,i}$ are i.i.d.

(ii) Let $0 \le s \le t$ and $f \in \mathcal{F}_+(S)$. Then by the cocycle property (5.16),

$$\begin{aligned} E\left[f\left(\eta_t^\xi\right) \,\middle|\, \mathcal{F}_s\right](\omega) \;&=\; E\left[f\left(\eta_t^{\left(s,\eta_s^\xi\right)}\right) \,\middle|\, \mathcal{F}_s\right](\omega) \;=\; E\left[f\left(\eta_t^{\left(s,\eta_s^\xi(\omega)\right)}\right)\right] \\ &=\; E\left[f\left(\eta_{t-s}^{\eta_s^\xi(\omega)}\right)\right] \;=\; (p_{t-s}f)\left(\eta_s^\xi(\omega)\right). \end{aligned}$$

Here we have used in the second step that $\eta_t^{\left(s,\eta_s^\xi\right)}$ is a function of η_s^ξ and the arrivals of the Poisson process after time s which are independent of $\mathcal{F}_s$.

(iii) We want to show that for $f \in C_b(S)$ and $t \ge 0$, $\xi \mapsto (p_t f)(\xi)$ is continuous. Hence let ξ_n be a sequence in S such that $\xi_n \to \xi$ w.r.t. the product topology. Then for every $x \in \mathbb{Z}^d$, $\xi_n(x) \to \xi(x)$. Since T is finite, this implies that $\xi_n = \xi$ eventually on each finite set $C \subset \mathbb{Z}^d$, and hence on each component of $\left(\mathbb{Z}^d, E_{0,\delta}\right)$. Thus if $t \le \delta$ then by the componentwise construction of the dynamics up to time δ,

$$\eta_t^{\xi_n}(x) \;=\; \eta_t^\xi(x) \quad \text{eventually}$$

holds with probability one for every $x \in \mathbb{Z}^d$, i.e.,

$$\eta_t^{\xi_n} \to \eta_t^{\xi} \quad \text{almost surely.}$$

Thus for $f \in C_b(S)$ and $t \le \delta$, $f\left(\eta_t^{\xi_n}\right) \to f\left(\eta_t^{\xi}\right)$, and hence

$$(p_t f)(\xi_n) \;=\; E\left[f\left(\eta_t^{\xi_n}\right)\right] \;\longrightarrow\; (p_t f)(\xi)$$

follows by dominated convergence. Thus the Feller property holds for $t \le \delta$. Finally, for arbitrary $t \ge 0$, the Feller property holds since by the semigroup property, $p_t = p_{t-\delta\lfloor t/\delta\rfloor} p_\delta^{\lfloor t/\delta\rfloor}$.

(iv) Spatial homogeneity follows from the assumption that the transition rates $c_i(x,\eta)$ are translation invariant (A2), and the fact that the random variables

$$\left(\left(T_n^{x,i}\right)_{n,i}, \left(U_n^{x,i}\right)_{n,i}\right), \qquad x \in \mathbb{Z}^d,$$

are identically distributed. ■

Identification of the generator

By Corollary 5.9, the interacting particle systems constructed above induce Feller semigroups on the Banach space $E = C(T^{\mathbb{Z}^d})$ endowed with the supremum norm. Our next goal is to identify the corresponding generators L. Although the formal expression for the generator is easy to guess, it will be a bit tricky to identify the correct domain. In infinite dimensions, the rôle of smooth test functions is often taken over by *cylinder functions*, i.e., finitely based functions. On the configuration space $T^{\mathbb{Z}^d}$, these take the form

$$f(\xi) \;=\; \varphi(\xi(x_1), \xi(x_2), \ldots, \xi(x_k)) \tag{5.17}$$

for some $k \in \mathbb{N}$, $x_1, \ldots x_k \in \mathbb{Z}^d$ and a function $\varphi : T^k \to \mathbb{R}$.

Lemma 5.10. *The linear subspace C consisting of all cylinder functions is dense in $C(T^{\mathbb{Z}^d})$ w.r.t. the supremum norm.*

Proof. Let $A_n = [-n,n]^d \cap \mathbb{Z}^d$. For $\xi \in T^{\mathbb{Z}^d}$ and $n \in \mathbb{N}$, we define $\xi_n \in T^{\mathbb{Z}^d}$ by

$$\xi_n(x) \;=\; \begin{cases} \xi(x) & \text{for } x \in A_n, \\ t_0 & \text{for } x \notin A_n, \end{cases}$$

where t_0 is an arbitrary fixed element in T. Since for every fixed $x \in \mathbb{Z}^d$, $\xi_n(x) = \xi(x)$ eventually, the sequence $(\xi_n)_{n\in\mathbb{N}}$ converges pointwise to ξ.

We now approximate an arbitrary function $f \in C(T^{\mathbb{Z}^d})$ by the cylinder functions $f_n(\xi) := f(\xi_n)$, $n \in \mathbb{N}$. By continuity of f (w.r.t. the product topology, i.e., pointwise convergence),

$$f_n(\xi) \;\to\; f(\xi) \qquad \text{for all } \xi \in T^{\mathbb{Z}^d}.$$

Since $T^{\mathbb{Z}^d}$ is compact, and all the functions are continuous, this pointwise convergence of functions on $T^{\mathbb{Z}^d}$ implies uniform convergence. Thus f is indeed a uniform limit of cylinder functions. ■

In a first step, we now identify the generator on cylinder functions. The advantage of cylinder functions is that the infinite sum over all possible jumps occurring in the formal expression (5.18) for the generator has only finitely many non-zero entries.

Theorem 5.11 (Generator on cylinder functions). Every cylinder function f of the form (5.17) is in the domain of the Feller generator L, and $Lf = \mathcal{L}f$ where

$$(\mathcal{L}f)(\xi) \;=\; \sum_{\substack{x\in\mathbb{Z}^d\\ i\in T}} c_i(x,\xi)\left(f(\xi^{x,i}) - f(\xi)\right). \tag{5.18}$$

Proof. The process $\bar{N}_t := \sum_{j=1}^k \sum_{i\in T} N_t^{x_j,i}$ counts all possible jumps that can change the value of the cylinder function f. By the superposition principle for Poisson processes, this process is again a Poisson process with finite intensity $k|T|\lambda$. Thus as $t\downarrow 0$, the probability $P[\bar{N}_t > 1]$ is of order $O(t^2)$. Therefore,

$$\begin{aligned}
(p_t f)(\xi) - f(\xi) &= E[f(\eta_t^\xi) - f(\xi)] \;=\; E[f(\eta_t^\xi) - f(\xi); \bar{N}_t = 1] \;+\; O(t^2)\\
&= \sum_{j,i}\left(f(\xi^{x_j,i}) - f(\xi)\right)\cdot P\left[N_t^{x_j,i} = 1,\; U_1^{x_j,i} \le c_i(x,\xi)/\lambda\right] + O(t^2)\\
&= \sum_{j,i} \lambda t \cdot \frac{c_i(x_j,\xi)}{\lambda}\cdot\left(f(\xi^{x_j,i}) - f(\xi)\right) + O(t^2)\\
&= f(\xi) + t\cdot(\mathcal{L}f)(\xi) + O(t^2)
\end{aligned}$$

where the constants of order $O(t^2)$ can be chosen independently of ξ. ■

Since the space C of all cylinder functions is dense in the Banach space $E = C(T^{\mathbb{Z}^d})$, Theorem 5.13 implies in particular that the transition function $(p_t)_{t\ge 0}$ induces a *strongly continuous* contraction semigroup $(P_t)_{t\ge 0}$ of linear operators on E. Moreover, the generator $(L, \mathrm{Dom}(L))$ extends the operator $(\mathcal{L}, C)$. In order to identify the precise domain of the generator, we introduce another norm on functions $f : T^{\mathbb{Z}^d} \to \mathbb{R}$. For $x\in\mathbb{Z}^d$ let

$$\Delta f(x) \;:=\; \sup\left\{|f(\xi^{x,i}) - f(\xi)| \;:\; i\in T,\, \xi\in T^{\mathbb{Z}^d}\right\}$$

denote the maximal change of the value of f if only the state of the particle at x is varied. Then the *triple norm* of f is defined as

$$|||f||| \;:=\; \sum_{x\in\mathbb{Z}^d} \Delta f(x) \;=\; \|\Delta f\|_{\ell^1(\mathbb{Z}^d)}.$$

The triple norm has some similarity to a Lipschitz norm, although the supremum and the sum are taken in a reverse order. Let

$$\mathcal{A} \;:=\; \left\{f\in C(T^{\mathbb{Z}^d}) \;:\; |||f||| < \infty\right\}.$$

Lemma 5.12. *The following bounds hold for every $f\in\mathcal{A}$.*

(i) For every $\xi\in T^{\mathbb{Z}^d}$, the series (5.18) *defining $(\mathcal{L}f)(\xi)$ is absolutely convergent, and*

$$\|\mathcal{L}f\|_{\sup} \;\le\; \lambda\,|T|\;|||\,f\,|||\,.$$

(ii) For every $t\in[0,\delta]$,

$$|||P_t f||| \;\le\; E\left[|C_0^{0,t}|\right]\;|||\,f\,|||\,.$$

Proof. (i) holds since by Assumption (A1),

$$\sum_{x\in\mathbb{Z}^d}\sum_{i\in T} c_i(x,\xi)\,|f(\xi^{x,i}) - f(\xi)| \;\le\; \lambda\,|T|\sum_{x\in\mathbb{Z}^d}\Delta f(x).$$

(ii) By construction, for $t \leq \delta$, a modification of the initial configuration at position x only effects the configuration at time t on the component $C_x^{0,t}$. Therefore,

$$\left|(P_t f)(\xi^{x,i}) - (P_t f)(\xi)\right| \;\leq\; E\left[|f(\eta_t^{\xi^{x,i}}) - f(\eta_t^{\xi})|\right] \;\leq\; E\left[\sum_{y \in C_x^{0,t}} \Delta f(y)\right]$$

holds for all x, i and ξ. Since y is in $C_x^{0,t}$ if and only if x is in $C_y^{0,t}$, we obtain

$$|||P_t f||| \;\leq\; \sum_{x\in\mathbb{Z}^d} E\left[\sum_{y\in C_x^{0,t}} \Delta f(y)\right] \;=\; \sum_{y\in\mathbb{Z}^d} E\left[\sum_{x\in C_y^{0,t}} \Delta f(y)\right] \;=\; \sum_{y\in\mathbb{Z}^d} \Delta f(y)\, E\left[|C_y^{0,t}|\right].$$

The assertion follows, because by translation invariance, $E[|C_y^{0,t}|]$ does not depend on y. ■

We are now ready to identify the generator of the particle system.

Theorem 5.13 (Domain of the generator). The generator $(L, \mathrm{Dom}(L))$ of the semigroup $(P_t)_{t\geq 0}$ on the Banach space $C(T^{\mathbb{Z}^d})$ extends the operator $(\mathcal{L}, \mathcal{A})$ defined by (5.18). Furthermore, both $\mathcal{A}$ and C are cores for the generator.

Proof. By Lemma 5.12, we can control the graph norm of $\mathcal{L}$ by the triple norm. More precisely, for any $f \in \mathcal{A}$,

$$\|f\|_{\mathcal{L}} \;=\; \|f\|_{\sup} + \|\mathcal{L}f\|_{\sup} \;\leq\; \|f\|_{\sup} + \lambda\,|T|\;|||\,f\,|||\,. \tag{5.19}$$

This bound implies that the cylinder functions are dense in $\mathcal{A}$ w.r.t. the graph norm. To verify this, we approximate an arbitrary function $f \in \mathcal{A}$ by the same sequence $(f_n)_{n\in\mathbb{N}}$ of cylinder functions as in the proof of Lemma 5.10. We already know that $f_n \to f$ uniformly. Hence for every $x \in \mathbb{Z}^d$,

$$\Delta(f - f_n)(x) \;\leq\; \sup_{\xi,i}\left(|(f-f_n)(\xi^{x,i})| + |(f-f_n)(\xi)|\right) \;\leq\; 2\,\|f - f_n\|_{\sup} \;\to\; 0$$

as $n \to \infty$. Moreover, $\Delta f_n(x) \leq \Delta f(x)$, and thus $\Delta(f - f_n)(x) \leq 2\Delta f(x)$. For $f \in \mathcal{A}$, $\Delta f(x)$ is summable. Therefore, we can apply dominated convergence to conclude that

$$|||f - f_n||| \;=\; \sum_x \Delta(f - f_n)(x) \;\to\; 0 \qquad \text{as } n \to \infty.$$

Thus the space C of cylinder functions is dense in $\mathcal{A}$ w.r.t. the graph norm. By Theorem 5.13, the generator $(L, \mathrm{Dom}(L))$ extends the operator $(\mathcal{L}, C)$, and thus it also extends $(\mathcal{L}, \mathcal{A})$.

Furthermore, by the second assertion in Lemma 5.12, $P_t(\mathcal{A}) \subseteq \mathcal{A}$ holds for $t \in [0, \delta]$, and hence for arbitrary $t \geq 0$ by the semigroup property. Therefore, by Theorem 4.15, $\mathcal{A}$ is a core for the generator. Finally, since C is dense in $\mathcal{A}$ w.r.t. the graph norm, C is a core as well. ■

5.4. Attractive particle systems and phase transitions

In this section, we consider attractive interacting particle systems over $\mathbb{Z}^d$. These particle systems preserve stochastic order, and therefore comparison methods can be applied to study ergodic properties. As a consequence, it will turn out that ergodicity is closely related to the absence of phase transitions. An additional reference for this section is the monograph [35] by Liggett.

Attractive particle systems

We consider an interacting particle system over $\mathbb{Z}^d$ with state space $T = \{0,1\}$. Here $\eta(x) = 1$ and $\eta(x) = 0$ stand for the presence or absence of a particle at position x. Then a *partial order* on the configuration space $S = \{0,1\}^{\mathbb{Z}^d}$ is defined by

$$\eta \leq \widetilde{\eta} \qquad :\Leftrightarrow \qquad \eta(x) \leq \widetilde{\eta}(x) \text{ for all } x \in \mathbb{Z}^d.$$

Correspondingly, a function $f\colon S \to \mathbb{R}$ is called **increasing** if and only if

$$f(\eta) \leq f(\widetilde{\eta}) \text{ whenever } \eta \leq \widetilde{\eta}.$$

Definition 5.14 (Stochastic dominance). For probability measures $\mu, \nu \in \mathcal{P}(S)$ we set

$$\mu \preccurlyeq \nu \qquad :\Leftrightarrow \qquad \int f\, d\mu \leq \int f\, d\nu \text{ for every increasing bounded function } f\colon S \to \mathbb{R}.$$

Exercise (Stochastic dominance I). Let μ and ν be probability measures on $\mathbb{R}$. Prove that the following statements are equivalent.

(i) $\mu \preccurlyeq \nu$, i.e., $\int f\, d\mu \leq \int f\, d\nu$ for every increasing function $f \in \mathcal{F}_b(\mathbb{R})$.

(ii) $F_\mu(c) = \mu((-\infty, c]) \geq F_\nu(c)$ for all $c \in \mathbb{R}$.

(iii) There exists a coupling realized by random variables X and Y with distributions μ and ν, respectively, such that $X \leq Y$ almost surely.

Exercise (Stochastic dominance II). Let μ and ν be probability measures on the configuration space $S = \{0,1\}^{\mathbb{Z}^d}$, endowed with the product topology.

a) Prove that $\mu = \nu$ if and only if

$$\int f\, d\mu = \int f\, d\nu \quad \text{for any increasing, bounded and continuous function } f : S \to \mathbb{R}.$$

b) Conclude that if $\mu \preccurlyeq \nu$ and $\nu \preccurlyeq \mu$ then $\mu = \nu$.

Now consider again the stochastic dynamics constructed in the last section with transition rates $c_0(x,\eta)$ and $c_1(x,\eta)$ satisfying Assumptions (A1) and (A2).

Definition 5.15 (Attractive particle system). The Markov process $\left((\eta_t^\xi)_{t\geq 0}, P\right)$ is called **attractive** if and only if for all $x \in \mathbb{Z}^d$,

$$\eta \leq \widetilde{\eta},\ \eta(x) = \widetilde{\eta}(x) \quad \Rightarrow \quad c_1(x,\eta) \leq c_1(x,\widetilde{\eta}) \text{ and } c_0(x,\eta) \geq c_0(x,\widetilde{\eta}). \tag{5.20}$$

Example. The contact process, the voter model, as well as the Metropolis and the heat-bath dynamics for the (ferromagnetic) Ising model are all attractive particle systems.

Theorem 5.16 (Attractive particle systems preserve stochastic order). If the dynamics is attractive then the following statements hold.

(i) If $\xi \le \widetilde{\xi}$ then almost surely, $\eta_t^{\xi} \le \eta_t^{\widetilde{\xi}}$ for all $t \ge 0$.

(ii) If $f : S \to \mathbb{R}_+$ is an increasing function then $p_t f$ is increasing for every $t \ge 0$.

(iii) *Monotonicity of p_t:* If $\mu \preccurlyeq \nu$ then $\mu p_t \preccurlyeq \nu p_t$ for all $t \ge 0$.

Proof. (i) It can be easily checked that attractiveness implies that for the dynamics constructed in the last section, every transition at a fixed position $x \in \mathbb{Z}^d$ is order preserving. If, for example, before a possible transition at time $T_n^{x,1}$, $\eta \le \widetilde{\eta}$ and $\eta(x) = \widetilde{\eta}(x) = 0$, then after the transition, $\eta(x) = 1$ if $U_n^{x,1} \le c_1(x,\eta)/\lambda$, but in this case also $\widetilde{\eta}(x) = 1$ since $c_1(x,\eta) \le c_1(x,\widetilde{\eta})$ by attractiveness. The other cases are checked similarly.

The claim now follows from the componentwise construction of the dynamics on time intervals of length $\le \delta$.

(ii) Suppose that f is increasing and $\xi \le \widetilde{\xi}$. Then by (i),

$$(p_t f)(\xi) = E\left[f(\eta_t^{\xi})\right] \le E\left[f(\eta_t^{\widetilde{\xi}})\right] = (p_t f)(\widetilde{\xi}).$$

(iii) Suppose that $\mu \preccurlyeq \nu$. If f is bounded and increasing, then $p_t f$ is increasing as well. Therefore, by Fubini's lemma,

$$\int f \, d(\mu p_t) = \int p_t f \, d\mu \le \int p_t f \, d\nu = \int f \, d(\nu p_t).$$

This implies $\mu p_t \preccurlyeq \nu p_t$. ■

The minimal respectively maximal probability measure on S w.r.t. stochastic order are the Dirac measures δ_0 and δ_1 where 0 and 1 denote the corresponding constant configurations in S.

Theorem 5.17 (Extremal invariant probability measures). For an attractive particle system on $\{0,1\}^{\mathbb{Z}^d}$, the following assertions hold.

(i) The functions $t \mapsto \delta_0 p_t$ and $t \mapsto \delta_1 p_t$ are increasing respectively decreasing w.r.t. $\preccurlyeq$.

(ii) The limits $\underline{\mu} := \lim_{t\to\infty} \delta_0 p_t$ and $\bar{\mu} := \lim_{t\to\infty} \delta_1 p_t$ exist w.r.t. weak convergence in $\mathcal{P}(S)$.

(iii) Both $\underline{\mu}$ and $\bar{\mu}$ are invariant probability measures for $(p_t)_{t\ge 0}$.

(iv) Every invariant probability measure π satisfies

$$\underline{\mu} \preccurlyeq \pi \preccurlyeq \bar{\mu}.$$

Proof. (i) Let $0 \le s \le t$. Since $\delta_0 \preccurlyeq \delta_0 p_{t-s}$, the monotonicity of p_s implies

$$\delta_0 p_s \preccurlyeq \delta_0 p_{t-s} p_s = \delta_0 p_t.$$

(ii) This follows by monotonicity and compactness: Since $S = \{0,1\}^{\mathbb{Z}^d}$ is compact w.r.t. the product topology, $\mathcal{P}(S)$ is compact w.r.t. weak convergence. Thus it suffices to show that any two subsequential limits μ_1 and μ_2 of $\delta_0 p_t$ coincide. Suppose that f is an increasing continuous function. Then by (i), $\int f\, d(\delta_0 p_t)$ is increasing in t, and hence

$$\int f\, d\mu_1 \;=\; \lim_{t\uparrow\infty} \int f\, d(\delta_0 p_t) \;=\; \int f\, d\mu_2\,.$$

By the exercise above, this implies $\mu_1 = \mu_2$.

(iii) Let $t \geq 0$. By Corollary 5.9, p_t is Feller. Therefore, for all $f \in C_b(S)$,

$$\int f\, d(\underline{\mu} p_t) = \int p_t f\, d\underline{\mu} = \lim_{s\to\infty} \int p_t f\, d(\delta_0 p_s) = \lim_{s\to\infty} \int f\, d(\delta_0 p_s p_t) = \lim_{s\to\infty} \int f\, d(\delta_0 p_s) = \int f\, d\underline{\mu},$$

and thus $\underline{\mu} = \underline{\mu} p_t$. Similarly, $\bar{\mu} = \bar{\mu} p_t$.

(iv) Since π is invariant,

$$\delta_0 p_t \preccurlyeq \pi p_t = \pi \preccurlyeq \delta_1 p_t$$

for all $t \geq 0$. Considering the limit as $t \to \infty$, we obtain $\underline{\mu} \preccurlyeq \pi \preccurlyeq \bar{\mu}$. ■

Corollary 5.18 (Invariant probability measures and ergodicity for attractive IPS). For an attractive particle system, the following statements are equivalent:

(i) There is a unique invariant probability measure.

(ii) $\underline{\mu} = \bar{\mu}$.

(iii) There exists a probability measure μ on S such that $\nu p_t \to \mu$ for any $\nu \in \mathcal{P}(S)$.

Proof. By Theorem 5.17, (i) and (ii) are equivalent.
Moreover, for every $\nu \in \mathcal{P}(S)$, we have $\delta_0 \preccurlyeq \nu \preccurlyeq \delta_1$, and hence

$$\delta_0 p_t \preccurlyeq \nu p_t \preccurlyeq \delta_1 p_t \qquad \text{for all } t \geq 0. \tag{5.21}$$

If (ii) holds then $\delta_0 p_t$ and $\delta_1 p_t$ both converge weakly to $\underline{\mu}$. Using (5.21) and the exercise above, one can now conclude that every subsequential limit of νp_t coincides with $\underline{\mu}$. By compactness of $\mathcal{P}(S)$, this implies that νp_t converges weakly to $\underline{\mu}$. This proves "(ii)⇒(iii)", and the converse implication is obviously true as well. ■

Contact process on $\mathbb{Z}^{\mathbf{d}}$

For the contact process, $c_0(x,\eta) = \delta$ and $c_1(x,\eta) = b \cdot N_1(x,\eta)$ where the birth rate b and the death rate δ are positive constants. Since the zero configuration is an absorbing state, $\underline{\mu} = \delta_0$ is the minimal invariant probability measure. The question now is if there is another (non-trivial) invariant probability measure, i.e., if $\bar{\mu} \neq \underline{\mu}$.

Theorem 5.19. If $2db < \delta$ then δ_0 is the only invariant probability measure, and ergodicity holds.

Proof. By the forward equation and translation invariance,

$$\begin{aligned}\frac{d}{dt}P\left[\eta_t^1(x)=1\right] &= -\delta P\left[\eta_t^1(x)=1\right] + \sum_{y\,:\,|x-y|=1} b\cdot P\left[\eta_t^1(x)=0,\ \eta_t^1(y)=1\right] \\ &\le (-\delta+2db)\cdot P\left[\eta_t^1(x)=1\right]\end{aligned}$$

for all $x\in\mathbb{Z}^d$. Hence if $2db<\delta$ then

$$\bar{\mu}(\{\eta\ :\ \eta(x)=1\}) \;=\; \lim_{t\to\infty}(\delta_1 p_t)(\{\eta\ :\ \eta(x)=1\}) \;=\; \lim_{t\to\infty} P\left[\eta_t^1(x)=1\right] \;=\; 0$$

for all $x\in\mathbb{Z}^d$, and thus $\bar{\mu}=\delta_0$. ■

Conversely, one can show that for b sufficiently small (or δ sufficiently large), there is a nontrivial invariant probability measure. The proof is more involved, see Liggett [35]. Thus a phase transition from ergodicity to non-ergodicity occurs as b increases.

Ising model on $\mathbb{Z}^{\mathbf{d}}$

We consider either the heat bath or the Metropolis dynamics for the Ising model on $S=\{-1,+1\}^{\mathbb{Z}^d}$ with inverse temperature $\beta\in[0,\infty)$.

Finite volume

Let $A\subseteq\mathbb{Z}^d$ be finite. Then the restricted configuration spaces

$$S_{+,A}\;:=\;\{\eta\in S:\eta=+1\text{ on }A^c\}\qquad\text{and}\qquad S_{-,A}\;:=\;\{\eta\in S:\eta=-1\text{ on }A^c\}$$

corresponding to "+" and "-" boundary conditions outside A are finite. For initial configurations $\xi\in S_{+,A}$ or $\xi\in S_{-,A}$, we denote by $\eta_t^{\xi,A}$ the dynamics constructed as above, but taking into account only transitions in A. Then $\left(\eta_t^{\xi,A},P\right)$ is a continuous time Markov chain on $S_{+,A}$, $S_{-,A}$, respectively, with generator

$$(\mathcal{L}f)(\eta)\;=\;\sum_{\substack{x\in A\\ i\in\{-1,+1\}}} c_i(x,\eta)\cdot\left(f(\eta^{x,i})-f(\eta)\right).$$

Let

$$H(\eta)\;=\;\frac{1}{4}\sum_{\substack{x,y\in\mathbb{Z}^d\\ |x-y|=1}}(\eta(x)-\eta(y))^2$$

denote the *Ising Hamiltonian*. Note that for $\eta\in S_{+,A}$ or $\eta\in S_{-,A}$, only finitely many summands do not vanish, and so $H(\eta)$ is finite. The probability measures

$$\mu_\beta^{+,A}(\eta)\;=\;\frac{1}{Z_\beta^{+,A}}e^{-\beta H(\eta)},\quad \eta\in S_{+,A},\qquad\text{where}\qquad Z_\beta^{+,A}\;=\;\sum_{\eta\in S_{+,A}}e^{-\beta H(\eta)}$$

on $S_{+,A}$, and $\mu_\beta^{-,A}$ on $S_{-,A}$ defined correspondingly satisfy the detailed balance conditions

$$\begin{aligned}\mu_\beta^{+,A}(\xi)\mathcal{L}(\xi,\eta) &= \mu_\beta^{+,A}(\eta)\mathcal{L}(\eta,\xi)\qquad\text{for all }\xi,\eta\in S_{+,A},\\ \mu_\beta^{-,A}(\xi)\mathcal{L}(\xi,\eta) &= \mu_\beta^{-,A}(\eta)\mathcal{L}(\eta,\xi)\qquad\text{for all }\xi,\eta\in S_{-,A}.\end{aligned}$$

Since $S_{+,A}$ and $S_{-,A}$ are finite and irreducible this implies that $\mu_\beta^{+,A}$ and $\mu_\beta^{-,A}$ are the invariant probability measures of $(\eta_t^{\xi,A},P)$ for $\xi\in S_{+,A}$, $\xi\in S_{-,A}$, respectively. Thus in finite volume there are several processes corresponding to different boundary conditions (which affect the Hamiltonian) but each of them has a unique invariant probability measure. Conversely, in infinite volume there is only one process, but it may have several invariant probability measures.

Infinite volume

To identify the invariant probability measures for the process on $\mathbb{Z}^d$, we use an approximation by the dynamics in finite volume. For $n \in \mathbb{N}$ let

$$A_n \; := \; [-n,n]^d \cap \mathbb{Z}^d.$$

The sequences μ_β^{+,A_n} and μ_β^{-,A_n} ($n \in \mathbb{N}$), are decreasing respectively increasing w.r.t. stochastic dominance. Hence by compactness of $\{-1,+1\}^{\mathbb{Z}^d}$, there exist weak limits

$$\mu_\beta^+ \; := \; \lim_{n\uparrow\infty} \mu_\beta^{+,A_n} \quad \text{and} \quad \mu_\beta^- \; := \; \lim_{n\uparrow\infty} \mu_\beta^{-,A_n}.$$

For a finite set $A \subset \mathbb{Z}^d$ and $\xi \in S$, the *Boltzmann-Gibbs distribution with boundary condition ξ on $\mathbb{Z}^d \setminus A$* is defined by

$$\mu_\beta^{\xi,A}(\eta) \; := \; \frac{1}{Z_\beta^{\xi,A}} e^{-\beta H(\eta)}, \qquad \eta \; \in \; S_{\xi,A} \; := \; \{\eta \in S : \eta = \xi \text{ on } A^c\},$$

where $Z_\beta^{\xi,A}$ is the corresponding normalizing constant. For $\xi \equiv +1$ and $\xi \equiv -1$, we recover the measures $\mu_\beta^{+,A}$ and $\mu_\beta^{-,A}$ introduced above.

Definition 5.20 (Gibbs measure). A probability measure μ_β on S is called a *Gibbs measure for the Ising Hamiltonian on $\mathbb{Z}^d$ at inverse temperature $\beta > 0$* if and only if for every finite set $A \subset \mathbb{Z}^d$, a regular version of the conditional probability given $\xi(x)$ for all $x \in \mathbb{Z}^d \setminus A$ is given by

$$\mu_\beta \left(\cdot \;\middle|\; \eta = \xi \text{ on } \mathbb{Z}^d \setminus A \right) \; = \; \mu_\beta^{\xi,A}.$$

For $\beta = 0$, the components are independent, and hence the product measure $\bigotimes_{x\in\mathbb{Z}^d} \mathrm{Unif}\{-1,+1\}$ is the unique Gibbs measure. In general, one can show that μ_β^+ and μ_β^- are the extremal Gibbs measures for the Ising model w.r.t. stochastic dominance, see e.g. [41]. We say that a **phase transition** occurs if for some inverse temperature $\beta > 0$, there are several Gibbs measures. This is the case if and only if $\mu_\beta^+ \neq \mu_\beta^-$.

The following theorem shows that phase transitions occur for Ising models in dimension $d \geq 2$.

Theorem 5.21 (Phase transition for two dimensional Ising model; Peierl). For $d = 2$ there exists $\beta_c \in (0,\infty)$ such that for $\beta > \beta_c$,

$$\mu_\beta^+ \left(\{\eta \;:\; \eta(0) = -1\}\right) \; < \; \frac{1}{2} \; < \; \mu_\beta^- \left(\{\eta \;:\; \eta(0) = -1\}\right),$$

and thus $\mu_\beta^+ \neq \mu_\beta^-$.

The cluster swapping argument in the following proof goes back to Peierl.

Proof. Let $C_0(\eta)$ denote the connected component of 0 in $\{x \in \mathbb{Z}^d : \eta(x) = -1\}$, and set $C_0(\eta) = \varnothing$ if $\eta(0) = +1$. Let $A \subseteq \mathbb{Z}^d$ be finite and non-empty. For $\eta \in S$ with $C_0(\eta) = A$ let $\widetilde{\eta}$ denote the configuration obtained by reversing all spins in A. Since the energy of a configuration is twice the length of the boundary between the plus and minus spins,

$$H(\widetilde{\eta}) \; = \; H(\eta) - 2|\partial A|.$$

Therefore, for any $n \in \mathbb{N}$,

$$\mu_\beta^{+,n}(C_0 = A) \;=\; \sum_{\eta\,:\,C_0(\eta)=A} \mu_\beta^{+,n}(\eta) \;\le\; e^{-2\beta|\partial A|} \underbrace{\sum_{\eta\,:\,C_0(\eta)=A} \mu_\beta^{+,n}(\widetilde{\eta})}_{\le 1} \;\le\; e^{-2\beta|\partial A|}.$$

Thus we obtain

$$\begin{aligned} \mu_\beta^{+,n}(\{\eta \;:\; \eta(0) = -1\}) \;&=\; \sum_{\substack{A\subset\mathbb{Z}^d \\ A\neq\emptyset}} \mu_\beta^{+,n}(C_0 = A) \;\le\; \sum_{L=4}^{\infty} e^{-2\beta L} \left|\{A \subset \mathbb{Z}^d \;:\; 0 \in A,\, |\partial A| = L\}\right| \\ &\le\; \sum_{L=4}^{\infty} e^{-2\beta L} L^2 \cdot 4 \cdot 3^{L-1}. \end{aligned}$$

Here we have used that ∂A is a self-avoiding path in $\mathbb{Z}^2$ of length L, starting in $\left(-\frac{L}{2}, \frac{L}{2}\right)^2$, and the number of such paths is bounded by $L^2 \cdot 4 \cdot 3^{L-1}$.

As $n \to \infty$, we see that there exists $\beta_c \in (0, \infty)$ such that for all $\beta > \beta_c$,

$$\mu_\beta^{+}(\{\eta \;:\; \eta(0) = -1\}) \;<\; \frac{1}{2}.$$

Similarly, by symmetry,

$$\mu_\beta^{-}(\{\eta \;:\; \eta(0) = -1\}) \;=\; \mu_\beta^{+}(\{\eta \;:\; \eta(0) = +1\}) \;>\; \frac{1}{2}$$

for $\beta > \beta_c$. ■

We now want to link the existence of a phase transition to the non-ergodicity of the stochastic dynamics. We start with a preparatory lemma. For $\xi \in S$, we define $\xi_n \in S_{+,A_n}$ by

$$\xi_n(x) := \begin{cases} \xi(x) & \text{for } x \in A_n, \\ +1 & \text{for } x \in \mathbb{Z}^d \setminus A_n. \end{cases}$$

Lemma 5.22. *For all $x \in \mathbb{Z}^d$,*

$$\lim_{n\to\infty} P\left[\eta_t^{\xi}(x) \neq \eta_t^{\xi_n, A_n}(x) \quad \text{for some } \xi \in S \text{ and some } t \in [0,\delta]\right] \;=\; 0. \tag{5.22}$$

Proof. Let $C_x^{0,\delta}$ denote the component containing x in the random graph $\left(\mathbb{Z}^d, E_{0,\delta}\right)$. If $C_x^{0,\delta} \subseteq A_n$ then the modifications in the initial condition and the transition mechanism outside A_n do not affect the value at x before time δ. Hence the probability in (5.22) can be estimated by

$$P\left[C_x^{0,\delta} \cap A_n^c \neq \emptyset\right],$$

which goes to 0 as $n \to \infty$ by Lemma 5.7 above. ■

Let p_t denote the transition semigroup of the process on $\{-1,1\}^{\mathbb{Z}^d}$. Since the dynamics is attractive,

$$\bar{\mu}_\beta = \lim_{t\to\infty} \delta_{+1} p_t \quad \text{and} \quad \underline{\mu}_\beta = \lim_{t\to\infty} \delta_{-1} p_t$$

are extremal invariant probability measures w.r.t. stochastic dominance. The following theorem identifies $\bar{\mu}_\beta$ and $\underline{\mu}_\beta$ as the extremal Gibbs measures for the Ising Hamiltonian on $\mathbb{Z}^d$.

Theorem 5.23 (Phase transitions and ergodicity). The upper and lower invariant probability measures coincide with the extremal Gibbs measures, i.e.,

$$\bar{\mu}_\beta = \mu_\beta^+ \quad \text{and} \quad \underline{\mu}_\beta = \mu_\beta^-.$$

In particular, ergodicity holds if and only if there is a unique Gibbs measure (i.e., if and only if $\mu_\beta^+ = \mu_\beta^-$).

Proof. We show the following:

(i) $\bar{\mu}_\beta \preccurlyeq \mu_\beta^+$,

(ii) μ_β^+ is an invariant probability measure for $(p_t)_{t\geq 0}$.

This implies $\bar{\mu}_\beta = \mu_\beta^+$, since by (ii) and the corollary above, $\mu_\beta^+ \preccurlyeq \bar{\mu}_\beta$, and thus $\mu_\beta^+ = \bar{\mu}_\beta$ by (i) and the exercise above. $\mu_\beta^- = \underline{\mu}_\beta$ follows similarly. It remains to prove (i) and (ii).

(i) It can be shown by a similar argument as in the proof of Theorem 5.16 that the attractiveness of the dynamics implies that almost surely,

$$\eta_t^1 \;\leq\; \eta_t^{1,A_n} \qquad \text{for all } n \in \mathbb{N} \text{ and } t \geq 0. \tag{5.23}$$

As $t \to \infty$, $\eta_t^1 \xrightarrow{\mathcal{D}} \bar{\mu}_\beta$, and, by ergodicity of the dynamics in finite volume, $\eta_t^{1,A_n} \xrightarrow{\mathcal{D}} \mu_\beta^{+,A_n}$. Therefore, (5.23) implies that

$$\bar{\mu}_\beta \;\preccurlyeq\; \mu_\beta^{+,A_n} \qquad \text{for all } n \in \mathbb{N}.$$

The assertion follows as $n \to \infty$.

(ii) By the semigroup property of $(p_t)_{t\geq 0}$, it is enough to show

$$\mu_\beta^+ p_t \;=\; \mu_\beta^+ \quad \text{for all } t < \delta. \tag{5.24}$$

Hence fix $t \in [0, \delta)$ and consider the transition semigroup

$$\left(p_t^n f\right)(\xi) \;:=\; E\left[f(\eta_t^{\xi_n, A_n})\right]$$

on S_{ξ_n, A_n}. By detailed balance, we know that for all $n \in \mathbb{N}$,

$$\mu_\beta^{+,n} p_t^n \;=\; \mu_\beta^{+,n} \tag{5.25}$$

We want to pass to the limit as $n \to \infty$. Let $f(\eta) = \phi\left(\eta(x_1), \ldots, \eta(x_k)\right)$ be a cylinder function on S. Then by (5.25),

$$\int p_t f \, d\mu_\beta^{+,n} = \int p_t^n f \, d\mu_\beta^{+,n} + \int \left(p_t^n f - p_t f\right) \, d\mu_\beta^{+,n} = \int f \, d\mu_\beta^{+,n} + \int \left(p_t^n f - p_t f\right) \, d\mu_\beta^{+,n}. \tag{5.26}$$

But for $t \leq \delta$ and $\xi \in S$,

$$\left|(p_t^n f)(\xi) - (p_t f)(\xi)\right| \leq E\left[\left|f\left(\eta_t^{\xi_n, A_n}\right) - f\left(\eta_t^{\xi}\right)\right|\right] \leq 2 \sup|f| \cdot P\left[\exists\, i \leq k : \eta_t^{\xi_n, A_n}(x_i) \neq \eta_t^{\xi}(x_i)\right],$$

and by Lemma 5.22, this expression converges to 0 uniformly in ξ as $n \to \infty$. Since $\mu_\beta^{+,n} \xrightarrow{\mathrm{w}} \mu_\beta^+$, and f and $p_t f$ are continuous by the Feller property, taking the limit in (5.26) as $n \to \infty$ yields

$$\int f \, d\left(\mu_\beta^+ p_t\right) \;=\; \int p_t f \, d\mu_\beta^+ \;=\; \int f \, d\mu_\beta^+$$

for all cylinder functions f, which implies (5.24). ■

6. Limits of martingale problems

In this chapter, we give constructions for solutions of several important classes of martingale problems. In Section 6.1 the necessary tools for weak convergence of stochastic processes are developed and applied to obtain Brownian motion as a universal scaling limit of random walks. In Section 6.2, the techniques are generalized and applied to prove an existence result for diffusion processes in $\mathbb{R}^d$.

6.1. From Random Walks to Brownian motion

Limits of martingale problems occur frequently in theoretical and applied probability. Examples include the approximation of Brownian motion by random walks and, more generally, the convergence of Markov chains to diffusion limits, the approximation of Feller processes by jump processes, the approximation of solutions of stochastic differential equations by solutions to more elementary SDEs or by processes in discrete time, the construction of processes on infinite-dimensional or singular state spaces as limits of processes on finite-dimensional or more regular state spaces etc. A general and frequently applied approach to this type of problems can be summarized in the following scheme:

1. Write down generators $\mathcal{L}_n$ of the approximating processes and identify a limit generator $\mathcal{L}$ (on an appropriate collection of test functions) such that $\mathcal{L}_n \to \mathcal{L}$ in an appropriate sense.

2. Prove tightness for the sequence (P_n) of laws of the solutions to the approximating martingale problems. Then extract a weakly convergent subsequence.

3. Prove that the limit solves the martingale problem for the limit generator.

4. Identify the limit process.

The technically most demanding steps are usually 2 and 4. Notice that Step 4 involves a uniqueness statement. Since uniqueness for solutions of martingale problems is often difficult to establish (and may not hold!), the last step can not always be carried out. In this case, there may be different subsequential limits of the sequence (P_n).

In this section, we introduce the necessary tools from weak convergence that are required to make the program outlined above rigorous. We then apply the techniques in a simple but important case: The approximation of Brownian motion by random walks.

Weak convergence of stochastic processes

An excellent reference on this subject is the book by Billingsley [3]. Let S be a polish space. We fix a metric d on S such that (S, d) is complete and separable. We consider the laws of stochastic processes either on the space $C = C([0,\infty), S)$ of continuous functions $x : [0,\infty) \to S$ or on the space $\mathcal{D} = \mathcal{D}([0,\infty), S)$ consisting of all càdlàg functions $x : [0,\infty) \to S$. The space C is again a polish space w.r.t. the topology of uniform convergence on compact time intervals:

$$x_n \xrightarrow{C} x \quad :\Leftrightarrow \quad \forall T \in \mathbb{R}_+ : x_n \to x \;\text{ uniformly on } [0,T].$$

On càdlàg functions, uniform convergence is too restrictive for our purposes. For example, the indicator functions $1_{[0,1+n^{-1})}$ do not converge uniformly to $1_{[0,1)}$ as $n \to \infty$. Instead, we endow the space $\mathcal{D}$ with the Skorokhod topology:

Definition 6.1 (Skorokhod topology). A sequence of functions $x_n \in \mathcal{D}$ is said to **converge to a limit $\mathbf{x} \in \mathcal{D}$ in the Skorokhod topology** if and only if for any $T \in \mathbb{R}_+$ there exist continuous and strictly increasing maps $\lambda_n : [0,T] \to [0,T]$ ($n \in \mathbb{N}$) such that

$$x_n(\lambda_n(t)) \to x(t) \quad \text{and} \quad \lambda_n(t) \to t \quad \text{uniformly on } [0,T].$$

It can be shown that the Skorokhod space $\mathcal{D}$ is again a polish space, cf. [3]. Furthermore, the Borel σ-algebras on both C and $\mathcal{D}$ are generated by the projections $X_t(x) = x(t)$, $t \in \mathbb{R}_+$.

Let $(P_n)_{n\in\mathbb{N}}$ be a sequence of probability measures (laws of stochastic processes) on $C, \mathcal{D}$ respectively. By Prokhorov's Theorem, every subsequence of (P_n) has a weakly convergent subsequence provided (P_n) is tight. Here tightness means that for every $\varepsilon > 0$, there exists a relatively compact subset $K \subseteq C$, $K \subseteq \mathcal{D}$ respectively, such that

$$\sup_{n\in\mathbb{N}} P_n[K^c] \;\le\; \varepsilon.$$

To verify tightness we need a characterization of the relatively compact subsets of the function spaces C and $\mathcal{D}$. In the case of C such a characterization is the content of the classical Arzelà-Ascoli Theorem. This result has been extended to the space $\mathcal{D}$ by Skorokhod. To state both results we define the modulus of continuity of a function $x \in C$ on the interval $[0,T]$ by

$$\omega_{\delta,T}(x) \;=\; \sup_{\substack{s,t\in[0,T]\\ |s-t|\le\delta}} d(x(s),x(t)).$$

For $x \in \mathcal{D}$ we define

$$\omega'_{\delta,T}(x) = \inf_{\substack{0=t_0<t_1<\cdots<t_{n-1}<T\le t_n\\ |t_i-t_{i-1}|>\delta}} \max_i \sup_{s,t\in[t_{i-1},t_i)} d(x(s),x(t)).$$

For any $x \in C$ and $T \in \mathbb{R}_+$, the modulus of continuity $\omega_{\delta,T}(x)$ converges to 0 as $\delta \downarrow 0$. For a discontinuous function $x \in \mathcal{D}$, $\omega_{\delta,T}(x)$ does not converge to 0. However, the modified quantity $\omega'_{\delta,T}(x)$ again converges to 0, since the partition in the infimum can be chosen in such a way that jumps of size greater than some constant ε occur only at partition points and are not taken into account in the inner maximum.

Exercise (Modulus of continuity and Skorokhod modulus). Let $x \in \mathcal{D}$.

1) Show that $\lim_{\delta\downarrow 0} \omega_{\delta,T}(x) = 0$ for any $T \in \mathbb{R}_+$ if and only if x is continuous.

2) Prove that $\lim_{\delta\downarrow 0} \omega'_{\delta,T}(x) = 0$ for any $T \in \mathbb{R}_+$.

Theorem 6.2 (Arzelà-Ascoli, Skorokhod). 1) A subset $K \subseteq C$ is relatively compact if and only if

(i) $\{x(0) : x \in K\}$ is relatively compact in S, and

(ii) $\sup_{x\in K} \omega_{\delta,T}(x) \to 0$ as $\delta \downarrow 0$ for any $T > 0$.

2) A subset $K \subseteq \mathcal{D}\mathcal{D}\mathcal{D}$ is relatively compact if and only if

(i) $\{x(t) : x \in K\}$ is relatively compact for any $t \in \mathbb{Q}_+$, and

(ii) $\sup_{x\in K} \omega'_{\delta,T}(x) \to 0$ as $\delta \downarrow 0$ for any $T > 0$.

The proofs can be found in Billingsley [3] or Ethier/Kurtz [19]. By combining Theorem 6.2 with Prokhorov's Theorem, one obtains:

Corollary 6.3 (Tightness of probability measures on function spaces).

1) A subset $\{P_n : n \in \mathbb{N}\}$ of $\mathcal{P}(C)$ is relatively compact w.r.t. weak convergence if and only if
   (i) For any $\varepsilon > 0$, there exists a compact set $K \subseteq S$ such that
   $$\sup_{n\in\mathbb{N}} P_n[X_0 \notin K] \le \varepsilon, \quad \text{and}$$
   (ii) For any $T \in \mathbb{R}_+$,
   $$\sup_{n\in\mathbb{N}} P_n[\omega_{\delta,T} > \varepsilon] \to 0 \quad \text{as } \delta \downarrow 0.$$
2) A subset $\{P_n : n \in \mathbb{N}\}$ of $\mathcal{P}(\mathcal{D})$ is relatively compact w.r.t. weak convergence if and only if
   (i) For any $\varepsilon > 0$ and $t \in \mathbb{R}_+$ there exists a compact set $K \subseteq S$ such that
   $$\sup_{n\in\mathbb{N}} P_n[X_t \notin K] \le \varepsilon, \quad \text{and}$$
   (ii) For any $\varepsilon > 0$ and $T \in \mathbb{R}_+$,
   $$\sup_{n\in\mathbb{N}} P_n[\omega'_{\delta,T} > \varepsilon] \to 0 \quad \text{as } \delta \downarrow 0.$$

In the sequel we restrict ourselves to convergence of stochastic processes with continuous paths. We point out, however, that many of the arguments can be carried out (with additional difficulties) for processes with jumps if the space of continuous functions is replaced by the Skorokhod space. A detailed study of convergence of martingale problems for discontinuous Markov processes can be found in Ethier/Kurtz [19].

To apply the tightness criterion we need upper bounds for the probabilities $P_n[\omega_{\delta,T} > \varepsilon]$. To this end we observe that $\omega_{\delta,T} \le \varepsilon$ if

$$\sup_{t\in[0,\delta]} d(X_{k\delta+t}, X_{k\delta}) \le \frac{\varepsilon}{3} \quad \text{for any } k \in \mathbb{Z}_+ \text{ such that } k\delta < T.$$

Therefore, we can estimate

$$P_n[\omega_{\delta,T} > \varepsilon] \;\le\; \sum_{k=0}^{\lfloor T/\delta \rfloor} P_n\left[\sup_{t\le\delta} d(X_{k\delta+t}, X_{k\delta}) > \varepsilon/3\right]. \tag{6.1}$$

Furthermore, on $\mathbb{R}^n$ we can bound the distances $d(X_{k\delta+t}, X_{k\delta})$ by the sum of the differences $|X^i_{k\delta+t} - X^i_{k\delta}|$ of the components X^i, $i = 1, \dots, n$. The suprema can then be controlled by applying a semimartingale decomposition and the maximal inequality to the component processes.

Donsker's invariance principle

As a first application of the tightness criterion we prove Donsker's invariance principle stating that rescaled random walks with square integrable increments converge in law to a Brownian motion. In particular, this is a way (although not the easiest one) to prove that Brownian motion exists. Let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. square-integrable random variables on a probability space $(\Omega, \mathfrak{A}, P)$ with $E[Y_i] = 0$ and $\mathrm{Var}[Y_i] = 1$, and consider the random walk

$$S_n = \sum_{i=1}^{n} Y_i \qquad (n \in \mathbb{N}).$$

We rescale diffusively, i.e., by a factor n in time and a factor $\sqrt{n}$ in space, and define

$$X_t^{(n)} := \frac{1}{\sqrt{n}} S_{nt} \qquad \text{for } t \in \mathbb{R}_+ \text{ such that } nt \in \mathbb{Z}.$$

In between the partition points $t = k/n,\ k \in \mathbb{Z}_+$, the process $(X_t^{(n)})$ is defined by linear interpolation so that $X^{(n)}$ has continuous paths.

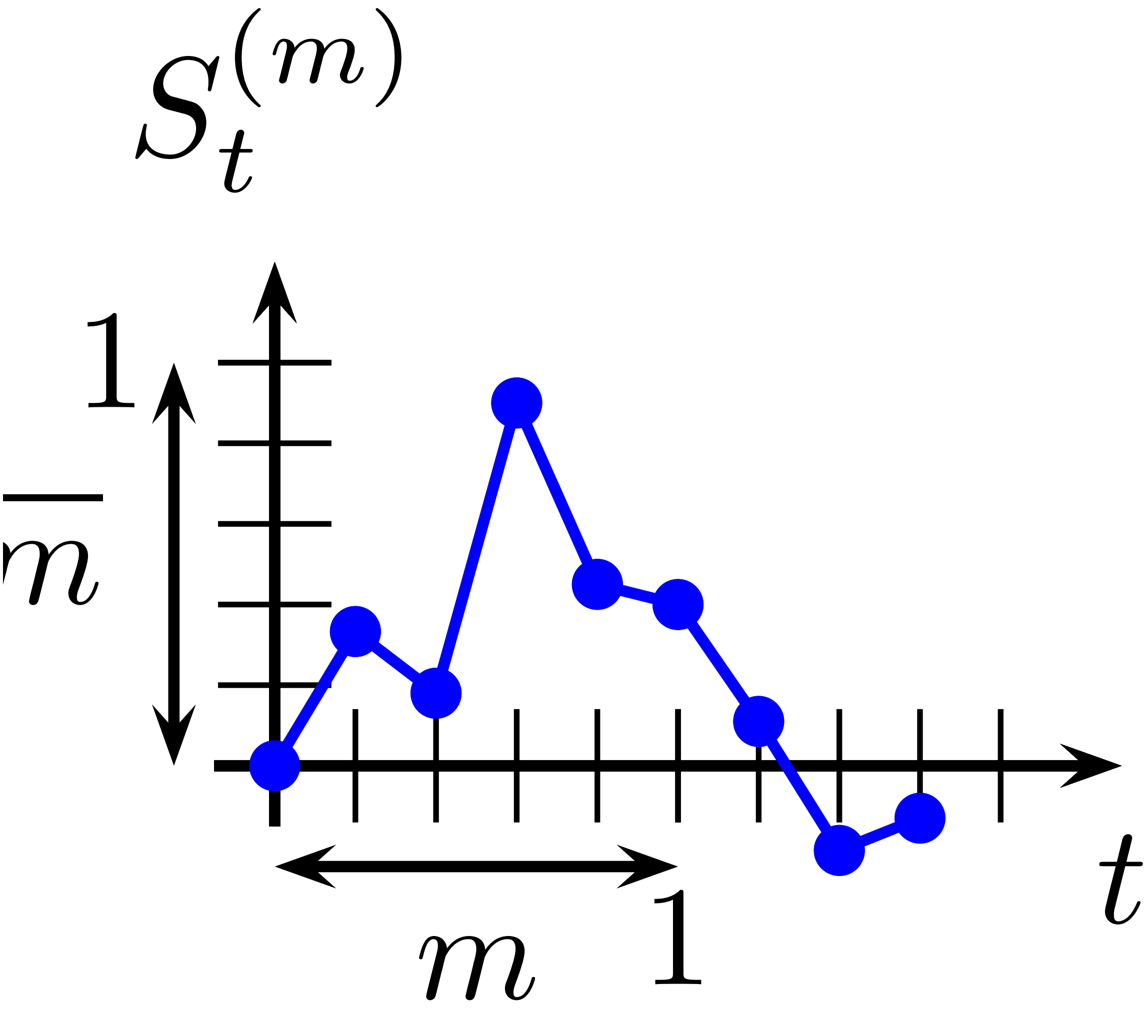


Figure 6.1.: Rescaling of a Random Walk.

The diffusive rescaling guarantees that the variances of $X_t^{(n)}$ converge to a finite limit as $n \to \infty$ for any fixed $t \in \mathbb{R}_+$. Indeed, the central limit theorem even shows that for any $k \in \mathbb{N}$ and $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$,

$$(X_{t_1}^{(n)} - X_{t_0}^{(n)}, X_{t_2}^{(n)} - X_{t_1}^{(n)}, \ldots, X_{t_k}^{(n)} - X_{t_{k-1}}^{(n)}) \xrightarrow{\mathcal{D}} \bigotimes_{i=1}^{k} N(0, t_i - t_{i-1}). \tag{6.2}$$

This implies that the marginals of the processes $X^{(n)}$ converge weakly to the marginals of a Brownian motion. Using tightness of the laws of the rescaled random walks on C, we can prove that not only the marginals but the whole processes converge in distribution to a Brownian motion:

Theorem 6.4 (Invariance principle, functional central limit theorem). Let P_n denote the law of the rescaled random walk $X^{(n)}$ on $C = C([0,\infty),\mathbb{R})$. Then $(P_n)_{n\in\mathbb{N}}$ converges weakly to Wiener measure, i.e., to the law of a Brownian motion starting at 0.

Proof. Since by (6.2), the marginals converge to the right limit, it suffices to prove tightness of the sequence $(P_n)_{n\in\mathbb{N}}$ of probability measures on C. Then by Prokhorov's Theorem, every subsequence has a weakly convergent subsequence, and all subsequential limits are equal to Wiener measure because the marginals coincide. Thus (P_n) also converges weakly to Wiener measure.
For proving tightness note that by (6.1) and time-homogeneity,

$$\begin{aligned} P_n[\omega_{\delta,T} > \varepsilon] &\leq \left(\left\lfloor \frac{T}{\delta} \right\rfloor + 1\right) \cdot P\left[\sup_{t\leq\delta}\left|X_t^{(n)} - X_0^{(n)}\right| \geq \frac{\varepsilon}{3}\right] \\ &\leq \left(\left\lfloor \frac{T}{\delta} \right\rfloor + 1\right) \cdot P\left[\max_{k\leq\lceil n\delta\rceil} |S_k| \geq \frac{\varepsilon}{3}\sqrt{n}\right] \end{aligned}$$

for any $\varepsilon,\delta > 0, T \in \mathbb{R}_+$ and $n \in \mathbb{N}$. By Corollary 6.3, tightness holds if the probability on the right hand side is of order $o(\delta)$ uniformly in n, i.e., if

$$\limsup_{m\to\infty} P\left[\max_{k\leq m} |S_k| \geq \frac{\varepsilon}{3}\frac{\sqrt{m}}{\sqrt{\delta}}\right] = o(\delta). \tag{6.3}$$

For the simple random walk, this follows from the reflection principle and the central limit theorem as

$$P\left[\max_{k\leq m} S_k \geq \frac{\varepsilon}{3}\frac{\sqrt{m}}{\sqrt{\delta}}\right] \leq P\left[|S_m| \geq \frac{\varepsilon}{3}\frac{\sqrt{m}}{\sqrt{\delta}}\right] \overset{m\uparrow\infty}{\longrightarrow} N(0,1)\left[|x| \geq \frac{\varepsilon}{3\sqrt{\delta}}\right],$$

cf. e.g. [15]. For general random walks one can show with some additional arguments that (6.3) also holds, see e.g. Billingsley [3]. ■

In the proof of Donsker's Theorem, convergence of the marginals was a direct consequence of the central limit theorem. In more general situations, other methods are required to identify the limit process. Therefore, we observe that instead of the central limit theorem, we could have also used the martingale problem to identify the limit as a Brownian motion. Indeed, the rescaled random walk $\left(X_{k/n}^{(n)}\right)_{k\in\mathbb{Z}_+}$ is a Markov chain (in discrete time) with generator

$$(\mathcal{L}^{(n)}f)(x) = \int \left(f\left(x + \frac{z}{\sqrt{n}}\right) - f(x)\right)\, \nu(dz)$$

where ν is the distribution of the increments $Y_i = S_i - S_{i-1}$. It follows that w.r.t. P_n, the process

$$f(X_t) - \sum_{i=0}^{nt-1} (n\mathcal{L}^{(n)}f)(X_{i/n}) \cdot \frac{1}{n}, \quad t = \frac{k}{n} \text{ with } k \in \mathbb{Z}_+,$$

is a martingale for any function $f \in C_b^\infty(\mathbb{R})$. As $n \to \infty$,

$$\begin{aligned} f\left(x + \frac{z}{\sqrt{n}}\right) - f(x) &= f'(x) \cdot \int \frac{z}{\sqrt{n}}\, \nu(dz) + \frac{1}{2}f''(x) \int \frac{z^2}{n}\, \nu(dz) + \mathrm{o}(n^{-1}) \\ &= \frac{1}{2n} f''(x) + \mathrm{o}(n^{-1}) \end{aligned}$$

by Taylor, and

$$(n\mathcal{L}^{(n)}f)(x) \to \frac{1}{2}f''(x) \quad \text{uniformly}.$$

Therefore, one can conclude that the process

$$f(X_t) - \int_0^t \frac{1}{2}f''(X_s)ds$$

is a martingale under P_∞ for any weak limit point of the sequence (P_n). Uniqueness of the martingale problem then implies that P_∞ is the law of a Brownian motion.

> **Exercise (Martingale problem proof of Donsker's Theorem).** Carry out carefully the arguments sketched above and give an alternative proof of Donsker's Theorem that avoids application of the central limit theorem.

6.2. Limits of general martingale problems

A broad class of diffusion processes on $\mathbb{R}^n$ can be constructed by stochastic analysis methods. Suppose that $((B_t)_{t\geq 0}, P)$ is a Brownian motion with values in $\mathbb{R}^n$ for some $n \in \mathbb{N}$, and $((X_t)_{t<\zeta}, P)$ is a solution to an Itô stochastic differential equation of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x_0, \tag{6.4}$$

up to the explosion time $\zeta = \sup T_k$ where T_k is the first exit time of (X_t) from the unit ball of radius k, cf. [16]. We assume that the coefficients are continuous functions $b : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n\cdot d}$. Then $((X_t)_{t<\zeta}, P)$ solves the **local martingale problem** for the operator

$$\mathcal{L}_t = b(t,x)\cdot\nabla_x + \frac{1}{2}\sum_{i,j=1}^n a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j}, \quad a := \sigma\sigma^T,$$

in the following sense: For any function $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$,

$$M_t^f = f(t, X_t) - \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{L}_s f\right)(s, X_s)\, ds$$

is a local martingale up to ζ. Indeed, by the Itô-Doeblin formula, M_t^f is a stochastic integral w.r.t. Brownian motion:

$$M_t^f = f(0, X_0) + \int_0^t \left(\sigma^T \nabla f\right)(s, X_s)\cdot dB_s.$$

If the explosion time ζ is almost surely infinite then M^f is even a **global martingale** provided the function $\sigma^T \nabla f$ is bounded.

In general, a solution of (6.4) is not necessarily a Markov process. If, however, the coefficients are Lipschitz continuous then by Itô's existence and uniqueness result there is a unique strong solution for any given initial value, and it can be shown that the strong Markov property holds, cf. [17].

By extending the methods developed in 6.1, we are now going to sketch another construction of diffusion processes in $\mathbb{R}^n$ that avoids stochastic analysis techniques to some extent. The raeson for our interest in this method is that the basic approach is very generally applicable – not only for diffusions in $\mathbb{R}^n$.

Regularity and tightness for solutions of martingale problems

We will extend the martingale argument for proving Donsker's Theorem that has been sketched above to limits of general martingale problems on the space $C = C([0,\infty), S)$ where S is a polish space. We first introduce a more general framework that allows to include non-Markovian processes. The reason is that it is sometimes convenient to approximate Markov processes by processes with a delay, see the proof of Theorem 6.8 below.

Suppose that $\mathcal{A}$ is a linear subspace of $\mathcal{F}_b(S)$, and

$$f \mapsto (\mathcal{L}_t f)_{t\geq 0}$$

is a linear map defined on $\mathcal{A}$ such that

$$(t,x) \mapsto (\mathcal{L}_t f)(x) \quad \text{is a function in } \mathcal{L}^2([0,T]\times C, \lambda \otimes P)$$

for any $T \in \mathbb{R}_+$ and $f \in \mathcal{A}$. The main example is still the one of time-homogeneous Markov processes with generator $\mathcal{L}$ where we set

$$\mathcal{L}_t f := (\mathcal{L} f)(X_t).$$

We say that the canonical process $X_t(\omega) = \omega(t)$ **solves the martingale problem** MP($\mathcal{L}_t, \mathcal{A}$) w.r.t. a probability measure P on C iff

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathcal{L}_r f \, dr$$

is a martingale under P for any $f \in \mathcal{A}$. Note that for $0 \leq s \leq t$,

$$f(X_t) - f(X_s) = M_t^f - M_s^f + \int_s^t \mathcal{L}_r f \, dr. \tag{6.5}$$

Therefore, martingale inequalities can be used to control the regularity of the process $f(X_t)$. As a first step in this direction we compute the angle-bracket process $\langle M^f \rangle$, i.e., the martingale part in the Doob-Meyer decomposition of $(M^f)^2$. Since we are considering processes with continuous paths, the angle-bracket process coincides with the quadratic variation $[M^f]$. The next theorem, however, is also valid for processes with jumps where $\langle M^f \rangle \neq [M^f]$:

Theorem 6.5 (Angle-bracket process for solutions of martingale problems). Let $f, g \in \mathcal{A}$ such that $f \cdot g \in \mathcal{A}$. Then

$$M_t^f \cdot M_t^g = N_t^{f,g} + \int_0^t \Gamma_r(f,g)\, dr \qquad \text{for any } t \geq 0,$$

where $N^{f,g}$ is a martingale, and

$$\Gamma_t(f,g) = \mathcal{L}_t(f\cdot g) - f(X_t)\mathcal{L}_t g - g(X_t)\mathcal{L}_t f.$$

Thus

$$\langle M^f, M^g \rangle_t = \int_0^t \Gamma_r(f,g)\, dr.$$

Example (Time-homogeneous Markov processes, Carré du champ operator).
Here $\mathcal{L}_t f = (\mathcal{L} f)(X_t)$, and therefore

$$\Gamma_t(f,g) = \Gamma(f,g)(X_t),$$

where $\Gamma : \mathcal{A} \times \mathcal{A} \to \mathcal{F}(S)$ is the **Carré du champ operator** defined by

$$\Gamma(f,g) = \mathcal{L}(f \cdot g) - f\mathcal{L}g - g\mathcal{L}f.$$

If $S = \mathbb{R}^d, \mathcal{A}$ is a subset of $C^\infty(\mathbb{R}^d)$, and

$$(\mathcal{L}f)(x) = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^d b_i(x)\frac{\partial f}{\partial x_i}(x) \quad \forall f \in \mathcal{A}$$

with measurable coefficients a_{ij}, b_i then

$$\Gamma(f,g)(x) = \sum_{i,j=1}^d a_{ij}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial g}{\partial x_j}(x) \quad \forall f,g \in \mathcal{A}.$$

In particular, for $a_{ij} \equiv \delta_{ij}, \Gamma(f,f) = |\nabla f|^2$ which explains the name "carré du champ" (= square field) operator. For general symmetric coefficients a_{ij} with $\det(a_{ij}) > 0$, the carré du champ is the square of the gradient w.r.t. the intrinsic metric $(g_{ij}) = (a_{ij})^{-1}$:

$$\Gamma(f,f) = \|\mathrm{grad}_g f\|_g^2.$$

Proof (Proof of Theorem 6.5). We may assume $f = g$, the general case follows by polarization. We write "$X \sim_s Y$" if $E[X|\mathcal{F}_s] = E[Y|\mathcal{F}_s]$ almost surely. To prove the claim we have to show that for $0 \le s \le t$ and $f \in \mathcal{A}$,

$$(M_t^f)^2 - (M_s^f)^2 \sim_s \int_s^t \Gamma_r(f,f)\,dr.$$

Since M^f is a square-integrable martingale, we have

$$\begin{aligned}(M_t^f)^2 - (M_s^f)^2 &\sim_s (M_t^f - M_s^f)^2 = \left(f(X_t) - f(X_s) - \int_s^t \mathcal{L}_r f\,dr\right)^2 \\ &= (f(X_t) - f(X_s))^2 - 2(f(X_t) - f(X_s))\int_s^t \mathcal{L}_r f\,dr + \left(\int_s^t \mathcal{L}_r f\,dr\right)^2 \\ &= I + II + III + IV\end{aligned}$$

where

$$\begin{aligned}I &:= f(X_t)^2 - f(X_s)^2 \sim_s \int_s^t \mathcal{L}_r f^2\,dr, \\ II &:= -2f(X_s)\left(f(X_t) - f(X_s) - \int_s^t \mathcal{L}_r f\,dr\right) \sim_s 0, \\ III &:= -2f(X_t)\int_s^t \mathcal{L}_r f\,dr = -2\int_s^t f(X_t)\mathcal{L}_r f\,dr, \quad \text{and} \\ IV &:= \left(\int_s^t \mathcal{L}_r f\,dr\right)^2 = 2\int_s^t \int_r^t \mathcal{L}_r f\,\mathcal{L}_u f\,du dr.\end{aligned}$$

Noting that $f(X_t)\mathcal{L}_r f \sim_r \left(f(X_r) + \int_r^t \mathcal{L}_u f\,du\right)\mathcal{L}_r f$, we see that for $s \le r \le t$ also the conditional expectations given $\mathcal{F}_s$ of these terms agree, and therefore

$$III \sim_s -2\int_s^t f(X_r)\mathcal{L}_r f\,dr - 2\int_s^t \int_r^t \mathcal{L}_r f\,\mathcal{L}_u f\,du dr.$$

Hence in total we obtain

$$(M_t^f)^2 - (M_s^f)^2 \sim_s \int_s^t \mathcal{L}_r f^2\,dr - 2\int_s^t f(X_r)\mathcal{L}_r f\,dr = \int_s^t \Gamma_r f\,dr.$$

■

We can now derive a bound for the modulus of continuity of $f(X_t)$ for a function $f \in \mathcal{A}$. Let

$$\omega^f_{\delta,T} := \omega_{\delta,T}(f \circ X), \qquad V^f_{s,t} := \sup_{r\in[s,t]} |f(X_r) - f(X_s)|.$$

Lemma 6.6 (Modulus of continuity of solutions to martingale problems). *For $p \in [2,\infty)$ there exist universal constants $C_p, \widetilde{C}_p \in (0,\infty)$ such that the following bounds hold for any solution (X_t, P) of a martingale problem as above and for any function $f \in \mathcal{A}$ such that the process $f(X_t)$ has continuous paths:*

1) For any $0 \le s \le t$,

$$\|V^f_{s,t}\|_{L^p(P)} \le C_p (t-s)^{1/2} \sup_{r\in[s,t]} \|\Gamma_r(f,f)\|^{1/2}_{L^{p/2}(P)} + (t-s) \sup_{r\in[s,t]} \|\mathcal{L}_r f\|_{L^p(P)}.$$

2) For any $\delta, \varepsilon, T \in (0,\infty)$,

$$P\left[\omega^f_{\delta,T} \ge \varepsilon\right] \le \widetilde{C}_p\, \varepsilon^{-p} \left(1 + \left\lceil \frac{T}{\delta} \right\rceil\right) \cdot \left(\delta^{p/2} \sup_{r\le T} \|\Gamma_r(f,f)\|^{p/2}_{L^{p/2}(P)} + \delta^p \sup_{r\le t} \|\mathcal{L}_r f\|_{L^p(P)}\right).$$

Proof. 1) By (6.5),

$$V^f_{s,t} \le \sup_{r\in[s,t]} |M^f_r - M^f_s| + \int_s^t |\mathcal{L}_u f|\, du.$$

Since $f(X_t)$ is continuous, M^f is a continuous martingale. Therefore, by **Burkholder's inequality**,

$$\begin{aligned}
\left\| \sup_{r\in[s,t]} \left| M^f_r - M^f_s \right| \right\|_{L^p(P)} &\le C_p \left\| \langle M^f\rangle_t - \langle M^f\rangle_s \right\|^{1/2}_{L^{p/2}(P)} \\
&= C_p \left\| \int_s^t \Gamma_r(f,f)\, dr \right\|^{1/2}_{L^{p/2}(P)} \\
&\le C_p (t-s)^{1/2} \sup_{r\in[s,t]} \|\Gamma_r(f,f)\|^{1/2}_{L^{p/2}(P)}.
\end{aligned}$$

For $p = 2$, Burkholder's inequality reduces to the usual maximal inequality for martingales - a proof for $p > 2$ can be found in many stochastic analysis textbooks, cf. e.g. [17].

2) We have already remarked above that the modulus of continuity $\omega^f_{\delta,T}$ can be controlled by bounds for $V^f_{s,t}$ on intervals $[s,t]$ of length δ. Here we obtain

$$\begin{aligned}
P\left[\omega^f_{\delta,T} \ge \varepsilon\right] &\le \sum_{k=0}^{\lfloor T/\delta \rfloor} P\left[V^f_{k\delta,(k+1)\delta} \ge \varepsilon/3\right] \\
&\le \sum_{k=0}^{\lfloor T/\delta \rfloor} \left(\frac{3}{\varepsilon}\right)^p \left\|V^f_{k\delta,(k+1)\delta}\right\|^p_{L^p(P)}.
\end{aligned}$$

The estimate in 2) now follows from 1). ∎

Remark. 1) The right-hand side in 2) converges to 0 as $\delta \downarrow 0$ if the suprema are finite and $p > 2$.

2) If $f(X_t)$ is not continuous then the assertion still holds for $p = 2$ but not for $p > 2$. The reason is that Burkholder's inequality for discontinuous martingales M_t is a bound in terms of the quadratic variation $[M]_t$ and not in terms of the angle bracket process $\langle M\rangle_t$. For continuous martingales, $\langle M\rangle_t = [M]_t$.

Example (Stationary Markov process).
If (X_t, P) is a stationary Markov process with generator extending $(\mathcal{L}, \mathcal{A})$ and stationary distribution $X_t \sim \mu$ then $\mathcal{L}_t f = (\mathcal{L}f)(X_t)$, $\Gamma_t(f,f) = \Gamma(f,f)(X_t)$, and therefore

$$\|\mathcal{L}_t f\|_{L^p(P)} = \|\mathcal{L} f\|_{L^p(\mu)}, \quad \|\Gamma_t(f,f)\|_{L^{p/2}(P)} = \|\Gamma(f,f)\|_{L^{p/2}(\mu)} \quad \text{for any } t \ge 0.$$

Construction of diffusion processes

The results above can be applied to prove the existence of diffusion processes generated by second order differential operators with continuous coefficients on $\mathbb{R}^d$. The idea is to obtain the law of the process as a weak limit of laws of processes with piecewise constant coefficients. The latter can be constructed from Brownian motion in an elementary way. The key step is again to establish tightness of the approximating laws.

Theorem 6.7 (Existence of diffusions in $\mathbb{R}^{\mathbf{d}}$). For $1 \le i, j \le d$ let $a_{ij}, b_i \in C_b(\mathbb{R}_+ \times \mathbb{R}^d)$ such that $a_{ij} = a_{ji}$. Then for any $x \in \mathbb{R}^d$ there exists a probability measure P_x on $C([0,\infty), \mathbb{R}^d)$ such that the canonical process (X_t, P_x) solves the martingale problem for the operator

$$\mathcal{L}_t f = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(t, X_t) \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) + \sum_{i=1}^{d} b_i(t, X_t) \frac{\partial f}{\partial x_i}(X_t)$$

with domain

$$\mathcal{A} = \left\{ f \in C^\infty(\mathbb{R}^d) : \frac{\partial f}{\partial x_i} \in C_b^\infty(\mathbb{R}^d) \text{ for } i = 1, \ldots, d \right\}$$

and initial condition $P_x[X_0 = x] = 1$.

Remark (Connections to SDE results). 1) If the coefficients are locally Lipschitz continuous then the existence of a diffusion process follows more easily from the Itô existence and uniqueness result for stochastic differential equations. The point is, however, that variants of the general approach presented here can be applied in many other situations as well.

2) The approximations used in the proof below correspond to Euler discretizations of the associated SDE.

Proof. 1) We first define the approximating generators and construct processes solving the corresponding martingale problems. For $n \in \mathbb{N}$ let

$$a_{ij}^{(n)}(t, X) = a_{ij}(\lfloor t \rfloor_n, X_{\lfloor t \rfloor_n}), \qquad b_i^n(t, X) = b_i(\lfloor t \rfloor_n, X_{\lfloor t \rfloor_n})$$

where $\lfloor t \rfloor_n := \max \left\{ s \in \frac{1}{n}\mathbb{Z} : s \le t \right\}$, i.e., for $t \in \left[\frac{k}{n}, \frac{k+1}{n}\right)$, we freeze the coefficients at their value at time $\frac{k}{n}$. Then the martingale problem for

$$\mathcal{L}_t^{(n)} f = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}^{(n)}(t, X) \frac{\partial^2 f}{\partial x_i \partial x_j}(X_t) + \sum_{i=1}^{d} b_i^{(n)}(t, X) \frac{\partial f}{\partial x_i}(X_t)$$

can be solved explicitly. Indeed let (B_t) be a Brownian motion on $\mathbb{R}^d$ defined on a probability space $(\Omega, \mathfrak{A}, P)$, and let $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ be measurable such that $\sigma\sigma^T = a$. Then the process $X_t^{(n)}$ defined recursively by

$$X_0^{(n)} = x, \quad X_t^{(n)} = X_{k/n}^{(n)} + \sigma\left(\frac{k}{n}, X_{k/n}^{(n)}\right)(B_t - B_{k/n}) + b\left(\frac{k}{n}, X_{k/n}^{(n)}\right)\frac{k}{n} \quad \text{for } t \in \left[0, \frac{1}{n}\right],$$

solves the martingale problem for $(\mathcal{L}_t^{(n)}, \mathcal{A})$ with initial condition δ_x. Hence the canonical process (X_t) on $C([0,\infty), \mathbb{R}^d)$ solves the same martingale problem w.r.t.

$$P^{(n)} = P \circ \left(X^{(n)}\right)^{-1}.$$

2) Next we prove tightness of the sequence $\{P^{(n)} : n \in \mathbb{N}\}$. For $i = 1, \dots, d$ let $f_i(x) := x_i$. Since $|x - y| \le \sum_{i=1}^d |f_i(x) - f_i(y)|$ for any $x, y \in \mathbb{R}^d$, we have

$$\omega_{\delta,T} \le \sum_{i=1}^d \omega_{\delta,T}^{f_i} \qquad \text{for any } \delta, T \in (0, \infty).$$

Furthermore, the functions

$$\mathcal{L}_t^{(n)} f_i = b_i^{(n)}(t, X) \quad \text{and} \quad \Gamma_t^{(n)}(f_i, f_i) = a_{ii}^{(n)}(t, X)$$

are uniformly bounded since the coefficients a_{ij} and b_i are bounded functions. Therefore, for any $\varepsilon, T \in (0, \infty)$,

$$P^{(n)}\left[\omega_{\delta,T} \ge \varepsilon\right] \le \sum_{i=1}^d P^{(n)}\left[\omega_{\delta,T}^{f_i} \ge \varepsilon/d\right] \to 0$$

uniformly in n as $\delta \downarrow 0$ by Lemma 6.6.
Hence by Theorem 6.4, the sequence $\{P^{(n)} : n \in \mathbb{N}\}$ is relatively compact, i.e., there exists a subsequential limit P^* w.r.t. weak convergence.

3) It only remains to show that (X_t, P^*) solves the limiting martingale problem. We know that $(X_t, P^{(n)})$ solves the martingale problem for $(\mathcal{L}_t^{(n)}, \mathcal{A})$ with initial law δ_x. In particular,

$$E^{(n)}\left[\left(f(X_t) - f(X_s) - \int_s^t \mathcal{L}_r^{(n)} f \, dr\right) g(X_{s_1}, \dots, X_{s_k})\right] = 0$$

for any $0 \le s_1 < s_2 < \cdots < s_k \le s \le t$ and $g \in C_b(\mathbb{R}^{k \cdot d})$. The assumptions imply that $\mathcal{L}_r^{(n)} f \to \mathcal{L}_r f$ pointwise as $n \to \infty$, and $\mathcal{L}_r^{(n)} f$ is uniformly bounded. This can be used to show that (X_t, P^*) solves the martingale problem for $(\mathcal{L}_t, f)$ - the details are left as an exercise. ■

Remark (Uniqueness). The assumptions in Theorem 6.7 are too weak to guarantee uniqueness of the solution. For example, the ordinary differential equation $dx = b(x)dt$ does not have a unique solution with $x_0 = 0$ when $b(x) = \sqrt{x}$. As a consequence, one can show that the trivial solution to the martingale problem for the operator $b(x)\frac{d}{dx}$ on $\mathbb{R}^1$ is not the only solution with initial law δ_0. A uniqueness theorem of Stroock and Varadhan states that the martingale problem has a unique solution for every initial law if the matrix $a(x)$ is strictly positive definite for each x, and the growth of the coefficients as $|x| \to \infty$ is at most of order $a_{ij}(x) = \mathrm{O}(|x|^2)$ and $b_i(x) = \mathrm{O}(|x|)$, cf. (24.1) in Roger&Williams II [48] for a sketch of the proof.

The general case

We finally state a general result on limits of martingale problems for processes with continuous paths. Let $(P^{(n)})_{n \in \mathbb{N}}$ be a sequence of probability measures on $C([0, \infty), S)$ where S is a polish space. Suppose that the canonical process $(X_t, P^{(n)})$ solves the martingale problem for $(\mathcal{L}_t^{(n)}, \mathcal{A})$ where $\mathcal{A}$ is a dense subspace of $C_b(S)$ such that $f^2 \in \mathcal{A}$ whenever $f \in \mathcal{A}$.

Theorem 6.8. Suppose that the following conditions hold:

(i) **Compact containment:** For any $T \in \mathbb{R}_+$ and $\gamma > 0$ there exists a compact set $K \subseteq S$ such that

$$P^{(n)}\left[\exists\, t \in [0, T] : X_t \notin K\right] \le \gamma \quad \text{for any } n \in \mathbb{N}.$$

(ii) **Uniform $\mathbf{L^p}$ bound:** There exists $p > 2$ such that for any $T \in \mathbb{R}_+$,

$$\sup_{n\in\mathbb{N}} \sup_{t\le T} \left(\left\| \Gamma_t^{(n)}(f,f) \right\|_{L^{p/2}(P^{(n)})} + \left\| \mathcal{L}_t^{(n)} f \right\|_{L^p(P^{(n)})} \right) < \infty.$$

Then $\{P^{(n)} : n \in \mathbb{N}\}$ is relatively compact. Furthermore, if

(iii) **Convergence of initial law:** There exists $\mu \in \mathcal{P}(S)$ such that

$$P^{(n)} \circ X_0^{-1} \xrightarrow{w} \mu \quad \text{as } n \to \infty, \text{ and}$$

(iv) **Convergence of generators:**

$$\mathcal{L}_t^{(n)} f \to \mathcal{L}_t f \qquad \text{uniformly for any } f \in \mathcal{A},$$

then any subsequential limit of $(P^{(n)})_{n\in\mathbb{N}}$ is a solution of the martingale problem for $(\mathcal{L}_t, \mathcal{A})$ with initial distribution μ.

The proof, including extensions to processes with discontinuous paths, can be found in Ethier and Kurtz [19].

Part III.

Markov Processes with given invariant measures

In this part we will study Markov processes that have a given probability measure μ as their invariant measure. The first important class of such processes are processes that are reversible when started with initial law μ. A reversible Markov process induces a strongly continuous contraction semigroup consisting of self-adjoint linear operators P_t on $L^2(S, \mu)$, and the corresponding generator is self-adjoint as well. This allows to describe and analyse reversible Markov processes efficiently in terms of the corresponding quadratic forms, which are called Dirichlet forms.

Beyond the reversible case, we will see that Markov processes with invariant measure μ induce C^0 contraction semigroups on $L^p(S, \mu)$ for any $p \in [1, \infty)$, and the corresponding generators can be decomposed into a symmetric part and an anti-symmetric part. In the diffusion case, the latter corresponds to a first order differential operator, i.e., to the generator of a deterministic flow. We will see that besides reversible Markov processes in the sense above, this class also includes processes satisfying a generalized reversibility condition. Typical examples are given by certain kinetic models such as the second order Langevin process.

Besides processes in continuous time, we will also consider Markov chains with a given invariant measure. These form the basis for Markov chain Monte Carlo methods. Moreover, we will consider measure preserving Markov processes on infinite dimensional state spaces.

7. Reversible Markov processes

In this chapter we will study Markov processes that are reversible when started with in equilibrium. Reversibility is equivalent to a detailed balance condition for the transition function. Therefore, a reversible Markov process induces a strongly continuous contraction semigroup consisting of self-adjoint linear operators P_t on the L^2 space w.r.t. the invariant measure, and the corresponding generator is self-adjoint as well. This allows the use of spectral theory. Moreover, it allows to describe and analyse reversible Markov processes efficiently in terms of the corresponding quadratic forms, which are called Dirichlet forms.

Besides processes in continuous time, we will also consider reversible Markov chains which form the basis for Markov chain Monte Carlo methods.

7.1. Reversibility and symmetry

Let S be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$. By $\mathcal{F}(S)$ and $\mathcal{F}_b(S)$ we denote the linear spaces of all measurable, bounded measurable functions $f\colon S \to \mathbb{R}$, respectively. Let $\mathcal{A}$ be a linear subspace of $\mathcal{F}_b(S)$, and let $E \subseteq \mathcal{F}_b(S)$ denote the closure of $\mathcal{A}$ with respect to the supremum norm. Then E is a Banach space. For example, if $S = \mathbb{R}^d$ and $\mathcal{A} = C_0^\infty(\mathbb{R}^d)$ then $E = \hat{C}(\mathbb{R}^d)$.

We assume that we are given a linear operator

$$\mathcal{L} : \mathcal{A} \subseteq E \to E,$$

and a right continuous time-homogeneous Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ with transition semigroup $(p_t)_{t\geq 0}$ such that for every $x \in S$, $(X_t)_{t\geq 0}$ is under P_x a *solution of the martingale problem for* $(\mathcal{L}, \mathcal{A})$ with $P_x[X_0 = x] = 1$. Thus for every $x \in S$ and $f \in \mathcal{A}$, the process

$$M_t^f = f(X_t) - \int_0^t (\mathcal{L}f)(X_s)\, ds$$

is an $(\mathcal{F}_t^X)$ martingale under P_x.

Moreover, we impose the following assumptions on the subspace $\mathcal{A}$ and the transition function (p_t).

(A1) For every $t \geq 0$ and $f \in \mathcal{A}$, $p_t f \in E$.

(A2) If μ is a signed measure on S with finite variation such that $\int f\, d\mu = 0$ for all $f \in \mathcal{A}$, then $\mu \equiv 0$.

Note that since p_t is contractive w.r.t. the sup norm, and $\mathcal{A}$ is dense in E, Assumption (A1) holds if and only if $p_t(E) \subseteq E$ for all $t \geq 0$. Assumption (A2) means that the subspace $\mathcal{A}$ is *separating*.

By Theorem 4.12, the assumptions (A1) and (A2) imply that (p_t) induces a C_0 contraction semigroup (P_t) of linear operators on the Banach space E such that the generator $(L, \mathrm{Dom}(L))$ is an extension of the operator $(\mathcal{L}, \mathcal{A})$. In particular, the forward equation

$$\frac{d}{dt} p_t f = p_t \mathcal{L} f \tag{7.1}$$

holds for any function $f \in \mathcal{A}$, and the backward equation

$$\frac{d}{dt} p_t f = \mathcal{L} p_t f \tag{7.2}$$

is satisfied for any function $f \in \mathcal{A}$ such that $p_t f$ is contained in the domain of L, which includes the completion of $\mathcal{A}$ w.r.t. the graph norm of $\mathcal{L}$. The derivatives in the forward and backward equation are taken on the Banach space E, i.e., they are defined as limits of difference quotients w.r.t. the supremum norm.

Notice that the *backward equation is valid only on a restricted class of functions*. This will cause technical difficulties in the proofs below. A common way to avoid these difficulties is to assume that $p_t(\mathcal{A}) \subseteq \mathcal{A}$ for any $t \geq 0$. However, this assumption is too restrictive for many applications. We will replace it by less restrictive conditions below, in particular we will assume that $\mathcal{A}$ is a *core* for the generator, see Definition 4.14 and Theorem 4.15. Often, it is still not easy to verify such an assumption in applications. In this case, one common strategy is to approximate the Markov process and its transition semigroup by more regular processes (e.g., by non-degenerate diffusions on $\mathbb{R}^d$), and to apply the results below to the approximating regularized processes. Unfortunately, there is no universal way how to implement this strategy. In the end, one has to decide on a case by case basis how to apply the techniques to a specific model.

Characterizations of reversibility

Theorem 7.1 (Characterizations of reversibility). Suppose (A1) and (A2) hold, and let $\mu \in \mathcal{P}(S)$. Then the following assertions are equivalent.

(i) The process (X_t, P_μ) is *invariant with respect to time reversal*, i.e., for every $t \geq 0$,

$$(X_s)_{0\leq s\leq t} \sim (X_{t-s})_{0\leq s\leq t} \quad \text{under } P_\mu.$$

(ii) The transition function (p_t) satisfies the *detailed balance condition*

$$\mu(dx)\, p_t(x,dy) \;=\; \mu(dy)\, p_t(y,dx) \quad \text{for all } t \geq 0.$$

(iii) The transition semigroup (P_t) is μ-symmetric, i.e.,

$$\int f\, P_t g\, d\mu \;=\; \int P_t f\, g\, d\mu \quad \text{for all } f,g \in \mathcal{F}_b(S) \text{ and } t \geq 0.$$

(iv) The generator $(L, \mathrm{Dom}(L))$ is μ-symmetric, i.e.,

$$\int f\, Lg\, d\mu \;=\; \int Lf\, g\, d\mu \quad \text{for all } f,g \in \mathrm{Dom}(L).$$

Moreover, is $\mathcal{A}$ is a core for the generator, then these conditions are also equivalent to:

(v) The operator $(\mathcal{L}, \mathcal{A})$ is μ-symmetric, i.e.,

$$\int f\, \mathcal{L}g\, d\mu \;=\; \int \mathcal{L}f\, g\, d\mu \quad \text{for all } f,g \in \mathcal{A}.$$

Remark (Reversibility and stationarity). A reversible process (X_t, P_μ) is stationary, since applying reversibility consecutively on the time intervals $[0, s+u]$ and $[0,u]$ shows that for all $s,u \geq 0$,

$$(X_{s+t})_{0\leq t\leq u} \sim (X_{u-t})_{0\leq t\leq u} \sim (X_t)_{0\leq t\leq u} \quad \text{with respect to } P_\mu.$$

Similarly, the detailed balance condition (ii) implies that μ is an invariant probability measure for (p_t), since for any $B \in \mathcal{B}$,

$$(\mu p_t)(B) \;=\; \int_{S\times B} \mu(dx)p_t(x,dy) \;=\; \int_{S\times B} p_t(y,dx)\mu(dy) \;=\; \mu(B).$$

In particular, if reversibility holds then the transition function (p_t) induces strongly contraction semigroups on $L^p(S,\mu)$ for every $p \in [1,\infty)$, see Section 4.1 above.

Proof (of Theorem 7.1).

(i)⇒(ii): Reversibility implies that for every $t \geq 0$,

$$\mu(dx)p_t(x,dy) = P_\mu \circ (X_0,X_t)^{-1} = P_\mu \circ (X_t,X_0)^{-1} = \mu(dy)p_t(y,dx).$$

(ii)⇒(i): Let $t \geq 0$. By induction, (ii) implies

$$\begin{aligned}&\mu(dx_0)p_{t_1-t_0}(x_0,dx_1)p_{t_2-t_1}(x_1,dx_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},dx_n)\\ &\quad= \;\mu(dx_n)p_{t_n-t_{n-1}}(x_n,dx_{n-1})\cdots p_{t_2-t_1}(x_2,dx_1)p_{t_1-t_0}(x_1,dx_0)\end{aligned}$$

for all $n \in \mathbb{N}$ and $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$, and thus

$$E_\mu[f(X_0,X_{t_1},X_{t_2},\ldots,X_{t_{n-1}},X_t)] \;=\; E_\mu[f(X_t,X_{t-t_1},\ldots,X_{t-t_{n-1}},X_0)]$$

for all measurable functions $f \geq 0$. Hence the time-reversed distribution coincides with the original one on cylinder sets, and thus everywhere.

(ii)⇔(iii): By Fubini's Theorem, for all $t \geq 0$ and $f,g \in \mathcal{F}_b(S)$,

$$(f,P_t g)_{L^2(\mu)} \;=\; \int f p_t g\, d\mu \;=\; \iint f(x)\,g(y)\;\mu(dx)\,p_t(x,dy) \tag{7.3}$$

This expression is symmetric in f and g if and only if $\mu \otimes p_t$ is a symmetric measure on $S \times S$.

(iii)⇒(iv): Suppose that (P_t) is symmetric on $L^2(\mu)$, and let $f,g \in \mathrm{Dom}(L)$. Then

$$(f,Lg)_{L^2(\mu)} - (Lf,g)_{L^2(\mu)} \;=\; \int (f\,Lg - g\,Lf)\,d\mu \;=\; \frac{d}{dt}\int (f\,P_t g - g\,P_t f)\,d\mu\Big|_{t=0+} \;=\; 0.$$

Here we have used that the functions are bounded, and for $g \in \mathrm{Dom}(L)$, Lg is a uniform limit of the difference quotients $(P_t g - g)/t$ as $t \downarrow 0$.

(iv)⇒(ii): Fix $t > 0$, let $f,g \in \mathrm{Dom}(L)$, and consider the function

$$u(s) \;:=\; (P_s f,\, P_{t-s} g)_{L^2(\mu)}, \qquad s \in [0,t].$$

Then the backward equation implies that u is differentiable, and

$$\frac{d}{ds}u(s) \;=\; (LP_s f, P_{t-s}g)_{L^2(\mu)} - (P_s f, LP_{t-s}g)_{L^2(\mu)}\,.$$

If (iv) holds then the right hand side vanishes, and hence u is constant. Thus

$$(f,P_t g)_{L^2(\mu)} \;=\; u(0) \;=\; u(t) \;=\; (P_t f, g)_{L^2(\mu)}\,.$$

Hence by (7.3), the identity

$$\iint f(x)\,g(y)\;\mu(dx)\,p_t(x,dy) \;=\; \iint f(x)\,g(y)\;\mu(dy)\,p_t(y,dx)$$

holds for all functions f,g in the domain of the generator, and thus, in particular, for all $f,g \in \mathcal{A}$. Applying the separability assumption (A2) twice, one concludes that detailed balance holds, and hence (iii) is satisfied for all $f,g \in \mathcal{F}_b(S)$.

(iv)⇔(v): If $\mathcal{A}$ is a core for the generator then any function $f \in \mathrm{Dom}(L)$ can be approximated in the graph norm by functions in $\mathcal{A}$. Therefore, in this case, the symmetry of the generator extends from function in $\mathcal{A}$ to general functions in the domain of the generator. ■

Now suppose that μ is an invariant probability measure for the transition semigroup (p_t). Then (p_t) induces a strongly continuous contraction semigroup of linear operators on the Hilbert space $L^2(\mu)$, and the corresponding generator is an extension of the operator $(\mathcal{L}, \mathcal{A})$. Instead of considering this operator, we can also consider the corresponding bilinear form

$$\mathcal{E}(f,g) \;:=\; -(f,\, \mathcal{L}g)_{L^2(\mu)} \;=\; -\frac{d}{dt}(f,\, P_t g)_{L^2(\mu)}\Big|_{t=0+}, \qquad f,g \in \mathcal{A}.$$

By Theorem 7.1, if $\mathcal{A}$ is a core for the generator, then this bilinear form is symmetric if and only if the corresponding Markov process with initial law μ is reversible. $(\mathcal{E}, \mathcal{A})$ is called the *(pre-) Dirichlet form associated to the operator* $(\mathcal{L}, \mathcal{A})$. Later, we will see that a reversible Markov process can be characterized by an extension $(\mathcal{E}, \mathrm{Dom}(\mathcal{E}))$ of the symmetric bilinear form $(\mathcal{E}, \mathcal{A})$. This extension will be called the *Dirichlet form associated to the Markov process*.

We conclude this section with several examples of pre-Dirichlet forms $(\mathcal{E}, \mathcal{A})$, the associated symmetric linear operators $(\mathcal{L}, \mathcal{A})$, and corresponding reversible Markov processes whose generator extends this operator.

Jump processes and Dirichlet forms on finite state spaces

If S is finite then $\mathcal{F}(S) = \mathbb{R}^S$ is a finite dimensional vector space. A Markov process on S is a pure jump process with non-negative jump rates $\mathcal{L}(x,y)$, $y \neq x$. The corresponding generator $\mathcal{L}$ is defined on the full space $\mathcal{A} = \mathcal{F}(S)$, and it can be identified with the matrix $(\mathcal{L}(x,y))_{x,y\in S}$, where $\mathcal{L}(x,x) = -\sum_{y\neq x} \mathcal{L}(x,y)$. The associated bilinear form on $L^2(\mu)$ is then given by

$$\mathcal{E}(f,g) \;=\; -\sum_{x\in S} f(x)\,(\mathcal{L}g)(x)\,\mu(x) \;=\; -\sum_x \sum_y \mu(x)\mathcal{L}(x,y)\,f(x)\,(g(y)-g(x)). \tag{7.4}$$

Since the space of all functions $f: S \to \mathbb{R}$ is spanned by the indicator functions 1_a with $a \in S$, a Markov process (X_t, P_μ) with generator $\mathcal{L}$ and initial law μ is reversible if and only if for all $a,b \in S$, we have $\mathcal{E}(1_a, 1_b) = \mathcal{E}(1_b, 1_a)$, or, equivalently, iff

$$\mu(a)\mathcal{L}(a,b) \;=\; \mu(b)\mathcal{L}(b,a) \qquad \text{for all } a,b \in S.$$

This is a detailed balance condition for the generator. If it is satisfied then the Dirichlet form $(\mathcal{E}, \mathcal{A})$ can be rewritten in a symmetric form:

$$\mathcal{E}(f,g) \;=\; -\sum_{x,y\in S} \mu(x)\mathcal{L}(x,y)\,f(y)\,(g(x)-g(y)), \qquad \text{and thus} \tag{7.5}$$

$$\mathcal{E}(f,g) \;=\; \frac{1}{2}\sum_{x,y\in S} \mu(x)\mathcal{L}(x,y)\,(f(y)-f(x))\,(g(y)-g(x)). \tag{7.6}$$

Here the identity (7.6) follows by taking the average of (7.4) and (7.5). For $x \neq y$, the coefficient $\mu(x)\mathcal{L}(x,y)$ is called the equilibrium flow from x to y.

Gradient Dirichlet forms and overdamped Langevin dynamics

Next, we consider the prototype of a (pre-)Dirichlet form corresponding to a diffusion process. We assume that μ is an absolutely continuous probability measure on $\mathbb{R}^d$ that can be written in the form

$$\mu(dx) \;=\; \frac{1}{Z}\exp(-U(x))\,\lambda^d(dx)$$

with a function $U \in C^2(\mathbb{R}^d)$ and a normalizing constant $Z \in (0,\infty)$, and we consider the symmetric bilinear form

$$\mathcal{E}(f,g) \;=\; \frac{1}{2}\int_{\mathbb{R}^d} \nabla f \cdot \nabla g \, d\mu , \qquad f,g \in C_0^\infty(\mathbb{R}^d).$$

The corresponding symmetric linear operator $(\mathcal{L},\mathcal{A})$ can be computed by integration by parts. Since f and g are compactly supported functions, there is no boundary term, and we obtain $\mathcal{E}(f,g) = -(f,\mathcal{L}g)_{L^2(\mu)}$ where

$$\mathcal{L}g \;=\; \frac{1}{2}\Delta g \;-\; \frac{1}{2}\nabla U \cdot \nabla g, \qquad g \in C_0^\infty(\mathbb{R}^d).$$

Solving the corresponding martingale problem is equivalent to finding weak solutions for the stochastic differential equation

$$dX_t \;=\; b(X_t)\,dt \;+\; dB_t\,, \qquad b \;=\; -\frac{1}{2}\nabla U. \tag{7.7}$$

A diffusion process solving this equation is called *Kolmogorov process* or *overdamped Langevin dynamics*. It can be constructed by applying Girsanov's Theorem, see for example [16, 17].

Theorem 7.2 (Existence and reversibility of overdamped Langevin dynamics). Let $C = C([0,\infty);\mathbb{R}^d)$ denote the space of continuous paths on $\mathbb{R}^d$, endowed with the canonical σ-algebra $\mathcal{B} = \sigma(X_t : t \in [0,\infty))$, where $X_t(\omega) = \omega(t)$. Suppose that either

$$\liminf_{|x|\to\infty} \frac{x \cdot \nabla U(x)}{|x|^2} \;>\; -\infty, \qquad \text{or} \tag{7.8}$$

$$\lim_{|x|\to\infty} U(x) \;=\; \infty \quad \text{and} \quad \liminf_{|x|\to\infty} \left(|\nabla U(x)|^2 - \Delta U \right) \;>\; -\infty. \tag{7.9}$$

Then there exists a probability kernel $(x,A) \mapsto P_x[A]$ defined on $\mathbb{R}^d \times \mathcal{B}$ such that for every $x \in \mathbb{R}^d$, the canonical process (X_t,P_x) is a non-explosive Markov process that solves the martingale problem for the operator $(\mathcal{L},C_0^\infty(\mathbb{R}^d))$ wit initial law δ_x. Moreover, if we define $P_\mu[A] := \int P_x[A]\,\mu(dx)$, then the process (X_t,P_μ) is a reversible Markov process that solves the martingale problem with initial law μ.

We give an outline of the proof, for more details see for example [17, Chapter 1].

Proof. 1) *Construction of P_x by change of measure.* Let Q_x denote the law on C of Brownian motion starting at x, and let

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

where the integral is an Itô stochastic integral. An application of Itô's formula shows that the process $(Z_t)_{t\ge 0}$ is a martingale under Q_x if b is bounded. Since we do not assume boundedness of b, we need a localization procedure. For $n \in \mathbb{N}$ let $T_n := \inf\{t \ge 0 : |X_t| \ge n\}$. Then the stopped process $b(X_{\cdot\wedge T_n})$ is bounded, and hence $Z_{\cdot\wedge T_n}$ is a martingale under Q_x for every $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$. In particular, $\int Z_{t\wedge T_n}\,dQ_x = 1$ for all $t \ge 0$. Therefore it can be shown that for every x, there exists a unique probability measure P_x on $\sigma\left(\bigcup_n \mathcal{F}^X_{T_n}\right)$ such that

$$\left.\frac{dP_x}{dQ_x}\right|_{\mathcal{F}^X_{t\wedge T_n}} \;=\; Z_{t\wedge T_n} \qquad \text{for all } t \ge 0 \text{ and } n \in \mathbb{N}.$$

Furthermore, Girsanov's Theorem implies that under the measure P_x, the canonical process (X_t) is a weak solution of the SDE (7.7) for $t < \sup T_n$. Let $\zeta := \sup T_n$. We will show that our assumptions imply that

$$P_x[\zeta = \infty] \;=\; 1 \qquad \text{for all } x \in \mathbb{R}^d, \tag{7.10}$$

i.e., the process is almost surely non-explosive. Then for every x, the process (Z_t, Q_x) is a martingale for $t \in [0, \infty)$, and the measures P_x can be extended to $\mathcal{B}$ in such a way that $(x, A) \mapsto P_x[A]$ is a probability kernel, and for every $x \in \mathbb{R}^d$, the process (X_t, P_x) is a weak solution of the SDE (7.7) with initial law δ_x, and hence a solution of the martingale problem for $(\mathcal{L}, C_0^\infty(\mathbb{R}^d))$.

To show non-explosiveness, we apply Hasminskii's criterion with the Lyapunov function

$$V(x) \;=\; \begin{cases} |x|^2 + 1 & \text{if (7.8) is satisfied,} \\ U(x) - \min U + 1 & \text{if (7.9) is satisfied,} \end{cases}$$

respectively. One easily verifies that the drift conditions (7.8) and (7.9) imply $\mathcal{L}V \le \alpha V$ for some real constant α. Therefore, by Corollary 1.21, $\zeta = \infty$ holds P_x-almost surely, and

$$\int Z_t \, dQ_x \;\ge\; \lim_{n\to\infty} \int_{T_n \ge t} Z_{t\wedge T_n} \, dQ_x \;=\; \lim_{n\to\infty} P_x[T_n \ge t] \;=\; 1. \tag{7.11}$$

By Itô's formula and Fatou's lemma, the process (Z_t, Q_x) is a non-negative supermartingale with $\int Z_0 \, dQ_x = 1$, and thus (7.11) implies that it is a martingale satisfying $\int Z_t \, dQ_x = 1$ for every x.

2) *Markov property.* Let $x \in \mathbb{R}^d$, $0 \le s \le t$, and let $f : \mathbb{R}^d \to \mathbb{R}_+$ be a non-negative measurable function. Then, by the Markov property for Brownian motion,

$$\begin{aligned} E_{P_x}[f(X_t)|\mathcal{F}_s^X] \;&=\; E_{Q_x}[f(X_t) Z_t|\mathcal{F}_s^X]/Z_s \\ &=\; E_{Q_x}\left[f(X_t) \exp\left(\int_s^t b(X_r)\cdot dX_r - \frac{1}{2}\int_s^t |b(X_r)|^2 \, dr \right) \middle| \mathcal{F}_s^X \right] \\ &=\; E_{Q_{X_s}}[f(X_{t-s}) Z_{t-s}] \;=\; E_{P_{X_s}}[f(X_{t-s})] \end{aligned}$$

holds Q_x-almost surely, and hence also P_x-almost surely. This shows that (X_t, P_x) is again a time-homogeneous Markov process.

3) *Reversibility.* This follows from the reversibility of Brownian motion w.r.t. Lebesgue measure. Indeed, since $b(x) = -\nabla U(x)/2$, an application of Itô's formula for Brownian motion shows that Q_μ-almost surely,

$$Z_t \;=\; \exp\left(\frac{U(X_0) - U(X_t)}{2} + \frac{1}{4}\int_0^t (\Delta U - \frac{1}{2}|\nabla U|^2)(X_s) \, ds \right).$$

Therefore, for any $t \ge 0$ and any non-negative measurable function $F : C([0,t], \mathbb{R}^d) \to \mathbb{R}$, we obtain

$$\begin{aligned} & E_{P_\mu}[F(X_{0:t})] \;=\; E_{Q_\mu}[F(X_{0:t}) Z_t] \\ &= \int E_{Q_x}\left[F(X_{0:t}) \exp\left(\frac{U(X_0) - U(X_t)}{2} + \frac{1}{4}\int_0^t (\Delta U - \frac{1}{2}|\nabla U|^2)(X_s)\, ds \right) \right] \frac{\exp(-U(x))}{Z} \, dx \\ &= \frac{1}{Z} E_{Q_\lambda}\left[F(X_{0:t}) \exp\left(-\frac{1}{2}U(X_0) - \frac{1}{2}U(X_t) + \frac{1}{4}\int_0^t (\Delta U - \frac{1}{2}|\nabla U|^2)(X_s)\, ds \right) \right], \end{aligned}$$

where $X_{0:t} = (X_s)_{s\in[0,t]}$ and $Q_\lambda := \int Q_x \, \lambda(dx)$ is the σ-finite measure obtained by integrating Q_x w.r.t. Lebesgue measure. By the definition of Brownian motion, it can be verified that Q_λ is reversible in the sense that $Q_\lambda \circ (X_s)^{-1}_{s\in[0,t]} = Q_\lambda \circ (X_{t-s})^{-1}_{s\in[0,t]}$ for every $t \ge 0$. Since also the expression $-\frac{1}{2}U(X_0) - \frac{1}{2}U(X_t) + \frac{1}{4}\int_0^t (\Delta U - \frac{1}{2}|\nabla U|^2)(X_s)\, ds$ in the exponent of the density is invariant under time-reversal, we can conclude that $P_\mu \circ (X_s)^{-1}_{s\in[0,t]} = P_\mu \circ (X_{t-s})^{-1}_{s\in[0,t]}$ holds for every $t \ge 0$.

■

Remark (Non-gradient drifts). The construction of a Markov process solving the martingale problem for $\mathcal{L}f = \frac{1}{2}\Delta f + b \cdot \nabla f$ by change of measure can also be applied in a similar way if the drift b is not a gradient. However, in this case, the stochastic integral in the Girsanov density can not be eliminated by an application of Itô's formula, and so the argument used for proving reversibility fails. Indeed, it has been shown by Kolmogorov [30] that reversibility holds if and only if the drift is a gradient.

Non-constant diffusion coefficients and intrinsic metric

We now show that Dirichlet forms corresponding to more general reversible diffusion processes can be represented in a similar form as above if one introduces an appropriate Riemannian metric that is associated to the diffusion process. Because of this, it is often possible to carry over results for overdamped Langevin dynamics to other reversible diffusion processes.

Suppose that $a_{ij} : \mathbb{R}^d \to \mathbb{R}$, $1 \le i, j \le d$, are smooth functions such that for every $x \in \mathbb{R}^d$, the matrix $a(x) = (a_{ij}(x))_{1\le i,j\le d}$ is symmetric and positive definite with $\det(a(x)) > 0$. We consider the symmetric bilinear form

$$\mathcal{E}(f,h) \;=\; \frac{1}{2}\int \sum_{i,j=1}^{d} a_{ij}(x)\,\frac{\partial f}{\partial x_i}(x)\,\frac{\partial h}{\partial x_j}(x)\;\mu(dx), \qquad f,h \in C_0^\infty(\mathbb{R}^d). \tag{7.12}$$

In order to give a geometric interpretation, we define a Riemannian metric on $\mathbb{R}^d$ by setting

$$\langle v,w\rangle_x \;=\; v\cdot g(x)w \quad \text{for all } x\in\mathbb{R}^d, \quad \text{where} \quad g(x) \;:=\; a(x)^{-1}.$$

In this case, a Riemannian metric is just a map that assigns to every $x \in \mathbb{R}^d$ an inner product $\langle\cdot,\cdot\rangle_x$ such that for every $v, w \in \mathbb{R}^d$, the map $x \mapsto \langle v, w\rangle_x$ is smooth. The metric defined above is called the *intrinsic metric associated to the pre-Dirichlet form* $\mathcal{E}$. The *gradient* of a function $f \in C^1(\mathbb{R}^d)$ w.r.t. the intrinsic metric is the vector field $\operatorname{grad} f : \mathbb{R}^d \to \mathbb{R}^d$ defined by the identity

$$\langle v, \operatorname{grad} f(x)\rangle_x \;=\; (\partial_v f)(x) \;=\; v\cdot\nabla f(x) \qquad \text{for all } x, v\in\mathbb{R}^d,$$

where ∇ denotes the standard gradient w.r.t. the Euclidean metric. By definition of the metric, $\langle v, \operatorname{grad} f\rangle = v\cdot g\operatorname{grad} f$, and hence

$$\operatorname{grad} f(x) \;=\; g(x)^{-1}(\nabla f)(x) \;=\; (a\nabla f)(x) \qquad \text{for all } x\in\mathbb{R}^d.$$

In particular, for any $f, h \in C^1(\mathbb{R}^d)$, we obtain

$$\langle \operatorname{grad} f\,,\, \operatorname{grad} h\rangle \;=\; (a\nabla f)\cdot g a\nabla h \;=\; \nabla f\cdot a\nabla h,$$

and hence the bilinear form introduced in (7.12) can be written as

$$\mathcal{E}(f,h) \;=\; \frac{1}{2}\int \langle \operatorname{grad} f\,,\, \operatorname{grad} h\rangle\; d\mu, \qquad f,h\in C_0^\infty(\mathbb{R}^d).$$

Thus we see that by introducing an appropriate Riemannian geometry, the pre-Dirichlet form $\mathcal{E}$ can again be represented as a gradient Dirichlet form.

Remark (Carré du champ operator). The symmetric bilinear operator $\Gamma : C^1(\mathbb{R}^d)\times C^1(\mathbb{R}^d) \to \mathcal{F}(\mathbb{R}^d)$ defined by

$$\Gamma(f,h) \;=\; \langle \operatorname{grad} f\,,\, \operatorname{grad} h\rangle \;=\; \nabla f\cdot a\nabla h$$

is called the *Carré du champ operator (square field operator)* corresponding to the Dirichlet form $\mathcal{E}$. Carré du champ operators can be defined for a more broad class of diffusion processes, including in particular (but not exclusively) diffusions on Riemannian manifolds. They can be used to set up a formal calculus that is the basis for extending many results from diffusion processes on $\mathbb{R}^d$ to a more general context, see [2].

We now compute the symmetric linear operator $(\mathcal{L}, C_0^\infty(\mathbb{R}^d))$ associated to the symmetric bilinear form (7.12) on the Hilbert space $L^2(\mu)$. By integration by parts, we obtain

$$\mathcal{E}(f,h) \;=\; -\int f\,\mathcal{L}h\,d\mu \qquad \text{for all } f,h\in C_0^\infty(\mathbb{R}^d),$$

where $\mathcal{L}h$ is given by

$$\begin{aligned}
\mathcal{L}h &= \frac{1}{2}e^{U}\nabla\cdot\left(e^{-U}a\nabla h\right) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\frac{\partial^2 h}{\partial x_i\partial x_j} + \sum_{j=1}^{d} b_j\frac{\partial h}{\partial x_j},\\
b_j &= \frac{1}{2}\sum_{i=1}^{d} e^{U}\frac{\partial}{\partial x_i}\left(e^{-U}a_{ij}\right) = \frac{1}{2}\sum_{i=1}^{d}\left(\frac{\partial a_{ij}}{\partial x_i} - a_{ij}\frac{\partial U}{\partial x_i}\right).
\end{aligned}$$

The generator $\mathcal{L}$ is a second-order elliptic differential operator in divergence form. Alternatively, it can be written in a more geometric way by performing integration by parts on the Riemannian manifold $(\mathbb{R}^d, g)$. Let $d\mathrm{vol}_g = \sqrt{\det g}\, d\lambda$ denote the Riemannian volume measure. Then

$$d\mu = \frac{1}{Z}e^{-V}\, d\mathrm{vol}_g, \qquad \text{where} \quad V := U + \frac{1}{2}\log\det g.$$

By applying the integration by parts formula with respect to the Riemannian volume we obtain

$$\mathcal{E}(f,h) = \frac{1}{2}\int \langle \operatorname{grad} f, \operatorname{grad} h\rangle\, \frac{1}{Z}e^{-V}\, d\mathrm{vol}_g = -\int f\,\mathcal{L}h\,\frac{1}{Z}e^{-V}\, d\mathrm{vol}_g = -(f, \mathcal{L}h)_{L^2(\mu)},$$

where $\mathcal{L}h$ is given by

$$\mathcal{L}h = \frac{1}{2}e^{V}\mathrm{div}_g\left(e^{-V}\mathrm{grad}_g h\right) = \frac{1}{2}\Delta_g h - \frac{1}{2}\langle \operatorname{grad} V, \operatorname{grad} h\rangle.$$

Here the divergence operator div_g is defined as the adjoint of the gradient grad_g w.r.t. the Riemannian volume, and $\Delta_g := \mathrm{div}_g\mathrm{grad}_g$ is the *Laplace-Beltrami operator on the Riemannian manifold* $(\mathbb{R}^d, g)$.

Under appropriate assumptions on the coefficients a_{ij} and b_j, the existence of a diffusion process solving the martingale problem for the operator $(\mathcal{L}, C_0^\infty(\mathbb{R}^d))$ follows from Theorem 6.7. Alternatively, if $a(x)$ can be represented as $a(x) = \sigma(x)\sigma(x)^T$ with a locally Lipschitz continuous function $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, then a solution of the martingale problem can be obtained as a solution of the corresponding stochastic differential equation

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

driven by a Brownian motion in $\mathbb{R}^d$, see for instance [17]. Under an appropriate Lyapunov type drift condition, non-explosiveness can then be shown again via Hasminskii's criterion, see Corollary 1.21. If $V \equiv 0$ then $\mathcal{L} = \frac{1}{2}\Delta_g$, and an associated diffusion process is (by definition) a *Brownian motion on the Riemannian manifold* $(\mathbb{R}^d, g)$.

7.2. Self-adjoint operators and spectral theory

By Theorem 7.1, a reversible Markov process induces a C^0 contraction semigroup $(P_t)_{t\geq 0}$ of *symmetric* linear operators on the Hilbert space $L^2(S,\mu)$, where S is the state space and μ is the invariant law of the process at any given time t. We will see below that the corresponding generator $(L, \mathrm{Dom}(L))$ is a negative definite self-adjoint linear operator on this Hilbert space. As a consequence, spectral theory can be applied to study the transition semigroup and its relation to the generator. In particular, we will see that $P_t = \exp(tL)$ where the operator exponential is defined via the spectral theorem for self-adjoint linear operators. For further background on spectral theory for self-adjoint operators see e.g. Reed and Simon [45, 46] or Yosida [56].

Spectral decomposition for compact symmetric operators

Before stating the spectral theorem for general self-adjoint linear operators on a Hilbert space, we consider the case of compact operators.

Definition 7.3 (Bounded and compact linear operators). Suppose that E and F are Banach spaces, and let $B(0,1)$ denote the unit ball in E.

1) A linear operator $K : E \to F$ is called *bounded* iff $K(B(0,1))$ is bounded. The linear space of all bounded linear operators from E to F is denoted by $\mathcal{L}(E,F)$.

2) K is called *compact* iff the closure of $K(B(0,1))$ is compact in F.

A linear operator K is bounded if and only if K is continuous, or, equivalently, if and only if there exists a finite constant C such that $\|Kf\|_F \le C\|f\|_F$ for all $f \in E$. The smallest constant C with this property is the *operator norm*

$$\|K\|_{\mathcal{L}(E,F)} \;:=\; \sup\{\|Kf\|_F : f \in E \text{ with } \|f\|_F = 1\}\,.$$

Moreover, K is compact if and only if for every bounded sequence $(f_n)_{n\in\mathbb{N}}$, the sequence $(Kf_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Of course, every compact linear operator is bounded, but the converse is only true if the dimension of F is finite.

Example (Compact integral operators). Suppose that $E = F = L^2(S,\mu)$, and K is a *Hilbert-Schmidt* integral operator given by

$$(Kf)(x) \;=\; \int K(x,y)\, f(y)\, \mu(dy),$$

where $K : S \times S \to \mathbb{R}$ is a square integrable function, i.e.,

$$\int\int K(x,y)^2\, \mu(dx)\, \mu(dy) \;<\; \infty.$$

Then K is compact, see for example Yosida [56, p. X.2]. This applies for example if S is a bounded domain in $\mathbb{R}^d$ with smooth boundary, and $K = P_t$ or $K = G_\alpha$ for some $t, \alpha > 0$, where $(P_t)_{t\ge 0}$ and $(G_\alpha)_{\alpha>0}$ are the transition semigroup and the resolvent of Brownian motion on S with normal reflection or absorption at the boundary.

From now on, we assume that H is a separable real Hilbert space of infinite dimension with inner product $(\cdot,\cdot)$. Later, we will usually choose $H = L^2(S,\mu)$. A bounded linear operator $K : H \to H$ can be characterized by its associated bilinear form

$$Q(f,g) \;=\; (f,Kg), \qquad f,g \in H.$$

By definition, K is symmetric iff Q is symmetric, and K is non-negative definite iff $Q(f,f) \ge 0$ for all $f \in H$. An element $f \in H$ is an eigenfunction of K with eigenvalue λ iff $Kf = \lambda f$.

Theorem 7.4 (Rayleigh-Ritz variational principle). Suppose that $K : H \to H$ is a symmetric non-negative definite compact linear operator. Then there exists an orthonormal basis[1] $\{e_i : i \in \mathbb{N}\}$ consisting of eigenfunctions of K with real eigenvalues λ_n where $\lambda_1 = \|K\|_{\mathcal{L}(H,H)}$, $\lambda_n \ge \lambda_{n+1}$ for all n, and $\lim_{n\to\infty} \lambda_n = 0$. The eigenvalues and eigenfunctions are given inductively by the variational characterization

$$\lambda_n \;=\; \max\{Q(f,f) : f \in H \text{ with } \|f\| = 1, f \perp \operatorname{span}\{e_1,\dots,e_{n-1}\}\}\,,$$

where e_n is a maximizer of the variational expression for λ_n.

For the proof see for example [1, p. 10.14]. The key point is to show that the maxima are achieved. This follows by the Banach-Alaoglu Theorem (existence of a weakly convergent subsequence of a bounded sequence in H) and compactness of K. Of course, the same result holds if H is finite dimensional, except that in this case, there are only finitely many eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$.

Self-adjointness

Many generators of Markov processes are unbounded linear operators. Often, the resolvent is compact, and thus it is still possible to apply the spectral decomposition for compact linear operators. However, this is not always the case, for example the resolvent of Brownian motion on $\mathbb{R}^d$ is not compact. Moreover, it is tedious to verify compactness in each case. Therefore, it is important to extend the spectral theorem to unbounded linear operators. These can only be defined on a dense subspace of the Hilbert space, and it turns out that symmetry on its own is not sufficient to obtain a spectral decomposition.

Definition 7.5 (Adjoint operator, self-adjoint operator). Suppose that $L : \mathrm{Dom}(L) \subseteq H \to H$ is a densely defined linear operator on H.

1) The *adjoint operator* $(L^*, \mathrm{Dom}(L^*))$ is defined by

$$\mathrm{Dom}(L^*) \;=\; \{f \in H : \exists\, L^* f \in H : \ (f, Lg) = (L^* f, g) \text{ for all } g \in \mathrm{Dom}(L)\}\,.$$

2) The operator $(L, \mathrm{Dom}(L))$ is called *self-adjoint* iff

$$(L, \mathrm{Dom}(L)) \;=\; (L^*, \mathrm{Dom}(L^*)). \tag{7.13}$$

3) The operator $(L, \mathrm{Dom}(L))$ is called *essentially self-adjoint* iff its closure is a self-adjoint operator.

It is important to observe that *self-adjointness is stronger than symmetry*. Indeed, the definition of the adjoint shows that the operator $(L, \mathrm{Dom}(L))$ is symmetric if and only if

$$(L, \mathrm{Dom}(L)) \;\subseteq\; (L^*, \mathrm{Dom}(L^*)). \tag{7.14}$$

In general, however, there can be functions in the domain of the adjoint L^* of a symmetric linear operator that are not contained in the domain of L. On the other hand, (7.14) shows that a symmetric linear operator with $\mathrm{Dom}(L) = H$ is also self-adjoint. In particular, this applies to densely defined bounded linear operators, and thus to all symmetric linear operators on a finite dimensional Hilbert space.

Exercise (Generators of Brownian motions with boundary conditions on (0, 1)). Consider $Lf = f''/2$ as a linear operator on the Hilbert space $H = L^2(0,1)$. Prove the following assertions.

1) The linear operator $(L, C_0^\infty(0,1))$ is symmetric but not self-adjoint. Moreover, the domain of the adjoint operator L^* is the Sobolev space

$$H^{2,2}(0,1) \;=\; \left\{ f \in C^1([0,1]) : f' \text{ is absolutely continuous with } f'' \in L^2(0,1) \right\}.$$

2) The operator $(L, H^{2,2}(0,1))$ is not even symmetric.

3) Now consider the domains

$$\begin{aligned} \mathcal{D}_{\mathrm{Dirichlet}} &= \left\{ f \in H^{2,2}(0,1) : f(0) = f(1) = 0 \right\}, \\ \mathcal{D}_{\mathrm{Neumann}} &= \left\{ f \in H^{2,2}(0,1) : f'(0) = f'(1) = 0 \right\}, \\ \mathcal{D}_{\mathrm{periodic}} &= \left\{ f \in H^{2,2}(0,1) : f(0) = f(1) \text{ and } f'(0) = f'(1) \right\}. \end{aligned}$$

Show that the operator L is self-adjoint on each of these domains. Which reversible Markov processes correspond to these generators?

4) Give more examples of self-adjoint extensions of the operator $(L, C_0^\infty(0,1))$. Are all these extensions generators of transition semigroups of reversible Markov processes?

Theorem 7.6 (Generators of reversible Markov processes are self-adjoint). Suppose that $(P_t)_{t\geq 0}$ is a symmetric C^0 contraction semigroup of linear operators on H. Then the following assertions hold.

1) For every $t \geq 0$, P_t is a self-adjoint contraction.

2) The generator $(L, \mathrm{Dom}(L))$ is self-adjoint and negative definite, i.e., $(f, Lf) \leq 0$ for any $f \in \mathrm{Dom}(L)$.

Proof. 1) This follows from the remark above since P_t is symmetric and defined on all of H.

2) We first observe that L is symmetric and negative definite, since for any $f, g \in \mathrm{Dom}(L)$,

$$\begin{aligned} (f, Lf) &= \lim_{t\downarrow 0} \frac{1}{t} \{(f, P_t f) - (f, f)\} \leq 0, \qquad \text{and} \\ (f, Lg) - (Lf, g) &= \frac{d}{dt} \{(f, P_t g) - (P_t f, g)\}\Big|_{t=0+} = 0. \end{aligned}$$

Hence (7.14) is satisfied, and it "only" remains to show $\mathrm{Dom}(L) \supseteq \mathrm{Dom}(L^*)$.

To this end, let $g \in \mathrm{Dom}(L^*)$, and fix an arbitrary real $\alpha > 0$. By the Hille-Yosida Theorem, the range of $\alpha I - L$ is the whole Hilbert space H. Hence there exists $f \in \mathrm{Dom}(L)$ such that

$$\alpha g - L^* g = \alpha f - Lf = \alpha f - L^* f. \tag{7.15}$$

Here the last identity holds by (7.14). Equation (7.15) shows that $f - g$ is contained in the kernel of the operator $\alpha I - L^*$. It can be verified that this kernel is the orthogonal complement of the range of the operator $\alpha I - L$.. But by the Hille-Yosida Theorem, this range is the whole Hilbert space, and thus $f - g = 0$. Hence we have shown that $g = f$, and thus $g \in \mathrm{Dom}(L)$.

■

The spectral theorem for self-adjoint operators

We will now state a version of the spectral theorem for self-adjoint operators. The next definition is crucial.

Definition 7.7 (Resolution of the identity). A *resolution of the identity* on the Hilbert space H is a family of symmetric linear operators $E_\lambda : H \to H$ $(\lambda \in \mathbb{R})$ satisfying the following properties.

(i) *Projectivity:* For any $\lambda, \mu \in \mathbb{R}$, $E_\lambda E_\mu = E_{\min(\lambda,\mu)}$.

(ii) *Normalization:* For any $f \in H$, $\lim_{\lambda\downarrow-\infty} E_\lambda f = 0$ and $\lim_{\lambda\uparrow\infty} E_\lambda f = f$.

(iii) *Right continuity:* For any $f \in H$ and $\lambda \in \mathbb{R}$, $\lim_{h\downarrow 0} E_{\lambda+h} f = E_\lambda f$.

Notice that Condition (i) implies that each linear operator E_λ is a self-adjoint projection, i.e., $E_\lambda^2 = E_\lambda$. Furthermore, if $(E_\lambda)_{\lambda\in\mathbb{R}}$ is a resolution of the identity, then for every $f \in H$, the function $\lambda \mapsto (f, E_\lambda f)$ is the distribution function of a positive measure μ_f on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and, correspondingly, for every $f, g \in H$, $\lambda \mapsto (f, E_\lambda g)$ is the distribution function of a signed measure $\mu_{f,g}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We write

$$\mu_f(d\lambda) = d(f, E_\lambda f) \qquad \text{and} \qquad \mu_{f,g}(d\lambda) = d(f, E_\lambda g).$$

In this sense, one can view a symmetric resolution of the identity as the *distribution function of a projection-valued measure* dE_λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

The spectral theorem states that for every self-adjoint operator on H there is a resolution of the identity that provides a spectral decomposition of the operator. We first consider the special case of an operator with discrete spectrum. In this case, the spectral measure dE_λ is also discrete.

Example (Discrete spectrum). Suppose that $(L, \mathrm{Dom}(L))$ is a self-adjoint linear operator with *discrete spectrum*, i.e., there is an orthonormal basis $\{e_i : i \in \mathbb{N}\}$ consisting of eigenfunctions of L. Since the operator is self-adjoint, the eigenvalues λ_i are real, and a resolution of the identity is given by

$$E_\lambda f \;=\; \sum_{i:\lambda_i \le \lambda} e_i\,(e_i, f), \qquad \lambda \in \mathbb{R}.$$

Here, E_λ is the orthogonal projection onto the closed subspace $H_{(-\infty,\lambda]} \subseteq H$ that is defined as the completion of the linear span of all eigenfunctions with eigenvalues $\le \lambda$. Moreover,

$$(f, E_\lambda g) \;=\; \sum_{\alpha \le \lambda} (f, \Pi_\alpha g),$$

where the sum is over all eigenvalues α, and Π_α denotes the orthogonal projection onto the eigenspace $\ker(\alpha I - L)$ corresponding to the eigenvalue α. Thus, for $f, g \in H$, $\mu_{f,g}$ is the discrete measure given by

$$\mu_{f,g} \;=\; \sum_{\alpha \in \mathrm{spec}(L)} (f, \Pi_\alpha g)\,\delta_\alpha\,,$$

and the projection-valued spectral measure is formally given by

$$dE_\lambda \;=\; \sum_{\alpha \in \mathrm{spec}(L)} \Pi_\alpha\;\delta_\alpha(d\lambda)\,.$$

Finally, the operator L can be recovered from the spectral measure by the identity

$$\begin{aligned}
(f, Lg) \;&=\; \sum_{\alpha \in \mathrm{spec}(L)} \alpha\,(f, \Pi_\alpha g) \;=\; \int_{\mathbb{R}} \lambda\,\mu_{f,g}(d\lambda) \qquad \text{for all } f \in H \text{ and } g \in \mathrm{Dom}(L),\\
\mathrm{Dom}(L) \;&=\; \left\{ f \in H : \int_{\mathbb{R}} \lambda^2\,\mu_f(d\lambda) < \infty \right\}.
\end{aligned}$$

Exercise. Complete the missing details in the last example.

Although it is not always straightforward to verify, many generators of Markov processes that we are interested in actually do have a discrete spectrum. On the contrary, the generator of Brownian motion on the Hilbert space $H = L^2(\mathbb{R}^d)$ is $(\Delta/2, H^{2,2}(\mathbb{R}^d))$. This operator does not have a discrete spectrum, and its spectral decomposition is given by the Fourier transform. Because of this and other examples, the next theorem is crucial.

Theorem 7.8 (Spectral theorem for self-adjoint linear operators).

7.3. Symmetric Markov semigroups and Dirichlet forms

Closed quadratic forms

Definition 7.9 (Closed form; closability and closure).

Lemma 7.10 (Conditions for closability).

Example ().

From the quadratic form to generator and resolvent

Theorem 7.11 (Resolvent and generator of a quadratic form).

Self-adjoint extensions

Theorem 7.12 (Friedrichs extension).

Example (Overdamped Langevin dynamics).

From semigroup and resolvent to the quadratic form

Lemma 7.13 (Monotonicity).

Theorem 7.14 (Regularization of quadratic forms via the semigroup).

Exercise (Regularization via the resolvent).

One-to-one correspondences

Theorem 7.15 (Closed symmetric forms, self-adjoint operators and symmetric contraction semigroups).

Definition 7.16 (Dirichlet form; Dirichlet operator).

Theorem 7.17 (Dirichlet forms, Dirichlet operators and symmetric sub-Markov semigroups).

The assertion in the following exercise is often useful to verify in concrete examples that a given closed quadratic form is a Dirichlet form.

Exercise (Checking the Dirichlet property in practice). Show that a closed densely defined quadratic form $(\mathcal{E}, \mathrm{Dom}(\mathcal{E}))$ on $L^2(S,\mu)$ is a Dirichlet form if and only if there is a dense subspace $\mathcal{A} \subseteq \mathrm{Dom}(\mathcal{E})$ such that for any $f \in \mathcal{A}$ and for any function $\varphi \in C_b^\infty(\mathbb{R})$ with $\varphi(0) = 0$ and $0 \le \varphi' \le 1$, it holds

$$\varphi \circ f \in \mathrm{Dom}(\mathcal{E}) \quad \text{and} \quad \mathcal{E}(\varphi \circ f, \varphi \circ f) \le \mathcal{E}(f,f)\,.$$

Usually, the form $\mathcal{E}$ and the generator L are first given on an appropriate space $\mathcal{A}$ of smooth functions. Closability can then be verified by Lemma 7.10, and the exercise above can be applied to show that the closure is a Dirichlet form, and the associated Friedrichs extension of $(L,\mathcal{A})$ generates a sub-Markovian semigroup $(P_t)_{t\ge 0}$. The Markov representation theorem then ensures that each P_t is almost surely given by integration w.r.t. a sub-probability kernel $p_t(x,dy)$.

Theorem 7.18 (Markov representation theorem). Let μ be a probability measure on a Polish space S endowed with its Borel σ-algebra $\mathcal{B}$, and suppose that $P : \mathcal{F}_+(S) \to \mathcal{F}_+(S)$ is a *sub-Markovian* linear operator.

1) Suppose that P is bounded on $L^1(\mu)$, i.e., there exists a finite constant C such that

$$\|Pf\|_{L^1(\mu)} \le C \, \|f\|_{L^1(\mu)} \qquad \text{for all } f \in \mathcal{F}_+(S). \tag{7.16}$$

Then there exists a sub-probability kernel $p(x,dy)$ on $(S,\mathcal{B})$ such that for any $f \in \mathcal{F}_+(S)$,

$$(Pf)(x) = \int f(y)\, p(x,dy) \qquad \text{for } \mu\text{-almost every } x \in S. \tag{7.17}$$

In particular, this is the case if μ is sub-invariant for P.

2) If P is even bounded as an operator from $L^1(\mu)$ to $L^\infty(\mu)$, i.e., there exists a finite constant M such that

$$\|Pf\|_{L^\infty(\mu)} \le M \, \|f\|_{L^1(\mu)} \qquad \text{for all } f \in \mathcal{F}_+(S), \tag{7.18}$$

then there exists a non-negative measurable function $p : S \times S \to [0,M]$ such that for any $f \in \mathcal{F}_+(S)$,

$$(Pf)(x) = \int f(y)\, p(x,y)\, \mu(dy) \qquad \text{for } \mu\text{-almost every } x \in S. \tag{7.19}$$

Exercise (Proof of Markov representation theorem). Prove Theorem 7.18 by proceeding in the following steps.

a) Show that it suffices to prove the result under the additional assumption that P is Markov, i.e., $P1 = 1$. To this end extend a given sub-Markovian linear operator P on $\mathcal{F}_+(S)$ to a Markovian linear operator $\hat{P}$ on $\mathcal{F}_+(S \dot\cup \{\Delta\})$.

b) Show that if (7.16) holds then

$$Q(A,B) := \int 1_A \, P1_B \, d\mu, \qquad A,B \in \mathcal{B},$$

is a *bimeasure* with total mass 1, i.e., Q is a measure in each of its variables if the other variable is fixed, and $Q(S,S) = 1$.

c) You may assume without proof that for every such bimeasure, there is a probability measure q on the product space $S \times S$ such that

$$Q(A,B) = q(A \times B) \qquad \text{for all } A,B \in \mathcal{B},$$

see [42, 11]. Conclude that under (7.16), there exists a probability kernel $p(x, dy)$ such that

$$\int f\, Pg\, d\mu \;=\; \int f\, pg\, d\mu \qquad \text{for all } f, g \in \mathcal{F}_+(S),$$

and show that (7.17) holds.

2) Since S is a Polish space, there exists a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $\mathcal{B} = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$ and each $\mathcal{F}_n$ is generated by a finite partition of S. Let $P_n f$ denote the conditional expectation given $\mathcal{F}_n$ of Pf w.r.t. μ. Show that if (7.18) holds then for every $n \in \mathbb{N}$, there exists an $\mathcal{F}_n \otimes \mathcal{F}_n$ measurable function $p_n : S \times S \to [0, M]$ such that (7.19) holds with P and p replaced by P_n and p_n. Moreover, show that p_n is an $\mathcal{F}_n \otimes \mathcal{F}_n$ martingale under $\mu \otimes \mu$. Conclude that $p = \lim_{n\to\infty} p_n$ exists $\mu \otimes \mu$-almost surely, and (7.19) holds.

7.4. Model examples of reversible diffusions

Brownian motion on a flat torus

Brownian motion on a bounded domain

Standard Ornstein-Uhlenbeck processes

General Ornstein-Uhlenbeck processes

Brownian motion on the unit sphere

7.5. General reversible diffusions and transformations

One-dimensional diffusions

Overdamped Langevin dynamics

Sub-Markov semigroups and h-transform

Additive functionals

Time change

8. Markov Chain Monte Carlo methods

8.1. General Metropolis-Hastings method

Metropolis-Hastings algorithm

Two examples

Independence Sampler

8.2. Random Walk Metropolis

Gaussian case

General case

Examples

8.3. Langevin algorithms

Unadjusted Langevin Algorithm

ULA in the general case

Metropolis Adjusted Langevin Algorithm

Conductance bound for Metropolis-Hastings algorithms

8.4. “Dimension-free” Metropolis-Hastings

Gaussian measures on Hilbert spaces

Preconditioned Crank-Nicholson

Preconditioned MALA

9. Functional inequalities and convergence to equilibrium

Our goal in the following sections is to relate the long time asymptotics ($t \uparrow \infty$) of a time-homogeneous Markov process and its transition semigroup to its infinitesimal characteristics which describe the short-time behavior ($t \downarrow 0$):

$$\begin{array}{ccc} \text{Asymptotic properties} & \leftrightarrow & \text{Infinitesimal behavior, generator} \\ t \uparrow \infty & & t \downarrow 0 \end{array}$$

Although this is usually limited to the time-homogeneous case, some of the results can be applied to time-inhomogeneous Markov processes by considering the space-time process (t, X_t), which is always time-homogeneous. Useful additional references for this chapter are the books by Royer [50] and Bakry, Gentil, Ledoux [2], and the lecture notes by Malrieu [38].

Let S again be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$. As before, we assume that we are given a right continuous time-homogeneous Markov process $((X_t)_{t\geq 0}, (\mathcal{F}_t)_{t\geq 0}, (P_x)_{x\in S})$ with transition semigroup $(p_t)_{t\geq 0}$ such that for any $x \in S$, $(X_t)_{t\geq 0}$ is under P_x a solution of the martingale problem for $(\mathcal{L}, \mathcal{A})$ with $P_x[X_0 = x] = 1$. Here $\mathcal{A}$ is a linear subspace of $\mathcal{F}_b(S)$, and $\mathcal{L} : \mathcal{A} \to \mathcal{F}(S)$ is a linear operator. Furthermore, we assume from now on that μ is an invariant probability measure for $(p_t)_{t\geq 0}$. Then p_t is a contraction on $L^p(S, \mu)$ for all $p \in [1, \infty]$ since

$$\int |p_t f|^p \, d\mu \leq \int p_t |f|^p \, d\mu = \int |f|^p \, d\mu \quad \forall f \in \mathcal{F}_b(S)$$

by Jensen's inequality and invariance of μ. The assumptions on $\mathcal{A}_0$ and $\mathcal{A}$ imposed above can be relaxed in the following way:

(A0) If μ is a signed measure on S with finite variation such that $\int f \, d\mu = 0$ for any $f \in \mathcal{A}$, then $\mu \equiv 0$.

(A1') $f,\ \mathcal{L}f \in \mathcal{L}^p(S, \mu)$ for all $1 \leq p < \infty$

(A2') $\mathcal{A}_0$ is dense in $\mathcal{A}$ with respect to the $L^p(S, \mu)$ norms, $1 \leq p < \infty$, and $p_t f \in \mathcal{A}$ for all $f \in \mathcal{A}_0$

In addition, we assume for simplicity

(A3) $1 \in \mathcal{A}$

Remark. Condition (A0) implies that $\mathcal{A}$, and hence $\mathcal{A}_0$, is dense in $L^p(S, \mu)$ for all $p \in [1, \infty)$. In fact, if $g \in \mathcal{L}^q(S, \mu)$, $\frac{1}{q} + \frac{1}{p} = 1$, with $\int fg \, d\mu = 0$ for all $f \in \mathcal{A}$, then $g \, d\mu = 0$ by (A0) and hence $g = 0$ μ-a.e. Similarly as above, the conditions (A0), (A1') and (A2') imply that $(p_t)_{t\geq 0}$ induces a C_0 semigroup on $L^p(S, \mu)$ for all $p \in [1, \infty)$, and the generator $(L^{(p)}, \mathrm{Dom}(L^{(p)}))$ extends $(\mathcal{L}, \mathcal{A})$, i.e.,

$$\mathcal{A} \subseteq \mathrm{Dom}(L^{(p)}) \quad \text{and} \quad L^{(p)} f = \mathcal{L}f \quad \mu\text{-a.e. for all } f \in \mathcal{A}$$

In particular, the Kolmogorov forward equation

$$\frac{d}{dt} p_t f = p_t \mathcal{L} f \quad \forall f \in \mathcal{A}$$

and the backward equation

$$\frac{d}{dt} p_t f = \mathcal{L} p_t f \quad \forall f \in \mathcal{A}_0$$

hold with the derivative taken in the Banach space $L^p(S,\mu)$.

Example. (i) **Finite state space:** Suppose $\mu(x) > 0$ for all $x \in S$.

Generator:

$$(\mathcal{L}f)(x) = \sum_y \mathcal{L}(x,y) f(y) = \sum_y \mathcal{L}(x,y)(f(y) - f(x))$$

Adjoint:

$$\mathcal{L}^{*\mu}(y,x) = \frac{\mu(x)}{\mu(y)} \mathcal{L}(x,y)$$

Proof.

$$\begin{aligned}
(\mathcal{L}f,g)_\mu &= \sum_{x,y} \mu(x)\mathcal{L}(x,y) f(y) g(x) \\
&= \sum \mu(y) f(y) \frac{\mu(x)}{\mu(y)} \mathcal{L}(x,y) g(x) \\
&= (f, \mathcal{L}^{*\mu} g)_\mu
\end{aligned}$$

■

Symmetric part:

$$\begin{aligned}
\mathcal{L}_s(x,y) &= \frac{1}{2}\left(\mathcal{L}(x,y) + \mathcal{L}^{*\mu}(x,y)\right) = \frac{1}{2}\left(\mathcal{L}(x,y) + \frac{\mu(y)}{\mu(x)} \mathcal{L}(y,x)\right) \\
\mu(x)\mathcal{L}_s(x,y) &= \frac{1}{2}\left(\mu(x)\mathcal{L}(x,y) + \mu(y)\mathcal{L}(y,x)\right)
\end{aligned}$$

Dirichlet form:

$$\begin{aligned}
\mathcal{E}_s(f,g) = -(\mathcal{L}_s f, g) &= -\sum_{x,y} \mu(x)\mathcal{L}_s(x,y)\left(f(y) - f(x)\right) g(x) \\
&= -\sum_{x,y} \mu(y)\mathcal{L}_s(y,x)\left(f(x) - f(y)\right) g(y) \\
&= -\frac{1}{2}\sum \mu(x)\mathcal{L}_s(x,y)\left(f(y) - f(x)\right)\left(g(y) - g(x)\right)
\end{aligned}$$

Hence

$$\mathcal{E}(f,f) = \mathcal{E}_s(f,f) = \frac{1}{2}\sum_{x,y} Q(x,y)\left(f(y) - f(x)\right)^2$$

where

$$Q(x,y) = \mu(x)\mathcal{L}_s(x,y) = \frac{1}{2}\left(\mu(x)\mathcal{L}(x,y) + \mu(y)\mathcal{L}(y,x)\right)$$

(ii) **Diffusions in $\mathbb{R}^n$:** Let

$$\mathcal{L} = \frac{1}{2}\sum_{i,j} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + b \cdot \nabla,$$

and $\mathcal{A} = C_0^\infty$, $\mu = \varrho\, dx$, $\varrho, a_{ij} \in C^1$, $b \in C$ $\varrho \geq 0$,

$$\begin{aligned}
\mathcal{E}_s(f,g) &= \frac{1}{2}\int \sum_{i,j=1}^n a_{ij} \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\, d\mu \\
\mathcal{E}(f,g) &= \mathcal{E}_s(f,g) - (f, \beta \cdot \nabla g), \qquad \beta = b - \frac{1}{2\varrho} \div \left(\varrho a_{ij}\right)
\end{aligned}$$

Running example: Kolmogorov diffusions

9.1. Divergences and relaxation times

f-divergences

Definition 9.1 ("Distances" of probability measures). μ, ν probability measures on S, $\mu - \nu$ signed measure.

(i) **Total variation distance**:

$$\|\nu - \mu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{S}} |\nu(A) - \mu(A)|$$

(ii) χ^2**-divergence:**

$$\chi^2(\mu|\nu) = \begin{cases} \int \left(\frac{d\mu}{d\nu} - 1\right)^2 d\mu = \int \left(\frac{d\nu}{d\mu}\right)^2 d\mu - 1 & \text{if } \nu \ll \mu \\ +\infty & \text{else} \end{cases}$$

(iii) **Relative entropy (Kullback-Leibler divergence):**

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu}\, d\mu = \int \log \frac{d\nu}{d\mu}\, d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{else} \end{cases}$$

(where $0 \log 0 := 0$).

Remark. By Jensen's inequality,

$$H(\nu|\mu) \geq \int \frac{d\nu}{d\mu}\, d\mu \log \int \frac{d\nu}{d\mu}\, d\mu = 0$$

Properties of f-divergences

Bounds on the variation norm:

Lemma 9.2. *(i)*

$$\|\nu - \mu\|_{TV}^2 \leq \frac{1}{4} \chi^2(\nu|\mu)$$

(ii) Pinsker's inequality:

$$\|\nu - \mu\|_{TV}^2 \leq \frac{1}{2} H(\nu|\mu) \quad \forall\, \mu, \nu \in M_1(S)$$

Proof. If $\nu \not\ll \mu$, then $H(\nu|\mu) = \chi^2(\nu|\mu) = \infty$.
Now let $\nu \ll \mu$:

(i)

$$\|\nu - \mu\|_{\mathrm{TV}} = \frac{1}{2}\|\varrho - 1\|_{L^1(\mu)} \leq \frac{1}{2}\|\varrho - 1\|_{L^2(\mu)} = \frac{1}{2}\chi^2(\nu|\mu)^{\frac{1}{2}}$$

(ii) We have the inequality

$$3(x-1)^2 \le (4+2x)(x\log x - x + 1) \quad \forall\, x \ge 0$$

and hence

$$\sqrt{3}|x-1| \le (4+2x)^{\frac{1}{2}}(x\log x - x + 1)^{\frac{1}{2}}$$

and with the Cauchy Schwarz inequality

$$\begin{aligned}\sqrt{3}\int |\varrho - 1|\, d\mu &\le \left(\int (4+2\varrho)\, d\mu\right)^{\frac{1}{2}} \left(\int (\varrho\log\varrho - \varrho + 1)\, d\mu\right)^{\frac{1}{2}} \\ &= \sqrt{6}\cdot H(\nu|\mu)^{\frac{1}{2}}\end{aligned}$$

■

Remark. If S is finite and $\mu(x) > 0$ for all $x \in S$ then conversely

$$\begin{aligned}\chi^2(\nu|\mu) = \sum_{x\in S}\left(\frac{\nu(x)}{\mu(x)} - 1\right)^2 \mu(x) &\le \frac{\left(\sum_{x\in S}\left|\frac{\nu(x)}{\mu(x)} - 1\right|\mu(x)\right)^2}{\min_{x\in S}\mu(x)} \\ &= \frac{4\|\nu-\mu\|_{\mathrm{TV}}^2}{\min\mu}\end{aligned}$$

Variational characterizations

Lemma 9.3 (Variational characterizations).

(i)

$$\|\nu - \mu\| = \frac{1}{2}\sup_{\substack{f\in\mathcal{F}_b(S)\\ |f|\le 1}}\left(\int f\, d\nu - \int f\, d\mu\right)$$

(ii)

$$\chi^2(\nu|\mu) = \sup_{\substack{f\in\mathcal{F}_b(S)\\ \int f^2\, d\mu\le 1}}\left(\int f\, d\nu - \int f\, d\mu\right)^2$$

and by replacing f by $f - \int f\, d\mu$,

$$\chi^2(\nu|\mu) = \sup_{\substack{f\in\mathcal{F}_b(S)\\ \int f^2\, d\mu\le 1\\ \int f\, d\mu = 0}}\left(\int f\, d\nu\right)^2$$

(iii)

$$H(\nu|\mu) = \sup_{\substack{f\in\mathcal{F}_b(S)\\ \int e^f\, d\mu\le 1}}\int f\, d\nu = \sup_{f\in\mathcal{F}_b(S)}\int f\, d\nu - \log\int e^f\, d\mu$$

Remark. $\int e^f\, d\mu \le 1$, hence $\int f\, d\mu \le 0$ by Jensen and we also have

$$\sup_{\int e^f\, d\mu\le 1}\left(\int f\, d\nu - \int f\, d\mu\right) \le H(\nu|\mu)$$

Proof. (i)" $\leq$ "

$$\nu(A) - \mu(A) = \frac{1}{2}\left(\nu(A) - \mu(A) + \mu(A^c) - \nu(A^c)\right) = \frac{1}{2}\left(\int f\,d\nu - \int f\,d\mu\right)$$

and setting $f := I_A - I_{A^c}$ leads to

$$\|\nu - \mu\|_{\mathrm{TV}} = \sup_A \left(\nu(A) - \mu(A)\right) \leq \frac{1}{2}\sup_{|f|\leq 1}\left(\int f\,d\nu - \int f\,d\mu\right)$$

" $\geq$ " If $|f| \leq 1$ then

$$\begin{aligned}
\int f\,d(\nu-\mu) &= \int_{S_+} f\,d(\nu-\mu) + \int_{S_-} f\,d(\nu-\mu) \\
&\leq (\nu-\mu)(S_+) - (\nu-\mu)(S_-) \\
&= 2(\nu-\mu)(S_+) \qquad (\text{since } (\nu-\mu)(S_+) + (\nu-\mu)(S_-) = (\nu-\mu)(S) = 0) \\
&\leq 2\|\nu-\mu\|_{\mathrm{TV}}
\end{aligned}$$

where $S = S_+ \dot{\bigcup} S_-$, $\nu - \mu \geq 0$ on S_+, $\nu - \mu \leq 0$ on S_- is the Hahn-Jordan decomposition of the measure $\nu - \mu$.

(ii) If $\nu \ll \mu$ with density ϱ then

$$\chi^2(\nu|\mu)^{\frac{1}{2}} = \|\varrho - 1\|_{L^2(\mu)} = \sup_{\substack{f\in\mathcal{L}^2(\mu)\\ \|f\|_{L^2(\mu)}\leq 1}} \int f(\varrho-1)\,d\mu = \sup_{\substack{f\in\mathcal{F}_b(S)\\ \|f\|_{L^2(\mu)}\leq 1}} \left(\int f\,d\nu - \int f\,d\mu\right)$$

by the Cauchy-Schwarz inequality and a density argument.

If $\nu \not\ll \mu$ then there exists $A \in \mathcal{S}$ with $\mu(A) = 0$ and $\nu(A) \neq 0$. Choosing $f = \lambda \cdot I_A$ with $\lambda \uparrow \infty$ we see that

$$\sup_{\substack{f\in\mathcal{F}_b(S)\\ \|f\|_{L^2(\mu)}\leq 1}} \left(\int f\,d\nu - \int f\,d\mu\right)^2 = \infty = \chi^2(\nu|\mu).$$

This proves the first equation. The second equation follows by replacing f by $f - \int f\,d\mu$.

(iii) First equation:

" $\geq$ " By Young's inequality,

$$uv \leq u\log u - u + e^v$$

for all $u \geq 0$ and $v \in \mathbb{R}$, and hence for $\nu \ll \mu$ with density ϱ,

$$\begin{aligned}
\int f\,d\nu &= \int f\varrho\,d\mu \\
&\leq \int \varrho\log\varrho\,d\mu - \int \varrho\,d\mu + \int e^f\,d\mu \\
&= H(\nu|\mu) - 1 + \int e^f\,d\mu \qquad \forall f \in \mathcal{F}_b(S) \\
&\leq H(\nu|\mu) \qquad \text{if } \int e^f\,d\mu \leq 1
\end{aligned}$$

" ≤ " $\nu \ll \mu$ with density ϱ:

a) $\varepsilon \leq \varrho \leq \frac{1}{\varepsilon}$ for some $\varepsilon > 0$: Choosing $f = \log \varrho$ we have

$$H(\nu|\mu) = \int \log \varrho \, d\nu = \int f \, d\nu$$

and

$$\int e^f \, d\mu = \int \varrho \, d\mu = 1$$

b) General case by an approximation argument.

Second equation: cf. Deuschel, Stroock [12]. ■

Remark. If $\nu \ll \mu$ with density ϱ then

$$\|\nu - \mu\|_{\mathrm{TV}} = \frac{1}{2} \sup_{|f| \leq 1} \int f(\varrho - 1) \, d\mu = \frac{1}{2} \|\varrho - 1\|_{L^1(\mu)}$$

However, $\|\nu - \mu\|_{\mathrm{TV}}$ is finite even when $\nu \not\ll \mu$.

Relaxation times and mixing times

9.2. Poincaré inequalities and L^2 relaxation

We first restrict ourselves to the case $p = 2$. For $f, g \in \mathcal{L}^2(S, \mu)$ let

$$(f, g)_\mu = \int fg \, d\mu$$

denote the L^2 inner product.

Definition 9.4. The bilinear form

$$\mathcal{E}(f, g) := -(f, \mathcal{L}g)_\mu = -\frac{d}{dt}(f, p_t g)_\mu \Big|_{t=0},$$

$f, g \in \mathcal{A}$, is called the **Dirichlet form** associated to $(\mathcal{L}, \mathcal{A})$ on $L^2(\mu)$.

$$\mathcal{E}_s(f, g) := \frac{1}{2} \left(\mathcal{E}(f, g) + \mathcal{E}(g, f) \right)$$

is the **symmetrized Dirichlet form**.

Remark. More generally, $\mathcal{E}(f, g)$ is defined for all $f \in L^2(S, \mu)$ and $g \in \mathrm{Dom}(L^{(2)})$ by

$$\mathcal{E}(f, g) = -(f, L^{(2)}g)_\mu = -\frac{d}{dt}(f, p_t g)_\mu \Big|_{t=0}$$

Decay of variances and χ^2 divergences

Theorem 9.5. For all $f \in \mathcal{A}_0$ and $t \geq 0$

$$\frac{d}{dt} \operatorname{Var}_\mu(p_t f) = \frac{d}{dt} \int (p_t f)^2 \, d\mu = -2\mathcal{E}(p_t f, p_t f) = -2\mathcal{E}_s(p_t f, p_t f)$$

Remark. (i) In particular,

$$\mathcal{E}(f,f) = -\frac{1}{2} \int (p_t f)^2 \, d\mu = -\frac{1}{2} \frac{d}{dt} \operatorname{Var}_\mu(p_t f)\Big|_{t=0},$$

infinitesimal change of variance

(ii) The assertion extends to all $f \in \operatorname{Dom}(L^{(2)})$ if the Dirichlet form is defined with respect to the L^2 generator. In the symmetric case the assertion even holds for all $f \in L^2(S,\mu)$.

Proof. By the backward equation,

$$\frac{d}{dt} \int (p_t f)^2 \, d\mu = 2 \int p_t \mathcal{L} p_t f \, d\mu = -2\mathcal{E}(p_t f, p_t f) = -2\mathcal{E}_s(p_t f, p_t f)$$

Moreover, since

$$\int p_t f \, d\mu = \int f \, d(\mu p_t) = \int f \, d\mu$$

is constant,

$$\frac{d}{dt} \operatorname{Var}_\mu(p_t f) = \frac{d}{dt} \int (p_t f)^2 \, d\mu \qquad \blacksquare$$

Remark. (i) In particular,

$$\mathcal{E}(f,f) = -\frac{1}{2}\frac{d}{dt} \int (p_t f)^2 \, d\mu\Big|_{t=0} = -\frac{1}{2}\frac{d}{dt} \operatorname{Var}_\mu(p_t f)$$

$$\mathcal{E}_s(f,g) = \frac{1}{4}\left(\mathcal{E}_s(f+g,f+g) + \mathcal{E}_s(f-g,f-g)\right) = -\frac{1}{2}\frac{d}{dt} \operatorname{Cov}_\mu(p_t f, p_t g)$$

Dirichlet form = infinitesimal change of (co)variance.

(ii) Since p_t is a contraction on $\mathcal{L}^2(\mu)$, the operator $(\mathcal{L}, \mathcal{A})$ is negative-definite, and the bilinear form $(\mathcal{E}, \mathcal{A})$ is positive definite:

$$(-f, \mathcal{L} f)_\mu = \mathcal{E}(f,f) = -\frac{1}{2} \lim_{t \downarrow 0} \left(\int (p_t f)^2 \, d\mu - \int f^2 \, d\mu \right) \geq 0$$

Corollary 9.6 (Decay of variance). For $\lambda > 0$ the following assertions are equivalent:

(i) **Poincaré inequality**:

$$\operatorname{Var}_\mu(f) \leq \frac{1}{\lambda} \mathcal{E}_{(s)}(f,f) \quad \forall f \in \mathcal{A}$$

(ii) **Exponential decay of variance**:

$$\operatorname{Var}_\mu(p_t f) \leq e^{-2\lambda t} \operatorname{Var}_\mu(f) \quad \forall f \in L^2(S,\mu) \tag{9.1}$$

(iii) **Spectral gap**: (Only equivalent in reversible case!)

$$\operatorname{Real} \alpha \geq \lambda \quad \forall \alpha \in \operatorname{spec}\left(-L^{(2)}\Big|_{\operatorname{span}\{1\}^\perp} \right)$$

Remark. Optimizing over λ, the corollary says that (9.1) holds with

$$\lambda := \inf_{f\in\mathcal{A}} \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\mu(f)} = \inf_{\substack{f\in\mathcal{A}\\ f\perp 1 \text{ in } L^2(\mu)}} \frac{(f,-\mathcal{L}f)_\mu}{(f,f)_\mu}$$

Proof. $(i) \Rightarrow (ii)$

$$\mathcal{E}(f,f) \geq \lambda \cdot \mathrm{Var}_\mu(f) \quad \forall\, f \in \mathcal{A}$$

By the theorem above,

$$\frac{d}{dt}\mathrm{Var}_\mu(p_t f) = -2\mathcal{E}(p_t f, p_t f) \leq -2\lambda\,\mathrm{Var}_\mu(p_t f)$$

for all $t \geq 0$, $f \in \mathcal{A}_0$. Hence

$$\mathrm{Var}_\mu(p_t f) \leq e^{-2\lambda t}\,\mathrm{Var}_\mu(p_0 f) = e^{-2\lambda t}\,\mathrm{Var}_\mu(f)$$

for all $f \in \mathcal{A}_0$. Since the right hand side is continuous with respect to the $L^2(\mu)$ norm, and $\mathcal{A}_0$ is dense in $L^2(\mu)$ by (A0) and (A2), the inequality extends to all $f \in L^2(\mu)$.

$(ii) \Rightarrow (iii)$ For $f \in \mathrm{Dom}(L^{(2)})$,

$$\frac{d}{dt}\mathrm{Var}_\mu(p_t f)\Big|_{t=0} = -2\mathcal{E}(f,f).$$

Hence if (9.1) holds then

$$\mathrm{Var}_\mu(p_t f) \leq e^{-2\lambda t}\,\mathrm{Var}_\mu(f) \quad \forall\, t \geq 0$$

which is equivalent to

$$\mathrm{Var}_\mu(f) - 2t\mathcal{E}(f,f) + o(t) \leq \mathrm{Var}_\mu(f) - 2\lambda t\,\mathrm{Var}_\mu(f) + o(t) \quad \forall\, t \geq 0$$

Hence

$$\mathcal{E}(f,f) \geq \lambda\,\mathrm{Var}_\mu(f)$$

and thus

$$-(L^{(2)}f,f)_\mu \geq \lambda \int f^2\,d\mu \qquad \text{for } f\perp 1$$

which is equivalent to (iii) if reversibility holds.

$(iii) \Rightarrow (i)$ Follows by the equivalence above. ■

Remark. Since $(\mathcal{L},\mathcal{A})$ is negative definite, $\lambda \geq 0$. In order to obtain exponential decay, however, we need $\lambda > 0$, which is not always the case.

Corollary 9.7. The assertions $(i)-(iii)$ in the corollary above are also equivalent to

(iv) **Exponential decay of χ^2 divergence w.r.t. equilibrium measure:**

$$\chi^2(\nu p_t|\mu) \leq e^{-2\lambda t}\chi^2(\nu|\mu) \quad \forall\, \nu \in M_1(S)$$

Proof. We show $(ii) \Leftrightarrow (iv)$.

" $\Rightarrow$ " Let $f \in \mathcal{L}^2(\mu)$ with $\int f\, d\mu = 0$. Then

$$\begin{aligned}\int f\, d(\nu p_t) - \int f\, d\mu = \int f\, d(\nu p_t) &= \int p_t f\, d\nu \\ &\leq \|p_t f\|_{L^2(\mu)} \cdot \chi^2(\nu|\mu)^{\frac{1}{2}} \\ &\leq e^{-\lambda t} \|f\|_{L^2(\mu)} \cdot \chi^2(\nu|\mu)^{\frac{1}{2}}\end{aligned}$$

where we have used that $\int p_t f\, d\mu = \int f\, d\mu = 0$. By taking the supremum over all f with $\int f^2\, d\mu \leq 1$ we obtain

$$\chi^2(\nu p_t|\mu)^{\frac{1}{2}} \leq e^{-\lambda t} \chi^2(\nu|\mu)^{\frac{1}{2}}$$

" $\Leftarrow$ " For $f \in \mathcal{L}^2(\mu)$ with $\int f\, d\mu = 0$, (iv) implies

$$\begin{aligned}\int p_t f g\, d\mu \overset{\nu := g\mu}{=} \int f\, d(\nu p_t) &\leq \|f\|_{L^2(\mu)} \chi^2(\nu p_t|\mu)^{\frac{1}{2}} \\ &\leq e^{-\lambda t} \|f\|_{L^2(\mu)} \chi^2(\nu|\mu)^{\frac{1}{2}} \\ &= e^{-\lambda t} \|f\|_{L^2(\mu)} \|g\|_{L^2(\mu)}\end{aligned}$$

for all $g \in L^2(\mu), g \geq 0$. Hence

$$\|p_t f\|_{L^2(\mu)} \leq e^{-\lambda t} \|f\|_{L^2(\mu)}$$

■

Example (Gradient type diffusions in $\mathbb{R}^n$).

$$dX_t = dB_t + b(X_t)\, dt, \qquad b \in C(\mathbb{R}^n, \mathbb{R}^n)$$

Generator:

$$\mathcal{L} f = \frac{1}{2} \Delta f + b \nabla f, \quad f \in C_0^\infty(\mathbb{R}^n)$$

symmetric with respect to $\mu = \varrho\, dx$, $\quad \varrho \in C^1 \Leftrightarrow b = \frac{1}{2} \nabla \log \varrho$.
Corresponding Dirichlet form on $L^2(\varrho\, dx)$:

$$\mathcal{E}(f, g) = -\int \mathcal{L} f g \varrho\, dx = \frac{1}{2} \int \nabla f \nabla g \varrho\, dx$$

Poincaré inequality:

$$\mathrm{Var}_{\varrho\, dx}(f) \leq \frac{1}{2\lambda} \cdot \int |\nabla f|^2 \varrho\, dx$$

The one-dimensional case: $\quad n = 1, b = \frac{1}{2}(\log \varrho)'$ and hence

$$\varrho(x) = \text{const.}\ e^{\int_0^x 2b(y)\, dy}$$

e.g. $b(x) = -\alpha x, \varrho(x) = \text{const.}\ e^{-\alpha x^2}, \mu =$ Gauss measure.

Corollary 9.8. (i) If the Poincaré inequality

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\lambda} \mathcal{E}(f, f) \quad \forall f \in \mathcal{A}$$

holds then

$$\|\nu p_t - \mu\|_{\mathrm{TV}} \leq \frac{1}{2} e^{-\lambda t} \chi^2(\nu|\mu)^{\frac{1}{2}} \tag{9.2}$$

(ii) In particular, if S is finite then

$$\|\nu p_t - \mu\|_{\mathrm{TV}} \leq \frac{1}{\min_{x\in S}\mu(x)^{\frac{1}{2}}} e^{-\lambda t}\|\nu - \mu\|_{\mathrm{TV}}$$

where $\|\nu - \mu\|_{\mathrm{TV}} \leq 1$. This leads to a bound for the **Dobrushin coefficient** (contraction coefficient with respect to $\|\cdot\|_{\mathrm{TV}}$).

Proof.

$$\|\nu p_t - \mu\|_{\mathrm{TV}} \leq \frac{1}{2}\chi^2(\nu p_t|\mu)^{\frac{1}{2}} \leq \frac{1}{2}e^{-\lambda t}\chi^2(\nu|\mu)^{\frac{1}{2}} \leq \frac{2}{2}\frac{1}{\min\mu^{\frac{1}{2}}}e^{-\lambda t}\|\nu - \mu\|_{\mathrm{TV}}$$

if S is finite. ∎

Consequence: Total variation mixing time: $\varepsilon \in (0,1)$,

$$\begin{aligned}T_{\mathrm{mix}}(\varepsilon) &= \inf\{t \geq 0 \ : \ \|\nu p_t - \mu\|_{\mathrm{TV}} \leq \varepsilon \text{ for all } \nu \in M_1(S)\} \\ &\leq \frac{1}{\lambda}\log\frac{1}{\varepsilon} + \frac{1}{2\lambda}\log\frac{1}{\min\mu(x)}\end{aligned}$$

where the first summand is the **L^2 relaxation time** and the second is an upper bound for the **burn-in time**, i.e. the time needed to make up for a bad initial distribution.

Remark. On high or infinite-dimensional state spaces the bound (9.2) is often problematic since $\chi^2(\nu|\mu)$ can be very large (whereas $\|\nu - \mu\|_{\mathrm{TV}} \leq 1$). For example for product measures,

$$\chi^2(\nu^n|\mu^n) = \int\left(\frac{d\nu^n}{d\mu^n}\right)^2 d\mu^n - 1 = \left(\int\left(\frac{d\nu}{d\mu}\right)^2 d\mu\right)^n - 1$$

where $\int\left(\frac{d\nu}{d\mu}\right)^2 d\mu > 1$ grows exponentially in n.

Are there improved estimates?

$$\int p_t f\, d\nu - \int f\, d\mu = \int p_t f\, d(\nu - \mu) \leq \|p_t f\|_{\sup}\cdot\|\nu - \mu\|_{\mathrm{TV}}$$

Analysis: The Sobolev inequality implies

$$\|p_t f\|_{\sup} \leq c\cdot\|f\|_{L^p}$$

However, Sobolev constants are dimension dependent! This motivates a replacement by the log Sobolev inequality, see Section 9.4 below.

Spectral gap and relaxation time

Upper bounds for λ_P

The one-dimensional case

9.3. Lower bounds for spectral gaps

Factorization

Comparison

Isoperimetric bounds

Coupling bounds

9.4. Log Sobolev inequalities and entropy decay

We consider the setup from section 4.3. In addition, we now assume that $(\mathcal{L}, \mathcal{A})$ is symmetric on $L^2(S, \mu)$.

Decay of relative entropy

Theorem 9.9 (Exponential decay of relative entropy). (i) $H(\nu p_t|\mu) \leq H(\nu|\mu)$ for all $t \geq 0$ and $\nu \in M_1(S)$.

(ii) If a logarithmic Sobolev inequality with constant $\alpha > 0$ holds then

$$H(\nu p_t|\mu) \leq e^{-\frac{2}{\alpha}t} H(\nu|\mu)$$

Proof (Proof for gradient diffusions). $\mathcal{L} = \frac{1}{2}\Delta + b\nabla$, $b = \frac{1}{2}\nabla \log \varrho \in C(\mathbb{R}^n)$, $\mu = \varrho\, dx$ probability measure, $\mathcal{A}_0 = \operatorname{span}\{C_0^\infty(\mathbb{R}^n), 1\}$
. The Logarithmic Sobolev Inequality implies that

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}}\, d\mu \leq \frac{\alpha}{2} \int |\nabla f|^2\, d\mu = \alpha \mathcal{E}(f, f)$$

(i) Suppose $\nu = g \cdot \mu$, $0 < \varepsilon \leq g \leq \frac{1}{\varepsilon}$ for some $\varepsilon > 0$. Hence $\nu p_t \ll \mu$ with density $p_t g$, $\varepsilon \leq p_t g \leq \frac{1}{\varepsilon}$ (since $\int f\, d(\nu p_t) = \int p_t f\, d\nu = \int p_t f g d\mu = \int f p_t g\, d\mu$ by symmetry). This implies that

$$\frac{d}{dt} H(\nu p_t|\mu) = \frac{d}{dt} \int p_t g \log p_t g\, d\mu = \int \mathcal{L} p_t g (1 + \log p_t g)\, d\mu$$

by Kolmogorov and since $(x \log x)' = 1 + \log x$. We get

$$\frac{d}{dt} H(\nu p_t|\mu) = -\mathcal{E}(p_t g, \log p_t g) = -\frac{1}{2} \int \nabla p_t g \cdot \nabla \log p_t g\, d\mu$$

where $\nabla \log p_t g = \frac{\nabla p_t g}{p_t g}$. Hence

$$\frac{d}{dt} H(\nu p_t|\mu) = -2 \int \left|\nabla \sqrt{p_t g}\right|^2\, d\mu \tag{9.3}$$

(i) $-2 \int \left|\nabla \sqrt{p_t g}\right|^2\, d\mu \leq 0$

(ii) The Logarithmic Sobolev Inequality yields that

$$-2\int \left|\nabla\sqrt{p_t g}\right|^2 \, d\mu \le -\frac{4}{\alpha}\int p_t g \log \frac{p_t g}{\int p_t g \, d\mu} \, d\mu$$

where $\int p_t g \, d\mu = \int g \, d\mu = 1$ and hence

$$-2\int \left|\nabla\sqrt{p_t g}\right|^2 \, d\mu \le -\frac{4}{\alpha} H(\nu p_t|\mu)$$

(ii) Now for a general ν. If $\nu \not\ll \mu$, $H(\nu|\mu) = \infty$ and we have the assertion. Let $\nu = g \cdot \mu$, $g \in L^1(\mu)$ and

$$\begin{aligned} g_{a,b} &:= (g \vee a) \wedge b, \quad 0 < a < b, \\ \nu_{a,b} &:= g_{a,b} \cdot \mu. \end{aligned}$$

Then by (i),

$$H(\nu_{a,b} p_t|\mu) \le e^{-\frac{2t}{\alpha}} H(\nu_{a,b}|\mu)$$

The claim now follows for $a \downarrow 0$ and $b \uparrow \infty$ by dominated and monotone convergence. ■

Remark. (i) The proof in the general case is analogous, just replace (9.3) by inequality

$$4\mathcal{E}(\sqrt{f}, \sqrt{f}) \le \mathcal{E}(f, \log f)$$

(ii) An advantage of the entropy over the χ^2 distance is the good behavior in high dimensions. E.g. for product measures,

$$H(\nu^d|\mu^d) = d \cdot H(\nu|\mu)$$

grows only linearly in dimension.

Corollary 9.10 (Total variation bound). For all $t \ge 0$ and $\nu \in M_1(S)$,

$$\begin{aligned} \|\nu p_t - \mu\|_{\mathrm{TV}} &\le \frac{1}{\sqrt{2}} e^{-\frac{t}{\alpha}} H(\nu|\mu)^{\frac{1}{2}} \\ &\left(\le \frac{1}{\sqrt{2}} \log \frac{1}{\min \mu(x)} e^{-\frac{t}{\alpha}} \quad \text{if } S \text{ is finite} \right) \end{aligned}$$

Proof.

$$\|\nu p_t - \mu\|_{\mathrm{TV}} \le \frac{1}{\sqrt{2}} H(\nu p_t|\mu)^{\frac{1}{2}} \le \frac{1}{\sqrt{2}} e^{-\frac{t}{\alpha}} H(\nu|\mu)^{\frac{1}{2}}$$

where we use Pinsker's Theorem for the first inequality and Theorem 9.9 for the second inequality. Since S is finite,

$$H(\delta_x|\mu) = \log \frac{1}{\mu(x)} \le \log \frac{1}{\min \mu} \quad \forall\, x \in S$$

which leads to

$$H(\nu|\mu) \le \sum \nu(x) H(\delta_x|\mu) \le \log \frac{1}{\min \mu} \quad \forall\, \nu$$

since $\nu = \sum \nu(x)\delta_x$ is a convex combination. ■

Consequence for mixing time: (S finite)

$$\begin{aligned} T_{\mathrm{mix}}(\varepsilon) &= \inf\{t \ge 0 \;:\; \|\nu p_t - \mu\|_{\mathrm{TV}} \le \varepsilon \text{ for all } \nu \in M_1(S)\} \\ &\le \alpha \cdot \log \frac{1}{\sqrt{2}\varepsilon} + \log\log \frac{1}{\min_{x \in S} \mu(x)} \end{aligned}$$

Hence we have log log instead of log !

Extensions

Log Sobolev implies Poincaré

Theorem 9.11 (Rothaus). A logarithmic Sobolev inequality with constant α implies a Poincaré inequality with constant $\lambda = \frac{2}{\alpha}$.

Proof. Apply the logarithmic Sobolev-inequality to $f = 1 + \varepsilon g$ where $\int g d\mu = 0$. Then consider the limit $\varepsilon \to 0$ and use that $x \log x = x - 1 + \frac{1}{2}(x-1)^2 + \mathrm{O}(|x-1|^3)$. ■

Hypercontractivity

Theorem 9.12. With assumptions (A0)-(A3) and $\alpha > 0$, the following statements are equivalent:

(i) **Logarithmic Sobolev inequality (LSI)**

$$\int_S f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}} \, d\mu \le 2\alpha \mathcal{E}(f,f) \quad \forall f \in \mathcal{A}$$

(ii) **Hypercontractivity** For $1 \le p < q < \infty$,

$$\|p_t f\|_{L^q(\mu)} \le \|f\|_{L^p(\mu)} \quad \forall f \in L^p(\mu),\ t \ge \frac{\alpha}{2} \log \frac{q-1}{p-1}$$

(iii) Assertion (ii) holds for $p = 2$.

Remark. Hypercontractivity and Spectral gap implies

$$\|p_t f\|_{L^q(\mu)} = \left\|p_{t_0} p_{t-t_0} f\right\|_{L^q(\mu)} \le \left\|p_{t-t_0} f\right\|_{L^2(\mu)} \le e^{-\lambda(t-t_0)} \|f\|_{L^2(\mu)}$$

for all $t \ge t_0(q) := \frac{\alpha}{4} \log(q-1)$.

Proof. (i)⇒(ii). Idea: WLOG $f \in \mathcal{A}_0$, $f \ge \delta > 0$ (which implies that $p_t f \ge \delta \ \forall t \ge 0$).
Compute

$$\frac{d}{dt} \|p_t f\|_{L^{q(t)}(\mu)}, \quad q \colon \mathbb{R}^+ \to (1,\infty) \text{ smooth:}$$

(i) Kolmogorov:

$$\frac{d}{dt} p_t f = \mathcal{L} p_t f \quad \text{derivation with respect to sup-norm}$$

implies that

$$\frac{d}{dt} \int (p_t f)^{q(t)} \, d\mu = q(t) \int (p_t f)^{q(t)-1} \mathcal{L} p_t f \, d\mu + q'(t) \int (p_t f)^{q(t)} \log p_t f \, d\mu$$

where

$$\int (p_t f)^{q(t)-1} \mathcal{L} p_t f \, d\mu = -\mathcal{E}\left((p_t f)^{q(t)-1}, p_t f\right)$$

(ii) Stroock estimate:

$$\mathcal{E}\left(f^{q-1}, f\right) \geq \frac{4(q-1)}{q^2}\mathcal{E}\left(f^{\frac{q}{2}}, f^{\frac{q}{2}}\right)$$

Proof.

$$\begin{aligned}
\mathcal{E}(f^{q-1}, f) &= -\left(f^{q-1}, \mathcal{L}f\right)_\mu = \lim_{t\downarrow 0} \frac{1}{t}\left(f^{q-1}, f - p_t f\right)_\mu \\
&= \lim_{t\downarrow 0} \frac{1}{2t} \iint \left(f^{q-1}(y) - f^{q-1}(x)\right)(f(y) - f(x))\, p_t(x, dy)\, \mu(dx) \\
&\geq \frac{4(q-1)}{q^2} \lim_{t\downarrow 0} \frac{1}{2t} \iint \left(f^{\frac{q}{2}}(y) - f\frac{q}{2}(x)\right)^2 p_t(x, dy)\, \mu(dx) \\
&= \frac{4(q-1)}{q^2}\mathcal{E}\left(f^{\frac{q}{2}}, f^{\frac{q}{2}}\right)
\end{aligned}$$

where we have used that

$$\left(a^{\frac{q}{2}} - b\frac{q}{2}\right)^2 \leq \frac{q^2}{4(q-1)}\left(a^{q-1} - b^{q-1}\right)(a-b) \quad \forall\, a, b > 0,\ q \geq 1$$

Remark.

- The estimate justifies the use of functional inequalities with respect to $\mathcal{E}$ to bound L^p norms.
- For generators of diffusions, equality holds, e.g.:

$$\int \nabla f^{q-1} \nabla f\, d\mu = \frac{4(q-1)}{q^2} \int \left|\nabla f^{\frac{q}{2}}\right|^2 d\mu$$

  by the chain rule.

(iii) Combining the estimates:

$$q(t) \cdot \|p_t f\|_{q(t)}^{q(t)-1} \frac{d}{dt}\|p_t f\|_{q(t)} = \frac{d}{dt}\int (p_t f)^{q(t)}\, d\mu - q'(t)\int (p_t f)^{q(t)} \log \|p_t f\|_{q(t)}\, d\mu$$

where

$$\int (p_t f)^{q(t)}\, d\mu = \|p_t f\|_{q(t)}^{q(t)}$$

This leads to the estimate

$$\begin{aligned}
&q(t) \cdot \|p_t f\|_{q(t)}^{q(t)-1} \frac{d}{dt}\|p_t f\|_{q(t)} \\
&\leq -\frac{4(q(t)-1)}{q(t)}\mathcal{E}\left((p_t f)^{\frac{q(t)}{2}}, (p_t f)^{\frac{q(t)}{2}}\right) + \frac{q'(t)}{q(t)} \cdot \int (p_t f)^{q(t)} \log \frac{(p_t f)^{q(t)}}{\int (p_t f)^{q(t)}\, d\mu}\, d\mu
\end{aligned}$$

(iv) Applying the logarithmic Sobolev inequality: Fix $p \in (1, \infty)$. Choose $q(t)$ such that

$$\alpha q'(t) = 2(q(t) - 1), \quad q(0) = p$$

i.e.

$$q(t) = 1 + (p-1)e^{\frac{2t}{\alpha}}$$

Then by the logarithmic Sobolev inequality, the right hand side in the estimate above is negative, and hence $\|p_t f\|_{q(t)}$ is decreasing. Thus

$$\|p_t f\|_{q(t)} \leq \|f\|_{q(0)} = \|f\|_p \quad \forall\, t \geq 0.$$

Other implication: Exercise. (Hint: consider $\frac{d}{dt}\|p_t f\|_{L^{q(t)}(\mu)}$). ■

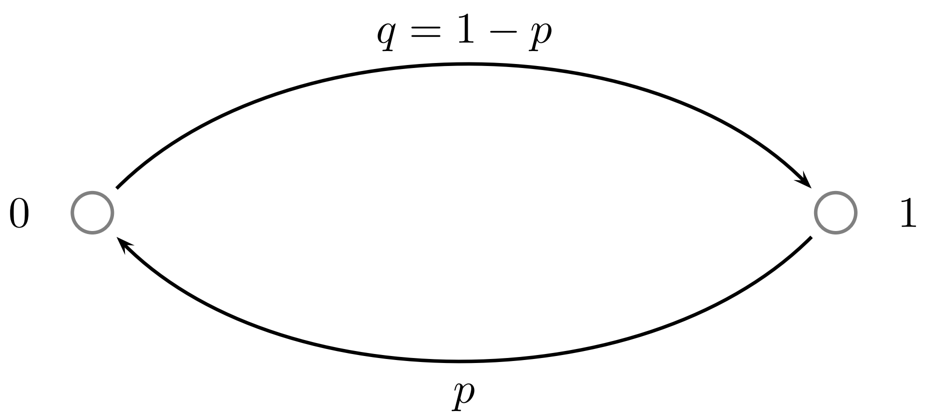
Basic examples

Example. Two-point space. $S = \{0, 1\}$. Consider a Markov chain with generator

$$\mathcal{L} = \begin{pmatrix} -q & q \\ p & -p \end{pmatrix}, \qquad\qquad p, q \in (0,1),\ p + q = 1$$

which is symmetric with respect to the Bernoulli measure,

$$\mu(0) = p, \quad \mu(1) = q$$



Dirichlet form:

$$\begin{aligned}\mathcal{E}(f,f) &= \frac{1}{2}\sum_{x,y} (f(y) - f(x))^2\, \mu(x)\mathcal{L}(x,y) \\ &= pq \cdot |f(1) - f(0)|^2 = \mathrm{Var}_\mu(f)\end{aligned}$$

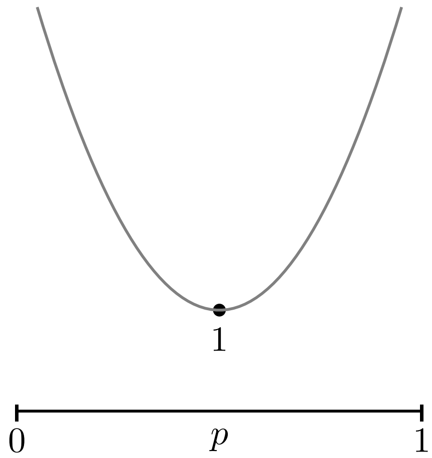
Spectral gap:

$$\lambda(p) = \inf_{f \text{not const.}} \frac{\mathcal{E}(f,f)}{\mathrm{Var}_\mu(f)} = 1 \quad \text{independent of } p\,!$$

Optimal Log Sobolev constant:

$$\alpha(p) = \sup_{\substack{f \perp 1 \\ \int f^2\, d\mu = 1}} \frac{\int f^2 \log f^2\, d\mu}{2\mathcal{E}(f,f)} = \begin{cases} 1 & \text{if } p = \frac{1}{2} \\ \frac{1}{2}\frac{\log q - \log p}{q - p} & \text{else} \end{cases}$$

goes to infinity as $p \downarrow 0$ or $p \uparrow \infty$!



9.5. Derivation of Log Sobolev inequalities

Factorization

Spectral gap and Logarithmic Sobolev Inequality for product measures:

$$\mathrm{Ent}_\mu(f) := \int f \log f \, d\mu, \; f > 0$$

Theorem 9.13 (Factorization property). $(S_i, \mathcal{S}_i, \mu_i)$ probability spaces, $\mu = \otimes_{i=1}^n \mu_i$. Then

(i)

$$\mathrm{Var}_\mu(f) \leq \sum_{i=1}^n E_\mu \left[\mathrm{Var}_{\mu_i}^{(i)}(f) \right]$$

where on the right hand side the variance is taken with respect to the i-th variable.

(ii)

$$\mathrm{Ent}_\mu(f) \leq \sum_{i=1}^n E_\mu \left[\mathrm{Ent}_{\mu_i}^{(i)}(f) \right]$$

Proof. (i) Exercise.

(ii)

$$\mathrm{Ent}_\mu(f) = \sup_{g \,:\, E_\mu[e^g]=1} E_\mu[fg], \quad \text{cf. above}$$

Fix $g\colon S^n \to \mathbb{R}$ such that $E_\mu\left[e^g\right] = 1$. Decompose:

$$\begin{aligned}g(x_1,\dots,x_n) &= \log e^{g(x_1,\dots,x_n)}\\ &= \log \frac{e^{g(x_1,\dots,x_n)}}{\int e^{g(y_1,x_2,\dots,x_n)}\,\mu_1(dy_1)} + \log \frac{\int e^{g(y_1,x_2,\dots,x_n)}\,\mu_1(dy_1)}{\iint e^{g(y_1,y_2,x_3,\dots,x_n)}\,\mu_1(dy_1)\mu_2(dy_2)} + \cdots\\ &=: \sum_{i=1}^n g_i(x_1,\dots,x_n)\end{aligned}$$

and hence

$$\begin{aligned}& E^i_{\mu_i}\left[e^{g_i}\right] = 1 \quad \forall, 1 \le i \le n\\ \Rightarrow\quad & E_\mu[fg] = \sum_{i=1}^n E_\mu\left[fg_i\right] = \sum_{i=1}^n E_\mu\left[E^{(i)}_{\mu_i}\left[fg_i\right]\right] \le \mathrm{Ent}^{(i)}_{\mu_i}(f)\\ \Rightarrow\quad & \mathrm{Ent}_\mu[f] = \sup_{E_\mu[e^g]=1} E_\mu[fg] \le \sum_{i=1}^n E_\mu\left[\mathrm{Ent}^{(i)}_{\mu_i}(f)\right]\end{aligned}$$

■

Corollary 9.14. (i) If the Poincaré inequalities

$$\mathrm{Var}_{\mu_i}(f) \le \frac{1}{\lambda_i}\mathcal{E}_i(f,f) \quad \forall f \in \mathcal{A}_i$$

hold for each μ_i then

$$\mathrm{Var}_\mu(f) \le \frac{1}{\lambda}\mathcal{E}(f,f) \quad \forall f \in \bigotimes_{i=1}^n \mathcal{A}_i$$

where

$$\mathcal{E}(f,f) = \sum_{i=1}^n E_\mu\left[\mathcal{E}^{(i)}_i(f,f)\right]$$

and

$$\lambda = \min_{1\le i\le n} \lambda_i$$

(ii) The corresponding assertion holds for Logarithmic Sobolev Inequalities with $\alpha = \max \alpha_i$

Proof.

$$\mathrm{Var}_\mu(f) \le \sum_{i=1}^n E_\mu\left[\mathrm{Var}^{(i)}_{\mu_i}(f)\right] \le \frac{1}{\min \lambda_i}\mathcal{E}(f,f)$$

since

$$\mathrm{Var}^{(i)}_{\mu_i}(f) \le \frac{1}{\lambda_i}\mathcal{E}_i(f,f)$$

■

Example. $S = \{0,1\}^n$, μ^n product of Bernoulli(p),

$$\begin{aligned}&\mathrm{Ent}_{\mu^n}(f^2)\\ &\le 2\alpha(p)\cdot p\cdot q\cdot \sum_{i=1}^n \int \left|f(x_1,\dots,x_{i-1},1,x_{i+1},\dots,x_n) - f(x_1,\dots,x_{i-1},0,x_{i+1},\dots,x_n)\right|^2\,\mu^n(dx)\end{aligned}$$

independent of n.

Example. Standard normal distribution $\gamma = N(0,1)$,

$$\phi_n \colon \{0,1\}^n \to \mathbb{R}, \qquad \phi_n(x) = \frac{\sum_{i=1}^n \left(x_i - \frac{1}{2}\right)}{\sqrt{\frac{n}{4}}}$$

The Central Limit Theorem yields that $\mu = \text{Bernoulli}(\frac{1}{2})$ and hence

$$\mu^n \circ \phi_n^{-1} \xrightarrow{w} \gamma$$

Hence for all $f \in C_0^\infty(\mathbb{R})$,

$$\begin{aligned}
\mathrm{Ent}_\gamma(f^2) &= \lim_{n\to\infty} \mathrm{Ent}_{\mu^n}(f^2 \circ \phi_n) \\
&\le \liminf \frac{1}{2} \sum_{i=1}^n \int |\Delta_i f \circ \phi_n|^2 \; d\mu^n \\
&\le \cdots \le 2 \cdot \int |f'|^2 \, d\gamma
\end{aligned}$$

Comparison

Theorem 9.15 (Bounded perturbations). $\mu, \nu \in M_1(\mathbb{R}^n)$ absolut continuous,

$$\frac{d\nu}{d\mu}(x) = \frac{1}{Z} e^{-U(x)}.$$

If

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}} \, d\mu \le 2\alpha \cdot \int |\nabla f|^2 \, d\mu \quad \forall\, f \in C_0^\infty$$

then

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\nu)}} \, d\nu \le 2\alpha \cdot e^{\mathrm{osc}(U)} \cdot \int |\nabla f|^2 \, d\nu \quad \forall\, f \in C_0^\infty$$

where

$$\mathrm{osc}(U) := \sup U - \inf U$$

Proof.

$$\int f^2 \log \frac{|f|^2}{\|f\|^2_{L^2(\nu)}} \, d\nu \le \int \left(f^2 \log f^2 - f^2 \log \|f\|^2_{L^2(\mu)} - f^2 + \|f\|^2_{L^2(\mu)} \right) d\nu \tag{9.4}$$

since

$$\int f^2 \log \frac{|f|^2}{\|f\|^2_{L^2(\nu)}} \, d\nu \le \int f^2 \log f^2 - f^2 \log t^2 - f^2 + t^2 \, d\nu \quad \forall\, t > 0$$

Note that in (9.4) the integrand on the right hand side is non-negative. Hence

$$
\begin{aligned}
\int f^2 \log \frac{|f|^2}{\|f\|^2_{L^2(\nu)}}\, d\nu &\leq \frac{1}{Z} \cdot e^{-\inf U} \int \left(f^2 \log f^2 - f^2 \log \|f\|^2_{L^2(\mu)} - f^2 + \|f\|^2_{L^2(\mu)} \right) d\mu \\
&= \frac{1}{Z} e^{-\inf U} \cdot \int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}}\, d\mu \\
&\leq \frac{2}{Z} \cdot e^{-\inf U} \alpha \int |\nabla f|^2\, d\mu \\
&\leq 2 e^{\sup U - \inf U} \alpha \int |\nabla f|^2\, d\nu
\end{aligned}
$$

■

Example. We consider the Gibbs measures μ from the example above

(i) **No interactions:**

$$H(x) = \sum_{i \in \Lambda} \left(\frac{x_i^2}{2} + V(x_i) \right), \qquad V : \mathbb{R} \to \mathbb{R} \text{ bounded}$$

Hence

$$\mu = \bigotimes_{i \in \Lambda} \mu_V$$

where

$$\mu_V(dx) \propto e^{-V(x)} \gamma(dx)$$

and $\gamma(dx)$ is the standard normal distribution. Hence μ satisfies the logarithmic Sobolev inequality with constant

$$\alpha(\mu) = \alpha(\mu_V) \leq e^{\operatorname{osc}(V)} \alpha(\gamma) = e^{\operatorname{osc}(V)}$$

by the factorization property. Hence we have independence of dimension!

(ii) **Weak interactions:**

$$H(x) = \sum_{i \in \Lambda} \left(\frac{x_i^2}{2} + V(x_i) \right) - \vartheta \sum_{\substack{i,j \in \Lambda \\ |i-j|=1}} x_i x_j - \vartheta \sum_{\substack{i \in \Lambda \\ j \notin \Lambda \\ |i-j|=1}} x_i z_j,$$

$\vartheta \in \mathbb{R}$. One can show:

Theorem 9.16. If V is bounded then there exists $\beta > 0$ such that for $\vartheta \in [-\beta, \beta]$ a logarithmic Sobolev inequality with constant independent of λ holds.

The proof is based on the exponential decay of correlations $\operatorname{Cov}_\mu(x_i, x_j)$ for Gibbs measures.

(iii) **Discrete Ising model:** One can show that for $\beta < \beta_c$ a logarithmic Sobolev inequality holds on$\{-N, \ldots, N\}^d$ with constant of Order $O(N^2)$ independent of the boundary conditions, whereas for $\beta > \beta_c$ and periodic boundary conditions the spectral gap, and hence the log Sobolev constant, grows exponentially in N, cf. [???].

Isoperimetric bounds

LSI for log-concave probability measures

Stochastic gradient flow in $\mathbb{R}^n$:

$$dX_t = dB_t - (\nabla H)(X_t)\,dt, \quad H \in C^2(\mathbb{R}^n)$$

Generator:

$$\mathcal{L} = \frac{1}{2}\Delta - \nabla H \cdot \nabla$$
$$\mu(dx) = e^{-H(x)}\,dx \text{ satisfies } \mathcal{L}^*\mu = 0$$

Assumption: There exists a $\kappa > 0$ such that

$$\partial^2 H(x) \geq \kappa \cdot I \quad \forall\, x \in \mathbb{R}^n$$
$$\text{i.e.} \quad \partial^2_{\xi\xi} H \geq \kappa \cdot |\xi|^2 \quad \forall\, \xi \in \mathbb{R}^n$$

Remark. The assumption implies the inequalities

$$x \cdot \nabla H(x) \geq \kappa \cdot |x|^2 - c, \tag{9.5}$$
$$H(x) \geq \frac{\kappa}{2}|x|^2 - \widetilde{c} \tag{9.6}$$

with constants $c, \widetilde{c} \in \mathbb{R}$. By (9.5) and a Lyapunov argument it can be shown that X_t does not explode in finite time and that $p_t(\mathcal{A}_0) \subseteq \mathcal{A}$ where $\mathcal{A}_0 = \operatorname{span}\left(C_0^\infty(\mathbb{R}^n), 1\right)$, $\mathcal{A} = \operatorname{span}\left(\mathcal{S}(\mathbb{R}^n), 1\right)$. By (9.6), the measure μ is finite, hence by our results above, the normalized measure is a stationary distribution for p_t.

Lemma 9.17. *If* $\operatorname{Hess} H \geq \kappa I$ *then*

$$|\nabla p_t f| \leq e^{-\kappa t} p_t\,|\nabla f| \quad f \in C_b^1(\mathbb{R}^n)$$

Remark. (i) Actually, both statements are equivalent.

(ii) If we replace $\mathbb{R}^n$ by an arbitrary Riemannian manifold the same assertion holds under the assumption

$$\operatorname{Ric} + \operatorname{Hess} H \geq \kappa \cdot I$$

(Bochner-Lichnerowicz-Weitzenböck).

Proof (Informal analytic proof:).

$$\begin{aligned}
\nabla \mathcal{L} f &= \nabla\left(\Delta - \nabla H \cdot \nabla\right) f \\
&= \left(\Delta - \nabla H \cdot \nabla - \partial^2 H\right) \nabla f \\
&=: \vec{\mathcal{L}} \quad \text{operator on one-forms (vector fields)}
\end{aligned}$$

This yields the evolution equation for $\nabla p_t f$:

$$\frac{\partial}{\partial t}\nabla p_t f = \nabla \frac{\partial}{\partial t} p_t f = \nabla \mathcal{L} p_t f = \vec{\mathcal{L}}\, \nabla p_t f$$

and hence

$$\begin{aligned}
\frac{\partial}{\partial t}\left|\nabla p_t f\right| &= \frac{\partial}{\partial t}\left(\nabla p_t f\cdot\nabla p_t f\right)^{\frac{1}{2}} = \frac{\left(\frac{\partial}{\partial t}\nabla p_t f\right)\cdot\nabla p_t f}{\left|\nabla p_t f\right|}\\
&= \frac{\left(\vec{\mathcal{L}}\,\nabla p_t f\right)\cdot\nabla p_t f}{\left|\nabla p_t f\right|} \leq \frac{\mathcal{L}\nabla p_t f\cdot\nabla p_t f}{\left|\nabla p_t f\right|} - \kappa\cdot\frac{\left|\nabla p_t f\right|^2}{\left|\nabla p_t f\right|}\\
&\leq \dots \leq \mathcal{L}\left|\nabla p_t f\right| - \kappa\left|\nabla p_t f\right|
\end{aligned}$$

We get that $v(t) := e^{\kappa t}p_{s-t}\left|\nabla p_t f\right|$ with $0 \leq t \leq s$ satisfies

$$v'(t) \leq \kappa v(t) - p_{s-t}\mathcal{L}\left|\nabla p_t f\right| + p_{s-t}\mathcal{L}\left|\nabla p_t f\right| - \kappa p_{s-t}\left|\nabla p_t f\right| = 0$$

and hence

$$e^{\kappa s}\left|\nabla p_s f\right| = v(s) \leq v(0) = p_s\left|\nabla f\right| \qquad \blacksquare$$

- The proof can be made rigorous by approximating $|\cdot|$ by a smooth function, and using regularity results for p_t, cf. e.g. Deuschel, Stroock[12].
- The assertion extends to general diffusion operators.

Proof (Probabilistic proof:). $p_t f(x) = E[f(X_t^x)]$ where X_t^x is the solution flow of the stochastic differential equation

$$\begin{aligned}
dX_t &= \sqrt{2}dB_t - (\nabla H)(X_t)\,dt, \quad \text{i.e.,}\\
X_t^x &= x + \sqrt{2}B_t - \int_0^t (\nabla H)(X_s^x)\,ds
\end{aligned}$$

By the assumption on H one can show that $x \longrightarrow X_t^x$ is smooth and the derivative flow $Y_t^x = \nabla_x X_t$ satisfies the differentiated stochastic differential equation

$$\begin{aligned}
dY_t^x &= -(\partial^2 H)(X_t^x)Y_t^x\,dt,\\
Y_0^x &= I
\end{aligned}$$

which is an ordinary differential equation. Hence if $\partial^2 H \geq \kappa I$ then for $v \in \mathbb{R}^n$,

$$\frac{d}{dt}\left|Y_t\cdot v\right|^2 = -2\left(Y_t\cdot v,\ (\partial^2 H)(X_t)Y_t\cdot v\right)_{\mathbb{R}^n} \leq 2\kappa\cdot\left|Y_t\cdot v\right|^2$$

where $Y_t\cdot v$ is the derivative of the flow in direction v. Hence

$$\begin{aligned}
&\left|Y_t\cdot v\right|^2 \leq e^{-2\kappa t}|v|\\
\Rightarrow\quad &\left|Y_t\cdot v\right| \leq e^{-\kappa t}|v|
\end{aligned}$$

This implies that for $f \in C_b^1(\mathbb{R}^n)$, $p_t f$ is differentiable and

$$\begin{aligned}
v\cdot\nabla p_t f(x) &= E\left[\left(\nabla f(X_t^x)\cdot Y_t^x\cdot v\right)\right]\\
&\leq E\left[\left|\nabla f(X_t^x)\right|\right]\cdot e^{-\kappa t}\cdot|v| \quad \forall\, v \in \mathbb{R}^n
\end{aligned}$$

i.e.

$$\left|\nabla p_t f(x)\right| \leq e^{-\kappa t}p_t\left|\nabla f\right|(x) \qquad \blacksquare$$

Theorem 9.18 (Bakry-Emery). Suppose that

$$\partial^2 H \geq \kappa \cdot I \quad \text{with } \kappa > 0$$

Then

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}} \, d\mu \leq \frac{2}{\kappa} \int |\nabla f|^2 \, d\mu \quad \forall \, f \in C_0^\infty(\mathbb{R}^n)$$

Remark. The inequality extends to $f \in H^{1,2}(\mu)$ where $H^{1,2}(\mu)$ is the closure of C_0^∞ with respect to the norm

$$\|f\|_{1,2} := \left(\int |f|^2 + |\nabla f|^2 \, d\mu \right)^{\frac{1}{2}}$$

Proof. $g \in \operatorname{span}(C_0^\infty, 1)$, $g \geq \delta \geq 0$.
Aim:

$$\int g \log g \, d\mu \leq \frac{1}{\kappa} \int \left|\nabla \sqrt{g}\right|^2 \, d\mu + \int g \, d\mu \log \int g \, d\mu$$

Then $g = f^2$ and we get the assertion.
Idea: Consider

$$u(t) = \int p_t g \log p_t g \, d\mu$$

Claim:

(i) $u(0) = \int g \log g \, d\mu$

(ii) $\lim_{t \uparrow \infty} u(t) = \int g \, d\mu \log \int g \, d\mu$

(iii) $-u'(t) \leq 4 e^{-2\kappa t} \int \left|\nabla \sqrt{g}\right|^2 \, d\mu$

By (i), (ii) and (iii) we then obtain:

$$\begin{aligned}
\int g \log g \, d\mu - \int g \, d\mu \log \int g \, d\mu &= \lim_{t \to \infty} (u(0) - u(t)) \\
&= \lim_{t \to \infty} \int_0^t -u'(t) \, ds \\
&\leq \frac{2}{\kappa} \int |\nabla \sqrt{g}|^2 \, d\mu
\end{aligned}$$

since $2 \int_0^\infty e^{-2\kappa s} \, ds = \frac{1}{\kappa}$.

Proof of claim:

(i) Obvious.

(ii) Ergodicity yields to

$$p_t g(x) \to \int g \, d\mu \quad \forall \, x$$

for $t \uparrow \infty$.

In fact:

$$|\nabla p_t g| \le e^{-\kappa t} p_t |\nabla g| \le e^{-\kappa t} |\nabla g|$$

and hence

$$|p_t g(x) - p_t g(y)| \le e^{-\kappa t} \sup |\nabla g| \cdot |x-y|$$

which leads to

$$\begin{aligned}\left| p_t g(x) - \int g \, d\mu \right| &= \left| \int (p_t g(x) - p_t g(y)) \; \mu(dy) \right| \\ &\le e^{-\kappa t} \sup |\nabla g| \cdot \int |x-y| \, \mu(dy) \to 0\end{aligned}$$

Since $p_t g \ge \delta \ge 0$, dominated convergence implies that

$$\int p_t g \log p_t \delta \, d\mu \to \int g \, d\mu \log \int g \, d\mu$$

(iii) **Key Step!** By the computation above (decay of entropy) and the lemma,

$$\begin{aligned}-u'(t) &= \int \nabla p_t g \cdot \nabla \log p_t g \, d\mu = \int \frac{|\nabla p_t g|^2}{p_t g} \, d\mu \\ &\le e^{-2\kappa t} \int \frac{(p_t |\nabla g|)^2}{p_t g} \, d\mu \le e^{-2\kappa t} \int p_t \frac{|\nabla g|^2}{g} \, d\mu \\ &= e^{-2\kappa t} \int \frac{|\nabla g|^2}{g} \, d\mu = 4 e^{-2\kappa t} \int \left| \nabla \sqrt{g} \right|^2 \, d\mu\end{aligned}$$

■

Example. An Ising model with real spin: (Reference: Royer [50])
$S = \mathbb{R}^\Lambda = \{(x_i)_{i \in \Lambda} \mid x_i \in \mathbb{R}\}$, $\Lambda \subset \mathbb{Z}^d$ finite.

$$\begin{aligned}\mu(dx) &= \frac{1}{Z} \exp(-H(x)) \, dx \\ H(x) &= \sum_{i \in \Lambda} \underbrace{V(x_i)}_{\text{potential}} - \frac{1}{2} \sum_{i,j \in \Lambda} \underbrace{\vartheta(i-j)}_{\text{interactions}} x_i x_j - \sum_{i \in \Lambda, j \in \mathbb{Z}^d \setminus \Lambda} \vartheta(i-j) x_i z_j,\end{aligned}$$

where $V \colon \mathbb{R} \to \mathbb{R}$ is a non-constant polynomial, bounded from below, and $\vartheta \colon \mathbb{Z} \to \mathbb{R}$ is a function such that $\vartheta(0) = 0$, $\vartheta(i) = \vartheta(-i) \; \forall \, i$, (symmetric interactions), $\vartheta(i) = 0 \; \forall \, |i| \ge R$ (finite range), $z \in \mathbb{R}^{\mathbb{Z}^d \setminus \Lambda}$ fixed boundary condition.

Glauber-Langevin dynamics:

$$dX_t^i = -\frac{\partial H}{\partial x_i}(X_t) \, dt + dB_t^i, \quad i \in \Lambda \tag{9.7}$$

Dirichlet form:

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{i \in \Lambda} \int \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_i} \, d\mu$$

Corollary 9.19. If

$$\inf_{x \in \mathbb{R}} V''(x) > \sum_{i \in \mathbb{Z}} |\vartheta(i)|$$

then $\mathcal{E}$ satisfies a log Sobolev inequality with constant independent of Λ.

Proof.

$$\frac{\partial^2 H}{\partial x_i \partial x_j}(x) = V''(x_i) \cdot \delta_{ij} - \vartheta(i-j)$$

$$\Rightarrow \quad \partial^2 H \geq \left(\inf V'' - \sum_i |\vartheta(i)| \right) \cdot I$$

in the sense of forms. ■

Consequence: There is a unique Gibbs measure on $\mathbb{Z}^d$ corresponding to H, cf. Royer [50]. What can be said if V is not convex?

9.6. Concentration of measure

$(\Omega, \mathcal{A}, P)$ probability space, $X_i : \Omega \to \mathbb{R}^d$ independent identically distributed, $\sim \mu$.
Law of large numbers:

$$\frac{1}{N} \sum_{i=1}^{N} U(X_i) \to \int U \, d\mu \qquad U \in \mathcal{L}^1(\mu)$$

Cramér:

$$P\left[\left| \frac{1}{N} \sum_{i=1}^{N} U(X_i) - \int U \, d\mu \right| \geq r \right] \leq 2 \cdot e^{-NI(r)},$$

$$I(r) = \sup_{t \in \mathbb{R}} \left(tr - \log \int e^{tU} \, d\mu \right) \quad \text{LD rate function.}$$

Hence we have

- Exponential concentration around mean value provided $I(r) > 0 \; \forall \, r \neq 0$
- $$P\left[\left| \frac{1}{N} \sum_{i=1}^{N} U(X_i) - \int U \, d\mu \right| \geq r \right] \leq e^{-\frac{Nr^2}{c}} \text{ provided } I(r) \geq \frac{r^2}{c}$$

  Gaussian concentration.

When does this hold? Extension to non independent identically distributed case? This leads to:
Bounds for $\log \int e^{tU} \, d\mu$!

Theorem 9.20 (Herbst). If μ satisfies a logarithmic Sobolev inequality with constant α then for any function $U \in C_b^1(\mathbb{R}^d)$ with $\|U\|_{\text{Lip}} \leq 1$:

(i)

$$\frac{1}{t} \log \int e^{tU} \, d\mu \leq \frac{\alpha}{2} t + \int U \, d\mu \quad \forall \, t > 0 \tag{9.8}$$

where $\frac{1}{t} \log \int e^{tU} \, d\mu$ can be seen as the **free energy at inverse temperature** t, $\frac{\alpha}{2}$ as a **bound for entropy** and $\int U \, d\mu$ as the **average energy**.

(ii)

$$\mu\left(U \geq \int U \, d\mu + r\right) \leq e^{-\frac{r^2}{2\alpha}}$$

Gaussian concentration inequality

In particular,

(iii)

$$\int e^{\gamma|x|^2} \, d\mu < \infty \quad \forall \gamma < \frac{1}{2\alpha}$$

Remark. Statistical mechanics:

$$F_t = t \cdot S - \langle U \rangle$$

where F_t is the **free energy**, t the **inverse temperature**, S the **entropy** and $\langle U \rangle$ the **potential**.

Proof. WLOG, $0 \leq \varepsilon \leq U \leq \frac{1}{\varepsilon}$. Logarithmic Sobolev inequality applied to $f = e^{\frac{tU}{2}}$:

$$\int tUe^{tU} \, d\mu \leq 2\alpha \int \left(\frac{t}{2}\right)^2 |\nabla U|^2 \, e^{tU} \, d\mu + \int e^{tU} \, d\mu \log \int e^{tU} \, d\mu$$

For $\Lambda(t) := \log \int e^{tU} \, d\mu$ this implies

$$t\Lambda'(t) = \frac{\int tUe^{tU} \, d\mu}{\int e^{tU} \, d\mu} \leq \frac{\alpha t^2}{2} \frac{\int |\nabla U|^2 e^{tU} \, d\mu}{\int e^{tU} \, d\mu} + \Lambda(t) \leq \frac{\alpha t^2}{2} + \Lambda(t)$$

since $|\nabla U| \leq 1$. Hence

$$\frac{d}{dt} \frac{\Lambda(t)}{t} = \frac{t\Lambda'(t) - \Lambda(t)}{t^2} \leq \frac{\alpha}{2} \quad \forall t > 0$$

Since

$$\Lambda(t) = \Lambda(0) + t \cdot \Lambda'(0) + O(t^2) = t \int U \, d\mu + O(t^2),$$

we obtain

$$\frac{\Lambda(t)}{t} \leq \int U \, d\mu + \frac{\alpha}{2} t,$$

i.e. (i).

(ii) follows from (i) by the Markov inequality, and (iii) follows from (ii) with $U(x) = |x|$. ■

Corollary 9.21 (Concentration of empirical measures). X_i independent identically distributed, $\sim \mu$. If μ satisfies a logarithmic Sobolev inequality with constant α then

$$P\left[\left|\frac{1}{N}\sum_{i=1}^{N} U(X_i) - E_\mu[U]\right| \geq r\right] \leq 2 \cdot e^{-\frac{Nr^2}{2\alpha}}$$

for any function $U \in C_b^1(\mathbb{R}^d)$ with $\|U\|_{\mathrm{Lip}} \leq 1$, $N \in \mathbb{N}$ and $r > 0$.

Proof. By the factorization property, μ^N satisfies a logarithmic Sobolev inequality with constant α as well. Now apply the theorem to

$$\widetilde{U}(x) := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} U(x_i)$$

noting that

$$\nabla \widetilde{U}(x_1, \ldots, x_n) = \frac{1}{\sqrt{N}} \begin{pmatrix} \nabla U(x_1) \\ \vdots \\ \nabla U(x_N) \end{pmatrix}$$

hence since U is Lipschitz,

$$\left| \nabla \widetilde{U}(x) \right| = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N} |\nabla U(x_i)|^2 \right)^{\frac{1}{2}} \leq 1$$

■

9.7. Central Limit Theorem for Markov processes in continuous time

When are stationary Markov processes in continuous time ergodic?
Let $(L, \mathrm{Dom}(L))$ denote the generator of $(p_t)_{t \geq 0}$ on $L^2(\mu)$.

Theorem 9.22. The following assertions are equivalent:

(i) P_μ is ergodic

(ii) $\ker L = \mathrm{span}\{1\}$, i.e.

$$h \in \mathcal{L}^2(\mu) \text{harmonic} \quad \Rightarrow \quad h = \text{const. } \mu\text{-a.s.}$$

(iii) p_t is μ-irreducible, i.e.

$$B \in \mathcal{S} \text{ such that } p_t 1_B = 1_B \quad \mu\text{-a.s. } \forall\, t \geq 0 \quad \Rightarrow \quad \mu(B) \in \{0, 1\}$$

If reversibility holds then (i)-(iii) are also equivalent to:

(iv) p_t is $L^2(\mu)$-ergodic, i.e.

$$\left\| p_t f - \int f \, d\mu \right\|_{L^2(\mu)} \to 0 \quad \forall\, f \in L^2(\mu)$$

CLT for continuous-time martingales

Let $(M_t)_{t \geq 0}$ be a continuous square-integrable $(\mathcal{F}_t)$ martingale where $(\mathcal{F}_t)$ is a filtration satisfying the usual conditions. Then M_t^2 is a submartingale and there exists a unique natural (e.g. continuous) increasing process $\langle M \rangle_t$ such that

$$M_t^2 = \text{martingale} + \langle M \rangle_t$$

(Doob-Meyer decomposition, cf. e.g. Karatzas, Shreve [28]).

Example. If N_t is a Poisson process then

$$M_t = N_t - \lambda t$$

is a martingale and

$$\langle M \rangle_t = \lambda t$$

almost sure.

Note: For discontinuous martingales, $\langle M \rangle_t$ is **not** the quadratic variation of the paths!

(X_t, P_μ) stationary Markov process, $L_L^{(2)}$, $L^{(1)}$ generator on $L^2(\mu)$, $L^1(\mu)$, $f \in \mathrm{Dom}(L^{(1)}) \supseteq \mathrm{Dom}(L^{(2)})$. Hence

$$f(X_t) = M_t^f + \int_0^t (L^{(1)} f)(X_s)\, ds \quad P_\mu\text{-a.s.}$$

and M^f is a martingale. For $f \in \mathrm{Dom}(L^{(2)})$ with $f^2 \in \mathrm{Dom}(L^{(1)})$,

$$\langle M^f \rangle_t = \int_0^t \Gamma(f,f)(X_s)\, ds \quad P_\mu\text{-a.s.}$$

where

$$\Gamma(f,g) = L^{(1)}(f \cdot g) - f L^2 g - g L^{(2)} f \in L^1(\mu)$$

is the **Carré du champ (square field) operator**.

Example. Diffusion in $\mathbb{R}^n$,

$$L = \frac{1}{2} \sum_{i,j} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + b(x) \cdot \nabla$$

Hence

$$\Gamma(f,g)(x) = \sum_{i,j} a_{ij}(x) \frac{\partial f}{\partial x_i}(x) \frac{\partial g}{\partial x_j}(x) = \left|\sigma^T(x) \nabla f(x)\right|_{\mathbb{R}^n}^2$$

for all $f, g \in C_0^\infty(\mathbb{R}^n)$. Results for gradient diffusions on $\mathbb{R}^n$ (e.g. criteria for log Sobolev) extend to general state spaces if $|\nabla f|^2$ is replaced by $\Gamma(f,g)$!

Connection to Dirichlet form:

$$\mathcal{E}(f,f) = -\int f L^{(2)} f\, d\mu + \underbrace{\left(\frac{1}{2}\int L^{(1)} f^2\, d\mu\right)}_{=0} = \frac{1}{2}\int \Gamma(f,f)\, d\mu$$

Theorem 9.23 (Central limit theorem for martingales). (M_t) square-integrable martingale on $(\Omega, \mathcal{F}, P)$ with stationary increments (i.e. $M_{t+s} - M_s \sim M_t - M_0$), $\sigma > 0$. If

$$\frac{1}{t}\langle M \rangle_t \to \sigma^2 \quad \text{in } L^1(P)$$

then

$$\frac{M_t}{\sqrt{t}} \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

CLT for Markov processes

Corollary 9.24 (Central limit theorem for Markov processes (elementary version)). Let (X_t, P_μ) be a stationary ergodic Markov process. Then for $f \in \text{range}(L)$, $f = Lg$:

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s)\, ds \xrightarrow{\mathcal{D}} N(0, \sigma_f^2)$$

where

$$\sigma_f^2 = 2 \int g(-L)g\, d\mu = 2\mathcal{E}(g,g)$$

Remark. (i) If μ is stationary then

$$\int f\, d\mu = \int Lg\, d\mu = 0$$

i.e. the random variables $f(X_s)$ are centered.

(ii) $\ker(L) = \text{span}\{1\}$ by ergodicity

$$(\ker L)^\perp = \left\{ f \in L^2(\mu) \;:\; \int f\, d\mu = 0 \right\} =: L_0^2(\mu)$$

If $L\colon L_0^2(\mu) \to L^2(\mu)$ is bijective with $G = (-L)^{-1}$ then the Central limit theorem holds for all $f \in L^2(\mu)$ with

$$\sigma_f^2 = 2(Gf, (-L)Gf)_{L^2(\mu)} = 2(f, Gf)_{L^2(\mu)}$$

(H^{-1} norm if symmetric).

Example. (X_t, P_μ) reversible, spectral gap λ, i.e.,

$$\text{spec}(-L) \subset \{0\} \cup [\lambda, \infty)$$

hence there is a $G = (-L\big|_{L_0^2(\mu)})^{-1}$, $\text{spec}(G) \subseteq [0, \frac{1}{\lambda}]$ and hence

$$\sigma_f^2 \leq \frac{2}{\lambda} \|f\|_{L^2(\mu)}^2$$

is a bound for asymptotic variance.

Proof (Proof of corollary.).

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s)\, ds = \frac{g(X_t) - g(X_0)}{\sqrt{t}} + \frac{M_t^g}{\sqrt{t}}$$

$$\langle M^g \rangle_t = \int_0^t \Gamma(g,g)(X_s)\, ds \quad P_\mu\text{-a.s.}$$

and hence by the ergodic theorem

$$\frac{1}{t} \langle M^g \rangle_t \overset{t\uparrow\infty}{\to} \int \Gamma(g,g)\, d\mu = \sigma_f^2$$

The central limit theorem for martingales gives

$$M_t^g \xrightarrow{\mathcal{D}} N(0, \sigma_f^2)$$

Moreover

$$\frac{1}{\sqrt{t}} \left(g(X_t) - g(X_0)\right) \to 0$$

in $L^2(P_\mu)$, hence in distribution. This gives the claim since

$$X_t \xrightarrow{\mathcal{D}} \mu, \quad Y_t \xrightarrow{\mathcal{D}} 0 \quad \Rightarrow \quad X_t + Y_t \xrightarrow{\mathcal{D}} \mu$$

■

Extension: range$(L) \neq L^2$, replace $-L$ by $\alpha - L$ (bijective), then $\alpha \downarrow 0$. Cf. Landim [32].

10. Beyond reversibility

10.1. Measure-preserving Markov processes

Infinitesimal invariance and stationarity

The next theorem gives a necessary and sometimes sufficient condition for invariance that can often be verified in concrete models:

Theorem 10.1 (Infinitesimal characterization of stationary distributions). For a probability measure μ on S, the following assertions are equivalent:

(i) The process (X_t, P_μ) is stationary, i.e., for any $s \geq 0$, $(X_{s+t})_{t\geq 0} \sim (X_t)_{t\geq 0}$ w.r.t. P_μ .

(ii) μ is an invariant probability measure for $(p_t)_{t\geq 0}$.

Moreover, (i) and (ii) imply

(iii) μ is **infinitesimally invariant**, i.e.,

$$\int \mathcal{L}f\, d\mu \;=\; 0 \qquad \text{for any } f \in \mathcal{A}.$$

Conversely, if Assumption (A2) is satisfied then (iii) implies (i) and (ii).

Proof. (i)⇒(ii) If (i) holds then for any $s \geq 0$,

$$\mu p_s \;=\; P_\mu \circ X_s^{-1} \;=\; P_\mu \circ X_0^{-1} \;=\; \mu.$$

(ii)⇒(i) By the Markov property, for any measurable subset $B \subseteq \mathcal{D}(\mathbb{R}^+, S)$,

$$P_\mu[(X_{s+t})_{t\geq 0} \in B \mid \mathcal{F}_s] \;=\; P_{X_s}[(X_t)_{t\geq 0} \in B] \quad P_\mu\text{-almost surely.}$$

Thus if (ii) holds then

$$P_\mu[(X_{s+t})_{t\geq 0} \in B] \;=\; E_\mu[P_{X_s}[(X_t)_{t\geq 0} \in B]] \;=\; P_{\mu p_s}[(X_t)_{t\geq 0} \in B] \;=\; P_\mu[X \in B]$$

(ii)⇒(iii) By Theorem 4.12, for any $f \in \mathcal{A}$,

$$\frac{p_t f - f}{t} \;\to\; \mathcal{L}f \qquad \text{uniformly as } t \downarrow 0.$$

Thus if μ is invariant for $(p_t)_{t\geq 0}$ then

$$\int \mathcal{L}f\, d\mu \;=\; \lim_{t\downarrow 0} \frac{\int (p_t f - f)\, d\mu}{t} \;=\; \lim_{t\downarrow 0} \frac{\int f\, d(\mu p_t) - \int f\, d\mu}{t} \;=\; 0.$$

(iii)⇒(ii) Suppose that (A2) holds, and let $f \in \mathcal{A}_0$. Then $p_t f \in \mathcal{A}$ for any $t \geq 0$. Hence, by the backward equation and (iii),

$$\frac{d}{dt}\int p_t f \, d\mu \;=\; \int \mathcal{L} p_t f \, d\mu \;=\; 0, \qquad \text{and thus}$$
$$\int f \, d(\mu p_t) \;=\; \int p_t f \, d\mu \;=\; \int f \, d\mu \qquad \text{for any } t \geq 0. \tag{10.1}$$

Since $\mathcal{A}_0$ is dense in $\mathcal{A}$ with respect to the supremum norm, (10.1) extends to all $f \in \mathcal{A}$. Hence by (A0), $\mu p_t = \mu$ for any $t \geq 0$. ■

Next, we apply Theorem 10.1 to diffusion processes on $\mathbb{R}^d$. Suppose that we are given non-explosive weak solutions $(X_t, P_x), x \in \mathbb{R}^d$, of a stochastic differential equation

$$dX_t = \sigma(X_t)\, dB_t + b(X_t)\, dt, \quad X_0 = x \quad P_x\text{-a.s.},$$

where $(B_t)_{t\geq 0}$ is a Brownian motion on $\mathbb{R}^d$, and the functions $\sigma\colon \mathbb{R}^n \to \mathbb{R}^{n\times d}$ and $b\colon \mathbb{R}^n \to \mathbb{R}^n$ are locally Lipschitz continuous. We assume that the diffusion coefficients $a(x) = \sigma(x)\sigma(x)^T$ and $b(x)$ are growing at most polynomially as $|x| \to \infty$. Then by Itô's formula, (X_t, P_x) solves the martingale problem for the operator $(\mathcal{L}, \mathcal{A})$ where

$$\mathcal{L} \;=\; \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x)\cdot\nabla, \qquad \text{and}$$

$$\mathcal{A} = \mathcal{S}(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n) : \lim_{|x|\to\infty} |x|^k (\partial^\alpha f)(x) = 0 \text{ for any } k > 0 \text{ and any multi-index } \alpha\}.$$

Moreover, the local Lipschitz condition implies uniqueness of strong solutions, and hence, by the Theorem of Yamade-Watanabe, uniqueness in distribution of weak solutions and uniqueness of the martingale problem for $(\mathcal{L}, \mathcal{A})$, cf. e.g. Rogers and Williams [49]. Therefore by Theorem 4.20, (X_t, P_x) is a Markov process.

Corollary 10.2 (Infinitesimal characterization of stationary distributions for diffusions on $\mathbb{R}^n$). Suppose that μ is an absolutely continuous probability measure on $\mathbb{R}^n$ with density $\varrho = d\mu/dx$. If μ is invariant for (X_t, P_x) then

$$\mathcal{L}^*\varrho \;:=\; \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}\varrho) - \div(b\varrho) \;=\; 0 \qquad \text{in the distributional sense.} \tag{10.2}$$

Conversely, if (10.2) holds and $p_t(C_0^\infty(\mathbb{R}^n)) \subseteq \mathcal{S}(\mathbb{R}^n)$ for any $t \geq 0$, then μ is invariant.

Proof. If μ is an invariant probability measure, then by Theorem 10.1,

$$0 \;=\; \int \mathcal{L} f \, d\mu = \int_{\mathbb{R}^n} \mathcal{L} f \;\varrho \, dx \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n), \tag{10.3}$$

i.e., $\mathcal{L}^*\varrho = 0$ holds in the distributional sense. Similarly, the converse implication follows by applying Theorem 10.1 with $\mathcal{A}_0 = C_0^\infty(\mathbb{R}^n)$ and $\mathcal{A} = \mathcal{S}(\mathbb{R}^n)$. ■

Remark. Even if μ is an invariant probability measure that is not absolutely continuous, it satisfies $\mathcal{L}^*\mu = 0$ in the distributional sense. This is precisely the statement of Theorem 10.1 (iii).

By Corollary 10.2, in order to show that a given measure is invariant for a diffusion process on $\mathbb{R}^n$, it is sufficient to verify that it is infinitesimally invariant and that $p_t(C_0^\infty(\mathbb{R}^n)) \subseteq \mathcal{S}(\mathbb{R}^n)$. The second condition has two parts: For a given test function $f \in C_0^\infty(\mathbb{R}^n)$ and $t \geq 0$, one has to check that $p_t f$ is smooth and that its derivatives are decaying faster than any polynomial as $|x| \to \infty$. This can be achieved by considering derivative flows if the coefficients are smooth, see the Master course on "Stochastic Analysis". Alternatively, even for non-smooth coefficients, smoothness of $p_t f$ follows by the regularity theory for partial differential equations if, for example, the operator $\mathcal{L}$ is strictly elliptic.

Example (Deterministic diffusions).

$$
\begin{aligned}
dX_t &= b(X_t)\,dt, && b \in C^2(\mathbb{R}^n) \\
\mathcal{L}f &= b \cdot \nabla f && \\
\mathcal{L}^*\varrho &= -\div(\varrho b) = -\varrho \div b - b \cdot \nabla \varrho, && \varrho \in C^1
\end{aligned}
$$

Lemma 10.3.

$$
\begin{aligned}
\mathcal{L}^*\varrho = 0 \quad &\Leftrightarrow \quad \div(\varrho b) = 0 \\
&\Leftrightarrow \quad (\mathcal{L}, C_0^\infty(\mathbb{R}^n)) \textit{ anti-symmetric on } L^2(\mu)
\end{aligned}
$$

Proof. First equivalence: cf. above

nd equivalence:

$$
\begin{aligned}
\int f\mathcal{L}g\,d\mu &= \int f b \cdot \nabla g \varrho\,dx = -\int \div(f b \varrho) g\,dx \\
&= -\int \mathcal{L}f g\,d\mu - \int \div(\varrho b) f g\,dx \qquad \forall f, g \in C_0^\infty
\end{aligned}
$$

Hence $\mathcal{L}$ is anti-symmetric if and only if $\div(\varrho b) = 0$ ■

Invariant measures of one-dimensional diffusions

In the one-dimensional case,

$$\mathcal{L}f = \frac{a}{2} f'' + b f',$$

and

$$\mathcal{L}^*\varrho = \frac{1}{2}(a\varrho)'' - (b\varrho)'$$

where $a(x) = \sigma(x)^2$. Assume $a(x) > 0$ for all $x \in \mathbb{R}$.

a) Harmonic functions and recurrence:

$$
\begin{aligned}
\mathcal{L}f = \frac{a}{2} f'' + b f' = 0 \quad &\Leftrightarrow \quad f' = C_1 \exp - \int_0^\bullet \frac{2b}{a}\,dx, \quad C_1 \in \mathbb{R} \\
&\Leftrightarrow \quad f = C_2 + C_1 \cdot s, \quad C_1, C_2 \in \mathbb{R}
\end{aligned}
$$

where

$$s := \int_0^\bullet e^{-\int_0^y \frac{2b(x)}{a(x)}\,dx}\,dy$$

is a strictly increasing harmonic function that is called the **scale function** or **natural scale of the diffusion**. In particular, $s(X_t)$ is a martingale with respect to P_x. The stopping theorem implies

$$P_x[T_a < T_b] = \frac{s(b)-s(x)}{s(b)-s(a)} \quad \forall\, a < x < b$$

As a consequence,

(i) If $s(\infty) < \infty$ or $s(-\infty) > -\infty$ then $P_x[|X_t| \to \infty] = 1$ for all $x \in \mathbb{R}$, i.e., (X_t, P_x) is *transient*.

(ii) If $s(\mathbb{R}) = \mathbb{R}$ then $P_x[T_a < \infty] = 1$ for all $x, a \in \mathbb{R}$, i.e., (X_t, P_x) is irreducible and *recurrent*.

b) Stationary distributions:

(i) $s(\mathbb{R}) \neq \mathbb{R}$: In this case, by the transience of (X_t, P_x), a stationary distribution does not exist. In fact, if μ is a finite stationary measure, then for all $t, r > 0$,

$$\mu(\{x : |x| \leq r\}) = (\mu p_t)(\{x : |x| \leq r\}) = P_\mu[|X_t| \leq r].$$

Since X_t is transient, the right hand side converges to 0 as $t \uparrow \infty$. Hence

$$\mu(\{x : |x| \leq r\}) = 0$$

for all $r > 0$, i.e., $\mu \equiv 0$.

(ii) $s(\mathbb{R}) = \mathbb{R}$: We can solve the ordinary differential equation $\mathcal{L}^* \varrho = 0$ explicitly:

$$\begin{aligned}
&& \mathcal{L}^* \varrho = \left(\frac{1}{2}(a\varrho)' - b\varrho\right)' &= 0 && \\
\Leftrightarrow && \frac{1}{2}(a\varrho)' - \frac{b}{a}a\varrho &= C_1 && \text{with } C_1 \in \mathbb{R} \\
\Leftrightarrow && \frac{1}{2}\left(e^{-\int_0^\bullet \frac{2b}{a}\,dx} a\varrho\right)' &= C_1 \cdot e^{-\int_0^\bullet \frac{2b}{a}\,dx} && \\
\Leftrightarrow && s' a\varrho &= C_2 + 2C_1 \cdot s && \text{with } C_1, C_2 \in \mathbb{R} \\
\Leftrightarrow && \varrho(y) = \frac{C_2}{a(y)s'(y)} &= \frac{C_2}{a(y)} e^{\int_0^y \frac{2b}{a}\,dx} && \text{with } C_2 \geq 0
\end{aligned}$$

Here the last equivalence holds since $s' a\varrho \geq 0$ and $s(\mathbb{R}) = \mathbb{R}$ imply $C_2 = 0$. Hence a stationary distribution μ can only exist if the measure

$$m(dy) := \frac{1}{a(y)} e^{\int_0^y \frac{2b}{a}\,dx}\, dy$$

is finite, and in this case $\mu = \frac{m}{m(\mathbb{R})}$. The measure m is called the **speed measure** of the diffusion.

Concrete examples:

(i) **Brownian motion:** $a \equiv 1, b \equiv 0, s(y) = y$. There is no stationary distribution. Lebesgue measure is an infinite stationary measure.

(ii) **Ornstein-Uhlenbeck process:**

$$\begin{aligned}
dX_t &= dB_t - \gamma X_t\, dt, && \gamma > 0, \\
\mathcal{L} &= \frac{1}{2}\frac{d^2}{dx^2} - \gamma x \frac{d}{dx}, && a \equiv 1, \\
b(x) &= -\gamma x, && s(y) = \int_0^y e^{\int_0^y 2\gamma x\,dx}\, dy = \int_0^y e^{\gamma y^2}\, dy \text{ recurrent}, \\
m(dy) &= e^{-\gamma y^2}\, dy, && \mu = \frac{m}{m(\mathbb{R})} = N\left(0, \frac{2}{\gamma}\right) \text{ is the unique stationary distribution}
\end{aligned}$$

(iii)

$$dX_t = dB_t + b(X_t)\,dt, \qquad b \in C^2, \qquad b(x) = \frac{1}{x} \text{ for } |x| \geq 1$$

transient, two independent non-negative solutions of $\mathcal{L}^*\varrho = 0$ with $\int \varrho\,dx = \infty$.

(**Exercise:** stationary distributions for $dX_t = dB_t - \frac{\gamma}{1+|X_t|}\,dt$)

Diffusion processes with a given invariant measure

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b \cdot \nabla, \qquad \mathcal{A} = C_0^\infty(\mathbb{R}^n)$$

μ probability measure on $\mathbb{R}^n$ (more generally locally finite positive measure)

Question: For which process is μ stationary?

Theorem 10.4. Suppose $\mu = \varrho\,dx$ with $\varrho_i a_{ij} \in C^1, b \in C, \varrho > 0$. Then

(i) We have
$$\mathcal{L}g = \mathcal{L}_s g + \mathcal{L}_a g$$
for all $g \in C_0^\infty(\mathbb{R}^n)$ where
$$\mathcal{L}_s g = \frac{1}{2}\sum_{i,j=1}^{n} \frac{1}{\varrho}\frac{\partial}{\partial x_i}\left(\varrho a_{ij}\frac{\partial g}{\partial x_i}\right)$$
$$\mathcal{L}_a g = \beta \cdot \nabla g, \quad \beta_j = b_j - \sum_i \frac{1}{2\varrho}\frac{\partial}{\partial x_i}\left(\varrho a_{ij}\right)$$

(ii) The operator $(\mathcal{L}_s, C_0^\infty)$ is symmetric with respect to μ.

(iii) The following assertions are equivalent:
   (i) $\mathcal{L}^*\mu = 0 \qquad$ (i.e. $\int \mathcal{L}f\,d\mu = 0$ for all $f \in C_0^\infty$).
   (ii) $\mathcal{L}_a^*\mu = 0$
   (iii) $\div(\varrho\beta) = 0$
   (iv) $(\mathcal{L}_a, C_0^\infty)$ is anti-symmetric with respect to μ

Proof. Let
$$\mathcal{E}(f,g) := -\int f\mathcal{L}g\,d\mu \qquad (f,g \in C_0^\infty)$$
denote the bilinear form of the operator $(\mathcal{L}, C_0^\infty(\mathbb{R}^n))$ on the Hilbert space $L^2(\mathbb{R}^n,\mu)$. We decompose $\mathcal{E}$ into a symmetric part and a remainder. An explicit computation based on the integration by parts formula in $\mathbb{R}^n$

shows that for $g \in C_0^\infty(\mathbb{R}^n)$ and $f \in C^\infty(\mathbb{R}^n)$:

$$\begin{aligned}
\mathcal{E}(f,g) &= -\int f\left(\frac{1}{2}\sum a_{ij}\frac{\partial^2 g}{\partial x_i \partial x_j} + b\cdot\nabla g\right)\varrho\, dt \\
&= \int \frac{1}{2}\sum_{i,j}\frac{\partial}{\partial x_i}\left(\varrho a_{ij} f\right)\frac{\partial g}{\partial x_j}\, dx - \int f b\cdot\nabla g \varrho\, dx \\
&= \int \frac{1}{2}\sum_{i,j} a_{i,j}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\varrho\, dx - \int f\beta\cdot\nabla g\varrho\, dx \qquad \forall\, f,g \in C_0^\infty
\end{aligned}$$

and set

$$\begin{aligned}
\mathcal{E}_s(f,g) &:= \int \frac{1}{2}\sum_{i,j} a_{i,j}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\varrho\, dx = -\int f\mathcal{L}_s g\, d\mu \\
\mathcal{E}_a(f,g) &:= \int f\beta\cdot\nabla g\varrho\, dx = -\int f\mathcal{L}_a g\, d\mu
\end{aligned}$$

This proves 1) and, since $\mathcal{E}_s$ is a symmetric bilinear form, also 2). Moreover, the assertions (i) and (ii) of 3) are equivalent, since

$$-\int \mathcal{L}g\, d\mu = \mathcal{E}(1,g) = \mathcal{E}_s(1,g) + \mathcal{E}_a(1,g) = -\int \mathcal{L}_a g\, d\mu$$

for all $g \in C_0^\infty(\mathbb{R}^n)$ since $\mathcal{E}_s(1,g) = 0$. Finally, the equivalence of (ii),(iii) and (iv) has been shown in the example above. ■

Example. $\mathcal{L} = \frac{1}{2}\Delta + b\cdot\nabla$, $b \in C(\mathbb{R}^n, \mathbb{R}^n)$,

$$\begin{aligned}
(\mathcal{L}, C_0^\infty)\ \mu\text{-symmetric} \quad &\Leftrightarrow \quad \beta = b - \frac{1}{2\varrho}\nabla\varrho = 0 \\
&\Leftrightarrow \quad b = \frac{\nabla\varrho}{2\varrho} = \frac{1}{2}\nabla\log\varrho
\end{aligned}$$

where $\log\varrho = -H$ if $\mu = e^{-H}\, dx$.

$$\begin{aligned}
\mathcal{L}\ \text{symmetrizable} \quad &\Leftrightarrow \quad b \text{ is a gradient} \\
\mathcal{L}^*\mu = 0 \quad &\Leftrightarrow \quad b = \frac{1}{2}\nabla\log\varrho + \beta
\end{aligned}$$

when $\div(\varrho\beta) = 0$.

Example (Langevin dynamics).

10.2. Non-symmetric Ornstein-Uhlenbeck processes

10.3. Couplings and Wasserstein bounds

10.4. Hypocoercive bounds

10.5. Hamiltonian Monte Carlo

Part IV.

Appendix

A. Conditioning and martingales

Let $(\Omega, \mathcal{A}, P)$ be a probability space, we denote by $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ ($\mathcal{L}^1(P)$) the space of measurable random variables $X : \Omega \to \mathbb{R}$ with $E[X^-] < \infty$ and $L^1(P) := \mathcal{L}^1(P)/\sim$ where two random variables a in relation to each other, if they are equal almost everywhere.

A.1. Conditional expectations

For more details and proofs of the following statements see [Eberle:Stochastic processes] [18].

Definition A.1 (Conditional expectations). Let $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ (or non-negative) and $\mathcal{F} \subset \mathcal{A}$ a σ-algebra. A random variable $Z \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ is called **conditional expectation** of X given $\mathcal{F}$ (written $Z = E[X|\mathcal{F}]$), if

- Z is $\mathcal{F}$-measurable, and
- for all $B \in \mathcal{F}$,

$$\int_B Z dP = \int_B X dP.$$

The random variable $E[X|\mathcal{F}]$ is P-a.s. unique. For a measurable Space $(S, \mathcal{S}$ and an abritatry random variable $Y : \Omega \to S$ we define $E[X|Y] := E[X|\sigma(Y)]$ and there exists a P-a.s. unique measurable function $g : S \to \mathbb{R}$ such that $E[X|\sigma(Y)] = g(Y)$. One also sometimes defines $E[X|Y = y] := g(y)$ μ_Y-a.e. (μ_Y law of Y).

Theorem A.2. Let X, Y and $X_n (n \in \mathbb{N})$ be non-negative or integrable random variables on $(\Omega, \mathcal{A}, P)$ and $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ two σ-algebras. The following statements hold:

(i) Linearity: $E[\lambda X + \mu Y|\mathcal{F}] = \lambda E[X|\mathcal{F}] + \mu E[Y|\mathcal{F}]$ P-almost surely for all $\lambda, \mu \in \mathbb{R}$.

(ii) Monotonicity: $X \geq 0$ P-almost surely implies that $E[X|\mathcal{F}] \geq 0$ P-almost surely.

(iii) $X = Y$ P-almost surely implies that $E[X|\mathcal{F}] = E[Y|\mathcal{F}]$ P-almost surely.

(iv) Monotone convergence: If (X_n) is growing monotone with $X_1 \geq 0$, then

$$E[\sup X_n|\mathcal{F}] = \sup E[X_n|\mathcal{F}] \ P\text{-almost surely.}$$

(v) Projectivity / Tower property: If $\mathcal{G} \subset \mathcal{F}$, then

$$E[E[X|\mathcal{F}]|\mathcal{G}] = E[X|\mathcal{G}] \ P\text{-almost surely.}$$

In particular:

$$E[E[X|Y, Z]|Y] = E[X|Y] \ P\text{-almost surely.}$$

(vi) Let Y be $\mathcal{F}$-measurable with $Y \cdot X \in \mathcal{L}^1$ or ≥ 0. This implies that

$$E[Y \cdot X|\mathcal{F}] = Y \cdot E[X|\mathcal{F}] \; P\text{-almost surely.}$$

(vii) Independence: If X is independent of $\mathcal{F}$, then $E[X|\mathcal{F}] = E[X]$ P-almost surely.

(viii) Let $(S, \mathcal{S}$ and $(T, \mathcal{T})$ be two measurable spaces. If $Y : \Omega \to S$ is $\mathcal{F}$-measurable, $X : \Omega \to T$ independent of $\mathcal{F}$ and $f : S \times T \to [0, \infty)$ a product measurable map, then it holds that

$$E[f(X,Y)|\mathcal{F}](\omega) = E[f(X, Y(\omega))] \text{ for } P\text{-almost all } \omega$$

Definition A.3 (Conditional probability). Let $(\Omega, \mathcal{A}, P)$ be a probability space, $\mathcal{F}$ a σ-algebra. The **conditional probability** is defined as

$$P[A|\mathcal{F}](\omega) := E[1_A|\mathcal{F}](\omega) \; \forall A \in \mathcal{F}, \omega \in \Omega.$$

A.2. Regular versions of conditional laws

Let $X, Y : \Omega \to S$ be random variables on a probability space $(\Omega, \mathfrak{A}, P)$ with polish state space S. A **regular version of the conditional distribution of Y given X** is a stochastic kernel $p(x, dy)$ on S such that

$$P[Y \in B|X] = p(X, B) \quad P-\text{a.s. for any } B \in \mathcal{B}.$$

If p is a regular version of the conditional distribution of Y given X then

$$P[X \in A, Y \in B] = E\left[P[Y \in B|X]; X \in A\right] = \int_A p(x, B)\mu_X(dx) \quad \text{for any } A, B \in \mathcal{B},$$

where μ_X denotes the law of X. For random variables with a polish state space, regular versions of conditional distributions always exist, cf. [XXX] []. Now let μ and p be a probability measure and a transition kernel on $(S, \mathcal{B})$. The first step towards analyzing a Markov chain with initial distribution μ and transition probability is to consider a single transition step:

Lemma A.4 (Two-stage model). *Suppose that X and Y are random variables on a probability space $(\Omega, \mathfrak{A}, P)$ such that $X \sim \mu$ and $p(X, dy)$ is a regular version of the conditional law of Y given X. Then*

$$(X, Y) \sim \mu \otimes p \quad \textit{and} \quad Y \sim \mu p,$$

where $\mu \otimes p$ and μp are the probability measures on $S \times S$ and S respectively defined by

$$(\mu \otimes p)(A) = \int \mu(dx) \left(\int p(x, dy) 1_A(x, y) \right) \quad \textit{for } A \in \mathcal{B} \otimes \mathcal{B},$$
$$(\mu p)(C) = \int \mu(dx) p(x, C) \quad \textit{for } C \in \mathcal{B}.$$

Proof. Let $A = B \times C$ with $B, C \in \mathcal{B}$. Then

$$\begin{aligned} P[(X,Y) \in A] &= P[X \in B, Y \in C] = E[P[X \in B, Y \in C|X]] \\ &= E[1_{\{X \in B\}} P[Y \in C|X]] = E[p(X,C); X \in B] \\ &= \int_B \mu(dx) p(x,C) = (\mu \otimes p)(A), \quad \textit{and} \\ P[Y \in C] &= P[(X,Y) \in S \times C] = (\mu p)(C). \end{aligned}$$

The assertion follows since the product sets form a generating system for the product σ-algebra that is stable under intersections. ■

A.3. Martingales

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

Filtrations

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

Definition A.5 (Filtration, adapted process). (i) A *filtration* on $(\Omega, \mathcal{A})$ is an increasing sequence

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

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

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

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

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

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

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

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

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

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

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

Martingales and supermartingales

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

Definition A.6 (Martingale, supermartingale, submartingale).

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

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

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

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

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

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

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

and correspondingly with "=" replaced by "≤" or "≥" for super- or submartingales.

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

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

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

Lemma A.7. *If (M_n) is a martingale (respectively a supermartingale) w.r.t. a filtration $(\mathcal{F}_n)$ then*

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

Doob Decomposition

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

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

Definition A.8 (Predictable process). A stochastic process $(A_n)_{n \geq 0}$ is called *predictable w.r.t.* $(\mathcal{F}_\mathbf{n})$ if and only if A_0 is constant and A_n is measurable w.r.t. $\mathcal{F}_{n-1}$ for any $n \in \mathbb{N}$.

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

Theorem A.9 (Doob decomposition). Every $(\mathcal{F}_n)$ adapted sequence of integrable random variables Y_n $(n \geq 0)$ has a unique decomposition (up to modification on null sets)

$$Y_n \quad = \quad M_n + A_n \tag{A.1}$$

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

$$A_n \quad = \quad \sum_{k=1}^{n} E[Y_k - Y_{k-1} \mid \mathcal{F}_{k-1}], \qquad \text{and} \qquad M_n = Y_n - A_n. \tag{A.2}$$

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

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

Proof (Proof of Theorem A.9). *Uniqueness:* For any decomposition as in (A.1) we have

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

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

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

This implies that (A.2) holds almost surely if $A_0 = 0$.

Existence: Conversely, if (A_n) and (M_n) are defined by (A.2) then (A_n) is predictable with $A_0 = 0$ and (M_n) is a martingale, since

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

■

A.4. Stopping times

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

Martingale transforms

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

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

i.e.,

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

Definition A.10 (Martingale transform). The stochastic process $C_\bullet M$ defined by

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

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

The process $C_\bullet M$ is a time-discrete version of the stochastic integral $\int_0^t C_s \, dM_s$ for continuous-time processes C and M, cf. [Introduction to Stochastic Analysis].

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

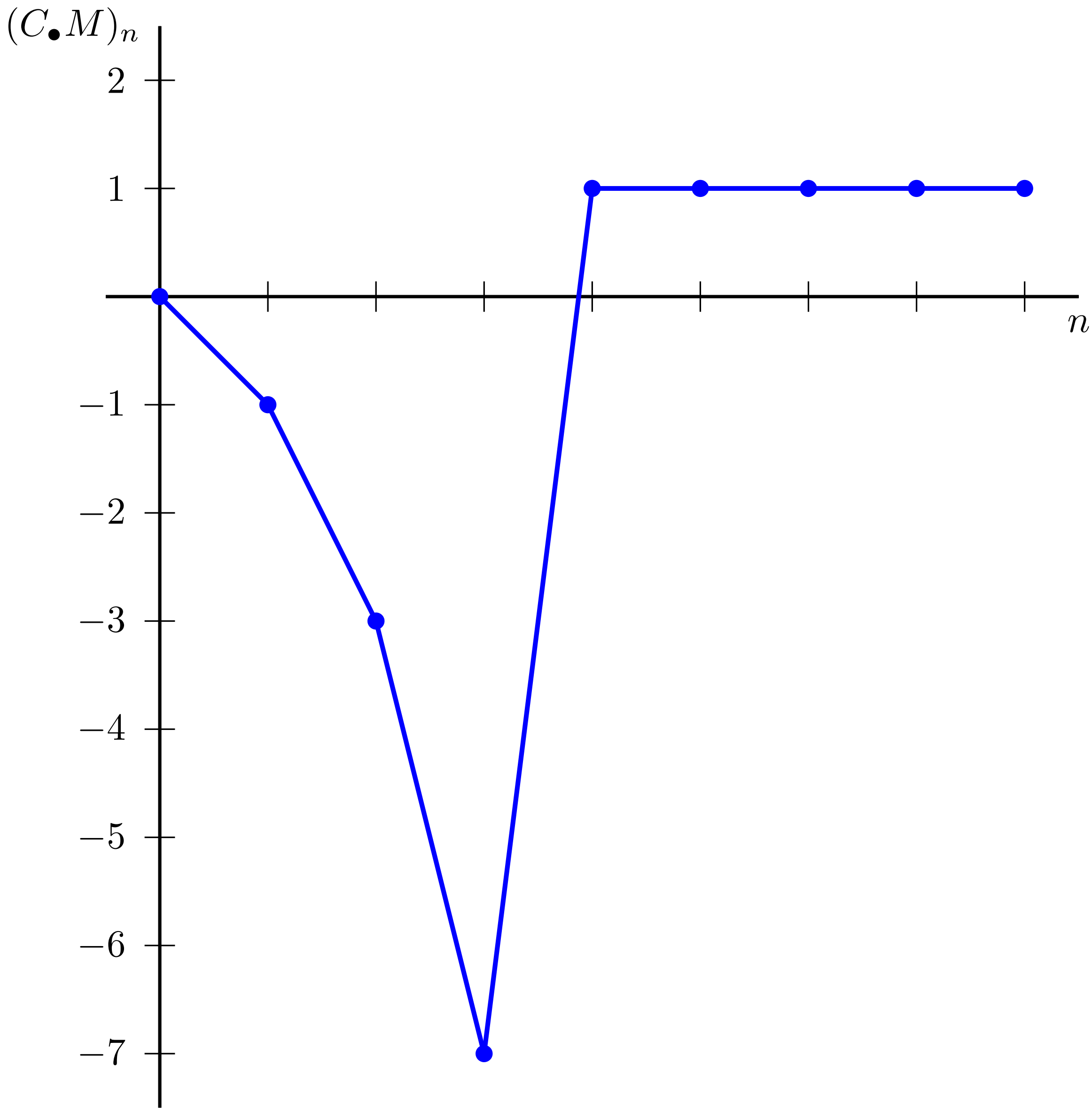
$$M_n \;=\; \eta_1 + \ldots + \eta_n, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = 1] = P[\eta_i = -1] = 1/2,$$

then the stake in the n-th round is

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

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

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



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

Theorem A.11 (You can't beat the system!). (i) If $(M_n)_{n\geq 0}$ is an $(\mathcal{F}_n)$ martingale, and $(C_n)_{n\geq 1}$ is predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ for any $n \geq 1$, then $C_\bullet M$ is again an $(\mathcal{F}_n)$ martingale.

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

Proof. For $n \geq 1$ we have

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

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

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

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

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

by the martingale property, although

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

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

Remark. The integrability assumption in Theorem A.11 is always satisfied if the random variables C_n are bounded, or if both C_n and M_n are square-integrable for any n.

Stopped Martingales

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

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

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

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

Theorem A.12 (Optional Stopping Theorem,Version 1). If $(M_n)_{n\geq 0}$ is a martingale (resp. a supermartingale) w.r.t. $(\mathcal{F}_n)$, and T is an $(\mathcal{F}_n)$-stopping time, then the stopped process $(M_{T\wedge n})_{n\geq 0}$ is again an $(\mathcal{F}_n)$-martingale (resp. supermartingale). In particular, we have

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

Proof. Consider the following strategy:

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

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

$$M_{T\wedge n} - M_0 = (C_\bullet M)_n \qquad \text{for any } n \geq 0. \tag{A.3}$$

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

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

and (A.3) follows by summing over n. Since the sequence (C_n) is predictable, bounded and non-negative, the process $C_\bullet M$ is a martingale, supermartingale respectively, provided the same holds for M. ■

Remark (IMPORTANT). (i) In general, it is NOT TRUE under the assumptions in Theorem A.12 that

$$E[M_T] = E[M_0], \quad E[M_T] \leq E[M_0] \quad \text{respectively.} \tag{A.4}$$

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

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

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

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

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

with initial value $X_0 = x$. We now show how to apply the optional stopping theorem to compute the distributions of the exit time

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

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

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

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

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

$$r(x) = \frac{b-x}{a-x}. \tag{A.5}$$

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

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

By monotone and dominated convergence, we obtain

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

Therefore, by (A.5),

$$\begin{aligned} E[T] &= E[X_T^2] - x^2 = a^2 \cdot r(x) + b^2 \cdot (1 - r(x)) - x^2 \\ &= (b-x) \cdot (x-a). \end{aligned} \tag{A.6}$$

(iii) *Mean passage time of b is infinite.*
The first passage time $T_b = \min\{n \geq 0 \ : \ X_n = b\}$ is greater or equal than the exit time from the interval (a,b) for any $a < x$. Thus by (A.6), we have

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

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

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

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

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

$$1 \ = \ E[M_0^\lambda] \ = \ E[M_{T\wedge n}^\lambda] \ = \ E[e^{\lambda X_{T\wedge n}}/(\cosh \lambda)^{T\wedge n}] \tag{A.7}$$

for any $n \in \mathbb{N}$. As $n \to \infty$, the integrands on the right hand side converge to $e^\lambda (\cosh \lambda)^{-T} \cdot I_{\{T<\infty\}}$. Moreover, they are uniformly bounded by e^λ, since $X_{T\wedge n} \leq 1$ for any n. Hence by the Dominated Convergence Theorem, the expectation on the right hand side of (A.7) converges to $E[e^\lambda/(\cosh \lambda)^T \ ; \ T < \infty]$, and we obtain the identity

$$E[(\cosh \lambda)^{-T} \ ; \ T < \infty] \ = \ e^{-\lambda} \qquad \text{for any } \lambda > 0. \tag{A.8}$$

Taking the limit as $\lambda \searrow 0$, we see that $P[T < \infty] = 1$. Taking this into account, and substituting $s = 1/\cosh \lambda$ in (A.8), we can now compute the generating function of T explicitly:

$$E[s^T] \ = \ e^{-\lambda} \ = \ (1 - \sqrt{1-s^2})/s \qquad \text{for any } s \in (0,1). \tag{A.9}$$

Developing both sides into a power series finally yields

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

Therefore, the distribution of the first passage time of 1 is given by $P[T = 2m] = 0$ and

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

for any $m \geq 1$.

Optional Stopping Theorems

Stopping times occurring in applications are typically not bounded, see the example above. Therefore, we need more general conditions guaranteeing that (A.4) holds nevertheless. A first general criterion is obtained by applying the Dominated Convergence Theorem:

Theorem A.13 (Optional Stopping Theorem, Version 2). Suppose that (M_n) is a martingale w.r.t. $(\mathcal{F}_n)$, T is an $(\mathcal{F}_n)$-stopping time with $P[T < \infty] = 1$, and there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ such that

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

Then

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

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

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

By Theorem A.12, $E[M_0] = E[M_{T\wedge n}]$, and by the Dominated Convergence Theorem, $E[M_{T\wedge n}] \longrightarrow E[M_T]$ as $n \to \infty$. ■

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

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

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

Theorem A.14 (Optional Stopping Theorem, Version 3). If (M_n) is a non-negative supermartingale w.r.t. $(\mathcal{F}_n)$, then

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

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

Proof. Since $M_T = \lim_{n\to\infty} M_{T\wedge n}$ on $\{T < \infty\}$, and $M_T \geq 0$, Theorem A.12 combined with Fatou's Lemma implies

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

A.5. Almost sure convergence of supermartingales

The strength of martingale theory is partially due to powerful general convergence theorems that hold for martingales, sub- and supermartingales. Let $(Z_n)_{n\geq 0}$ be a discrete-parameter supermartingale w.r.t. a filtration $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$. The following theorem yields a stochastic counterpart to the fact that any lower bounded decreasing sequence of reals converges to a finite limit:

Theorem A.15 (Supermartingale Convergence Theorem, Doob). If $\sup_{n\geq 0} E[Z_n^-] < \infty$ then (Z_n) converges almost surely to an integrable random variable $Z_\infty \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. In particular, supermartingales that are uniformly bounded from above converge almost surely to an integrable random variable.

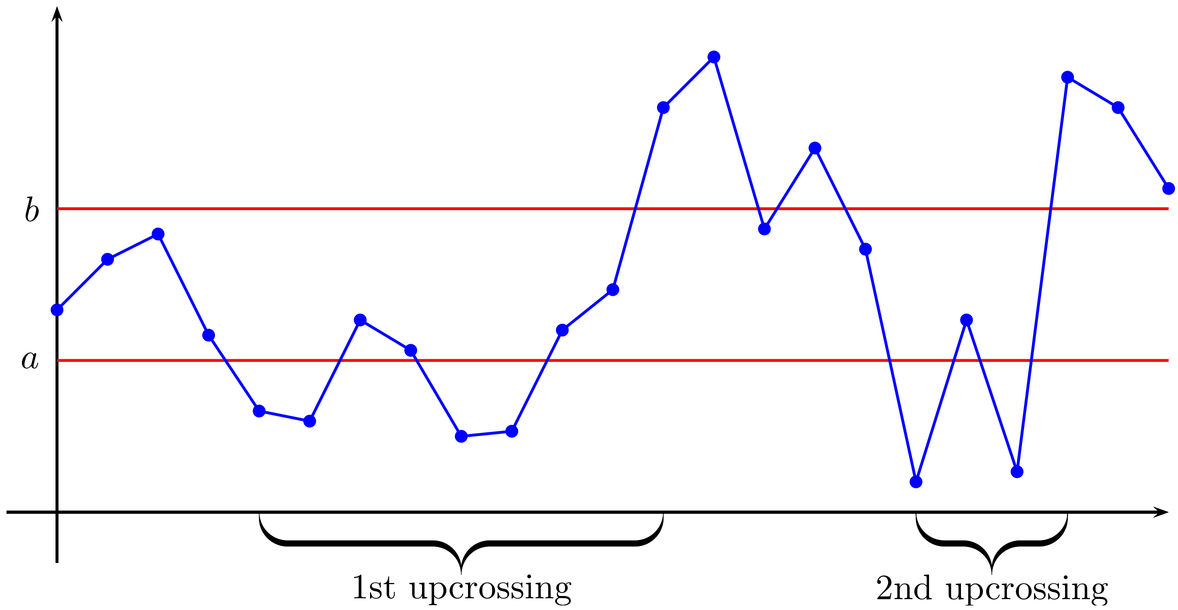
Remark (L^1 boundedness and L^1 convergence).

(i) Although the limit is integrable, L^1 convergence does *not* hold in general.

(ii) The condition $\sup E[Z_n^-] < \infty$ holds if and only if (Z_n) is bounded in L^1. Indeed, as $E[Z_n^+] < \infty$ by our definition of a supermartingale, we have

$$E[\,|Z_n|\,] \;=\; E[Z_n] + 2E[Z_n^-] \;\leq\; E[Z_0] + 2E[Z_n^-] \qquad \text{for any } n \geq 0.$$

For proving the Supermartingale Convergence Theorem, we introduce the number $U^{(a,b)}(\omega)$ of upcrossings over an interval (a, b) by the sequence $Z_n(\omega)$, cf. below for the exact definition.



Note that if $U^{(a,b)}(\omega)$ is finite for any non-empty bounded interval (a, b) then $\limsup Z_n(\omega)$ and $\liminf Z_n(\omega)$ coincide, i.e., the sequence $(Z_n(\omega))$ converges. Therefore, to show almost sure convergence of (Z_n), we derive an upper bound for $U^{(a,b)}$. We first prove this key estimate and then complete the proof of the theorem.

Doob's upcrossing inequality

For $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $a < b$ we define the number $U_n^{(a,b)}$ of upcrossings over the interval (a, b) before time n by

$$U_n^{(a,b)} = \max\left\{k \geq 0 : \exists\, 0 \leq s_1 < t_1 < s_2 < t_2 \ldots < s_k < t_k \leq n : Z_{s_i} \leq a, Z_{t_i} \geq b\right\}.$$

Lemma A.16 (Doob). *If (Z_n) is a supermartingale then*

$$(b - a) \cdot E[U_n^{(a,b)}] \leq E[(Z_n - a)^-] \qquad \textit{for any } a < b \textit{ and } n \geq 0.$$

Proof. We may assume $E[Z_n^-] < \infty$ since otherwise there is nothing to prove. The key idea is to set up a predictable gambling strategy that increases our capital by $(b - a)$ for each completed upcrossing. Since the net gain with this strategy should again be a supermartingale this yields an upper bound for the average number of upcrossings. Here is the strategy:

- Wait until $Z_k \leq a$.
- Then play unit stakes until $Z_k \geq b$.

The stake C_k in round k is

$$C_1 = \begin{cases} 1 & \text{if } Z_0 \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$C_k = \begin{cases} 1 & \text{if } (C_{k-1} = 1 \text{ and } Z_{k-1} \leq b) \text{ or } (C_{k-1} = 0 \text{ and } Z_{k-1} \leq a), \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, (C_k) is a predictable, bounded and non-negative sequence of random variables. Moreover, $C_k \cdot (Z_k - Z_{k-1})$ is integrable for any $k \le n$, because C_k is bounded and

$$E\big[\, |Z_k| \,\big] \;=\; 2E[Z_k^+] - E[Z_k] \;\le\; 2E[Z_k^+] - E[Z_n] \le 2E[Z_k^+] - E[Z_n^-]$$

for $k \le n$. Therefore, by Theorem A.11 and the remark below, the process

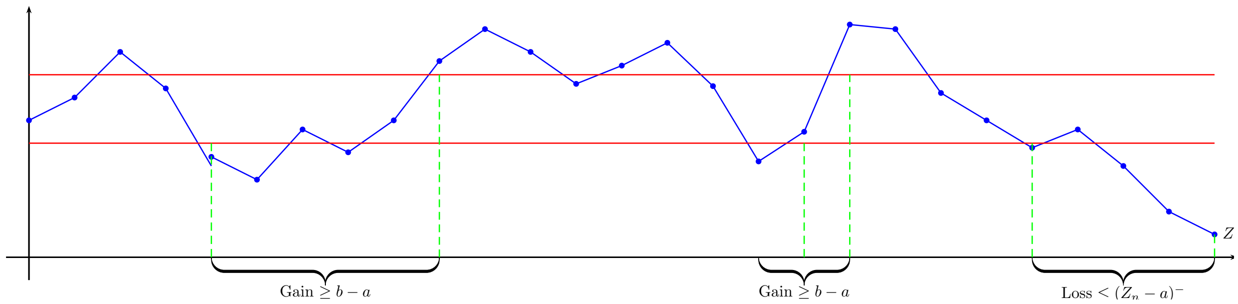
$$(C_\bullet Z)_k \;=\; \sum_{i=1}^{k} C_i \cdot (Z_i - Z_{i-1}), \qquad 0 \le k \le n,$$

is again a supermartingale.

Clearly, the value of the process $C_\bullet Z$ increases by at least $(b-a)$ units during each completed upcrossing. Between upcrossing periods, the value of $(C_\bullet Z)_k$ is constant. Finally, if the final time n is contained in an upcrossing period, then the process can decrease by at most $(Z_n - a)^-$ units during that last period (since Z_k might decrease before the next upcrossing is completed). Therefore, we have

$$(C_\bullet Z)_n \quad\ge\quad (b-a)\cdot U_n^{(a,b)} - (Z_n - a)^-, \qquad \text{i.e.,}$$

$$(b-a)\cdot U_n^{(a,b)} \quad\le\quad (C_\bullet Z)_n + (Z_n - a)^-.$$



Since $C_\bullet Z$ is a supermartingale with initial value 0, we obtain the upper bound

$$(b-a)E[U_n^{(a,b)}] \;\le\; E[(C_\bullet Z)_n] + E[(Z_n - a)^-] \;\le\; E[(Z_n - a)^-]. \qquad \blacksquare$$

Proof of Doob's Convergence Theorem

We can now complete the proof of Theorem A.15.

Proof. Let

$$U^{(a,b)} \;=\; \sup_{n\in\mathbb{N}} U_n^{(a,b)}$$

denote the total number of upcrossings of the supermartingale (Z_n) over an interval (a,b) with $-\infty < a < b < \infty$. By the upcrossing inequality and monotone convergence,

$$E[U^{(a,b)}] \;=\; \lim_{n\to\infty} E[U_n^{(a,b)}] \;\le\; \frac{1}{b-a}\cdot \sup_{n\in\mathbb{N}} E[(Z_n - a)^-]. \tag{A.10}$$

Assuming $\sup E[Z_n^-] < \infty$, the right hand side of (A.10) is finite since $(Z_n - a)^- \le |a| + Z_n^-$. Therefore,

$$U^{(a,b)} \;<\; \infty \qquad P\text{-almost surely,}$$

and hence the event

$$\{\liminf Z_n \neq \limsup Z_n\} = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{U^{(a,b)} = \infty\}$$

has probability zero. This proves almost sure convergence.

It remains to show that the almost sure limit $Z_\infty = \lim Z_n$ is an integrable random variable (in particular, it is finite almost surely). This holds true as, by the remark below Theorem A.15, $\sup E[Z_n^-] < \infty$ implies that (Z_n) is bounded in L^1, and therefore

$$E[\,|Z_\infty|\,] = E[\lim |Z_n|\,] \leq \liminf E[\,|Z_n|\,] < \infty$$

by Fatou's lemma. ■

Examples and first applications

We now consider a few prototypic applications of the almost sure convergence theorem:

Example (Sums of i.i.d. random variables). Consider a Random Walk

$$S_n = \sum_{i=1}^n \eta_i$$

on $\mathbb{R}$ with centered and bounded increments:

$$\eta_i \text{ i.i.d. with } |\eta_i| \leq c \text{ and } E[\eta_i] = 0, \quad c \in \mathbb{R}.$$

Suppose that $P[\eta_i \neq 0] > 0$. Then there exists $\varepsilon > 0$ such that $P[|\eta_i| \geq \varepsilon] > 0$. As the increments are i.i.d., the event $\{|\eta_i| \geq \varepsilon\}$ occurs infinitely often with probability one. Therefore, almost surely the martingale (S_n) does not converge as $n \to \infty$.

Now let $a \in \mathbb{R}$. We consider the first hitting time

$$T_a = \min\{n \geq 0 : S_n \geq a\}$$

of the interval $[a, \infty)$. By the Optional Stopping Theorem, the stopped Random Walk $(S_{T_a \wedge n})_{n\geq 0}$ is again a martingale. Moreover, as $S_k < a$ for any $k < T_a$ and the increments are bounded by c, we obtain the upper bound

$$S_{T_a\wedge n} < a + c \qquad \text{for any } n \in \mathbb{N}.$$

Therefore, the stopped Random Walk converges almost surely by the Supermartingale Convergence Theorem. As (S_n) does not converge, we can conclude that $P[T_a < \infty] = 1$ for any $a > 0$, i.e.,

$$\limsup S_n = \infty \quad \text{almost surely.}$$

Since (S_n) is also a submartingale, we obtain

$$\liminf S_n = -\infty \quad \text{almost surely}$$

by an analogue argument.

Remark (Almost sure vs. $\mathbf{L^p}$ convergence). In the last example, the stopped process does not converge in L^p for any $p \in [1, \infty)$. In fact,

$$\lim_{n\to\infty} E[S_{T_a\wedge n}] = E[S_{T_a}] \geq a \qquad \text{whereas} \qquad E[S_0] = 0.$$

Example (Products of non-negative i.i.d. random variables). Consider a growth process

$$Z_n \quad = \quad \prod_{i=1}^{n} Y_i$$

with i.i.d. factors $Y_i \geq 0$ with finite expectation $\alpha \in (0, \infty)$. Then

$$M_n \quad = \quad Z_n / \alpha^n$$

is a martingale. By the almost sure convergence theorem, a finite limit M_∞ exists almost surely, because $M_n \geq 0$ for all n. For the almost sure asymptotics of (Z_n), we distinguish three different cases:

(i) $\alpha < 1$ *(subcritical):* In this case,

$$Z_n \quad = \quad M_n \cdot \alpha^n$$

converges to 0 exponentially fast with probability one.

(ii) $\alpha = 1$ *(critical):* Here (Z_n) is a martingale and converges almost surely to a finite limit. If $P[Y_i \neq 1] > 0$ then there exists $\varepsilon > 0$ such that $Y_i \geq 1 + \varepsilon$ infinitely often with probability one. This is consistent with convergence of (Z_n) only if the limit is zero. Hence, if (Z_n) is not almost surely constant, then also in the critical case $Z_n \to 0$ almost surely.

(iii) $\alpha > 1$ *(supercritical):* In this case, on the set $\{M_\infty > 0\}$,

$$Z_n \; = \; M_n \cdot \alpha^n \quad \sim \quad M_\infty \cdot \alpha^n,$$

i.e., (Z_n) grows exponentially fast. The asymptotics on the set $\{M_\infty = 0\}$ is not evident and requires separate considerations depending on the model.

Although most of the conclusions in the last example could have been obtained without martingale methods (e.g. by taking logarithms), the martingale approach has the advantage of carrying over to far more general model classes. These include for example branching processes or exponentials of continuous time processes.

Example (Boundary behaviour of harmonic functions). Let $D \subseteq \mathbb{R}^d$ be a bounded open domain, and let $h : D \to \mathbb{R}$ be a harmonic function on D that is bounded from below:

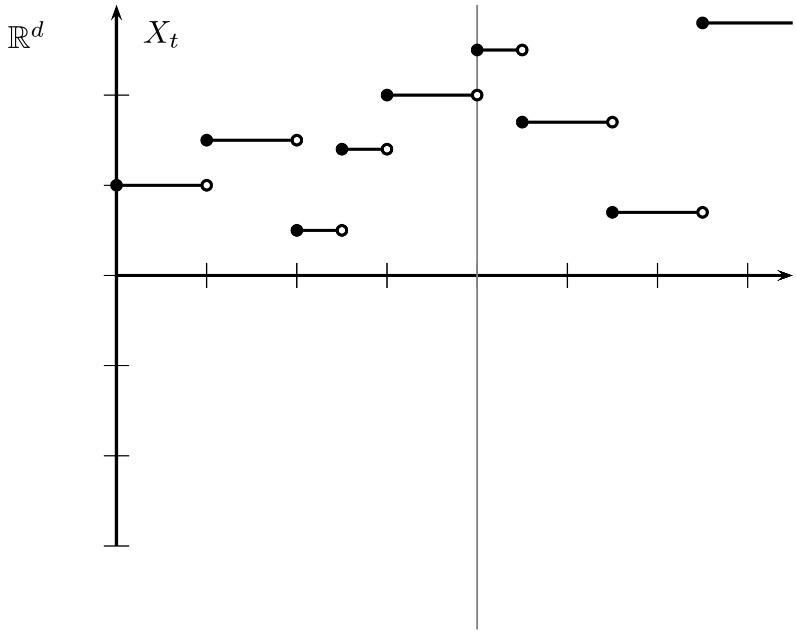
$$\Delta h(x) \; = \; 0 \qquad \text{for any } x \in D, \qquad \inf_{x \in D} h(x) > -\infty. \tag{A.11}$$

To study the asymptotic behavior of $h(x)$ as x approaches the boundary ∂D, we construct a Markov chain (X_n) such that $h(X_n)$ is a martingale: Let $r : D \to (0, \infty)$ be a continuous function such that

$$0 \; < \; r(x) \; < \; \operatorname{dist}(x, \partial D) \qquad \text{for any } x \in D, \tag{A.12}$$

and let (X_n) w.r.t P_x denote the canonical time-homogeneous Markov chain with state space D, initial value x, and transition probabilities

$$p(x, dy) \; = \; \text{Uniform distribution on the sphere } \{y \in \mathbb{R}^d \; : \; |y - x| = r(x)\}.$$



By (A.12), the function h is integrable w.r.t. $p(x, dy)$, and, by the mean value property,

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

Therefore, the process $h(X_n)$ is a martingale w.r.t. P_x for each $x \in D$. As $h(X_n)$ is lower bounded by (A.11), the limit as $n \to \infty$ exists P_x-almost surely by the Supermartingale Convergence Theorem. In particular, since the coordinate functions $x \mapsto x_i$ are also harmonic and lower bounded on $\overline{D}$, the limit $X_\infty = \lim X_n$ exists P_x-almost surely. Moreover, X_∞ is in ∂D, because r is bounded from below by a strictly positive constant on any compact subset of D.

Summarizing we have shown:

(i) *Boundary regularity:* If h is harmonic and bounded from below on D then the limit $\lim_{n\to\infty} h(X_n)$ exists along almost every trajectory X_n to the boundary ∂D.

(ii) *Representation of h in terms of boundary values:* If h is continuous on $\overline{D}$, then $h(X_n) \to h(X_\infty)$ P_x-almost surely and hence

$$h(x) = \lim_{n\to\infty} E_x[h(X_n)] = E[h(X_\infty)],$$

i.e., the distribution of X_∞ w.r.t. P_x is the harmonic measure on ∂D.

Note that, in contrast to classical results from analysis, the first statement holds without any smoothness condition on the boundary ∂D. Thus, although boundary values of h may not exist in the classical sense, they still do exist along almost every trajectory of the Markov chain!

B. Brownian motion and stochastic calculus

B.1. Brownian Motion

Definition B.1 (Brownian motion).

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

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

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

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

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

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

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

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

Index

Bibliography

[1] Hans Wilhelm Alt. *Linear functional analysis*. Universitext. An application-oriented introduction, Translated from the German edition by Robert Nürnberg. Springer-Verlag London, Ltd., London, 2016, pp. xii+435. ISBN: 978-1-4471-7279-6; 978-1-4471-7280-2. DOI: 10.1007/978-1-4471-7280-2. URL: https://doi.org/10.1007/978-1-4471-7280-2.

[2] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der mathematischen Wissenschaften. Springer, 2013. ISBN: 9783319002279. URL: https://books.google.de/books?id=gU3ABAAAQBAJ.

[3] P. Billingsley. *Convergence of probability measures*. Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics. Wiley, 1968. URL: http://books.google.de/books?id=O9oQAQAAIAAJ.

[4] E. Bolthausen, A.S. Sznitman, and Deutschen Mathematiker-Vereinigung. *Ten Lectures on Random Media*. DMV Seminar. Momenta, 2002. ISBN: 9783764367039. URL: https://books.google.de/books?id=N6iPAj6eitgC.

[5] Leo Breiman. *Probability*. Vol. 7. Classics in Applied Mathematics. Corrected reprint of the 1968 original. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, pp. xiv+421. ISBN: 0-89871-296-3. DOI: 10.1137/1.9781611971286. URL: https://doi.org/10.1137/1.9781611971286.

[6] Russ Bubley and Martin Dyer. "Path coupling: A technique for proving rapid mixing in Markov chains". In: *Proceedings 38th Annual Symposium on Foundations of Computer Science*. IEEE. 1997, pp. 223–231.

[7] Krzysztof Burdzy and Wilfrid S Kendall. "Efficient Markovian couplings: examples and counterexamples". In: *Annals of Applied Probability* (2000), pp. 362–409.

[8] M. Chen. *From Markov Chains to Non-equilibrium Particle Systems*. World Scientific, 2004. ISBN: 9789812562456. URL: http://books.google.de/books?id=zD50Y61NnKYC.

[9] E.B. Davies. *One-parameter semigroups*. L.M.S. monographs. Academic Press, 1980. ISBN: 9780122062803. URL: http://books.google.de/books?id=IQOFAAAAIAAJ.

[10] M. H. A. Davis. *Markov models and optimization*. Vol. 49. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1993, pp. xiv+295. ISBN: 0-412-31410-X. DOI: 10.1007/978-1-4899-4483-2. URL: https://doi.org/10.1007/978-1-4899-4483-2.

[11] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*. Vol. 29. North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978, pp. viii+189. ISBN: 0-7204-0701-X.

[12] J.D. Deuschel and D.W. Stroock. *Large deviations*. Pure and Applied Mathematics. Elsevier Science, 1989. ISBN: 9780080874579. URL: http://books.google.de/books?id=sKyCBxf%5C_RQcC.

[13] Rick Durrett. *Probability—theory and examples*. Vol. 49. Cambridge Series in Statistical and Probabilistic Mathematics. Fifth edition of [MR1068527]. Cambridge University Press, Cambridge, 2019, pp. xii+419. ISBN: 978-1-108-47368-2. DOI: 10.1017/9781108591034. URL: https://doi.org/10.1017/9781108591034.

[14] Rick Durrett. "Ten lectures on particle systems". In: *Lectures on probability theory (Saint-Flour, 1993)*. Vol. 1608. Lecture Notes in Math. Springer, Berlin, 1995, pp. 97–201. DOI: 10.1007/BFb0095747. URL: http://dx.doi.org/10.1007/BFb0095747.

[15] Andreas Eberle. *Lecture Notes „Einführung in die Wahrscheinlichkeitstheorie"*. Institute for Applied Mathematics, University of Bonn. http://wt.iam.uni-bonn.de/eberle/home/. 2013.

[16] Andreas Eberle. *Lecture Notes „Introduction to Stochastic Analysis"*. Institute for Applied Mathematics, University of Bonn. http://wt.iam.uni-bonn.de/eberle/home/. 2019.

[17] Andreas Eberle. *Lecture Notes „Stochastic Analysis"*. Institute for Applied Mathematics, University of Bonn. http://wt.iam.uni-bonn.de/eberle/home/. 2019.

[18] Andreas Eberle. *Lecture Notes „Stochastic processes"*. Institute for Applied Mathematics, University of Bonn. http://wt.iam.uni-bonn.de/eberle/home/. 2014.

[19] S.N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 1986. URL: http://books.google.de/books?id=BAWnAAAAIAAJ.

[20] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Reprint of the 1998 edition. Springer-Verlag, Berlin, 2001, pp. xiv+517. ISBN: 3-540-41160-7.

[21] Martin Hairer. *Lecture „Convergence of Markov processes"*. Mathematics Department, University of Warwick. http://www.hairer.org/notes/Convergence.pdf. 2010.

[22] Martin Hairer and Jonathan C. Mattingly. "Yet another look at Harris' ergodic theorem for Markov chains". In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Vol. 63. Progr. Probab. Birkhäuser/Springer Basel AG, Basel, 2011, pp. 109–117.

[23] Peter Holderrieth. *Cores for Piecewise-Deterministic Markov Processes used in Markov Chain Monte Carlo*. 2019. arXiv: 1910.11429 [math.ST].

[24] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Springer Science & Business Media, 2003.

[25] Galin L. Jones. "On the Markov chain central limit theorem". In: *Probability Surveys* 1 (2004), pp. 299–320. DOI: 10.1214/154957804100000051. URL: http://dx.doi.org/10.1214/154957804100000051.

[26] Aldéric Joulin and Yann Ollivier. "Curvature, concentration and error estimates for Markov chain Monte Carlo". In: *Ann. Probab.* 38.6 (2010), pp. 2418–2442. ISSN: 0091-1798. DOI: 10.1214/10-AOP541. URL: https://doi.org/10.1214/10-AOP541.

[27] Olav Kallenberg. *Foundations of modern probability*. Second. Probability and its Applications (New York). Springer-Verlag, New York, 2002, pp. xx+638. ISBN: 0-387-95313-2. DOI: 10.1007/978-1-4757-4015-8. URL: https://doi.org/10.1007/978-1-4757-4015-8.

[28] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991. ISBN: 9780387976556. URL: http://books.google.de/books?id=ATNy%5C_Zg3PSsC.

[29] R. Khasminskii. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability. Springer, 2011. ISBN: 9783642232800. URL: https://books.google.de/books?id=4hT60oIArOUC.

[30] A. Kolmogoroff. "Zur Umkehrbarkeit der statistischen Naturgesetze". In: *Math. Ann.* 113.1 (1937), pp. 766–772. ISSN: 0025-5831. DOI: 10.1007/BF01571664. URL: https://doi.org/10.1007/BF01571664.

[31] Kazumasa Kuwada et al. “Characterization of maximal Markovian couplings for diffusion processes”. In: *Electronic Journal of Probability* 14 (2009), pp. 633–662.

[32] Claudio Landim. “Central Limit Theorem for Markov Processes”. English. In: *From Classical to Modern Probability*. Ed. by Pierre Picco and Jaime San Martin. Vol. 54. Progress in Probability. Birkhaeuser Basel, 2003, pp. 145–205. ISBN: 978-3-0348-9422-7. DOI: 10.1007/978-3-0348-8053-4_5. URL: http://dx.doi.org/10.1007/978-3-0348-8053-4_5.

[33] S. Lang. *Analysis I*. Addison-Wesley world student series. Addison-Wesley, 1978. URL: https://books.google.de/books?id=8KDntgAACAAJ.

[34] T.M. Liggett. *Continuous Time Markov Processes: An Introduction*. Graduate studies in mathematics. American Mathematical Soc., 2010. ISBN: 9780821884195. URL: http://books.google.de/books?id=ClW%5C_cKYSON0C.

[35] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Reprint of the 1985 original. Springer-Verlag, Berlin, 2005, pp. xvi+496. ISBN: 3-540-22617-6.

[36] Georg Lindgren. *„Lecture notes on Stationary Stochastic Processes“*. Lund Institute of Technology Centre for Mathematical Sciences Mathematical Statistics. http://www.maths.lth.se/matstat/staff/georg/Publications/lecture2002.pdf. 2002.

[37] Torgny Lindvall. *Lectures on the coupling method*. Corrected reprint of the 1992 original. Dover Publications, Inc., Mineola, NY, 2002, pp. xiv+257. ISBN: 0-486-42145-7.

[38] Florent Malrieu. *Lecture Notes „Processus de Markov et inégalités fonctionelles“*. http://perso.univ-rennes1.fr/florent.malrieu/MASTER2/markov180406.pdf. 2005/2006.

[39] Robert J. McCann. “Exact solutions to the transportation problem on the line”. In: *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* 455.1984 (1999), pp. 1341–1380. ISSN: 1364-5021. DOI: 10.1098/rspa.1999.0364. URL: http://dx.doi.org/10.1098/rspa.1999.0364.

[40] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. 2nd. New York, NY, USA: Cambridge University Press, 2009. ISBN: 0521731828, 9780521731829.

[41] R. A. Minlos. *Introduction to mathematical statistical physics*. Vol. 19. University Lecture Series. American Mathematical Society, Providence, RI, 2000, pp. viii+103. ISBN: 0-8218-1337-4.

[42] Philippe Morando. “Mesures aléatoires”. In: *Séminaire de Probabilités, III (Univ. Strasbourg, 1967/68)*. Springer, Berlin, 1967/1968, pp. 190–229.

[43] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences. Springer New York, 1992. ISBN: 9780387908458. URL: http://books.google.de/books?id=sIAyOgM4R3kC.

[44] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory*. Mass Transportation Problems. Springer, 1998. ISBN: 9780387983509. URL: http://books.google.de/books?id=ZNb158VkIvoC.

[45] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. I. Functional analysis*. Methods of Modern Mathematical Physics Vol. 1. Academic Press, 1980. ISBN: 9780125850506. URL: https://books.google.de/books?id=fXX0j4qa8G8C.

[46] M. Reed and B. Simon. *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*. Fourier Nanlysis, Self-adjointness. Academic Press, 1975. ISBN: 9780125850025. URL: https://books.google.de/books?id=Kz7s7bgVe8gC.

[47] Gareth O. Roberts and Jeffrey S. Rosenthal. “General state space Markov chains and MCMC algorithms”. In: *Probab. Surv.* 1 (2004), pp. 20–71. ISSN: 1549-5787. DOI: 10.1214/154957804100000024. URL: http://dx.doi.org/10.1214/154957804100000024.

[48] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Itô calculus, Reprint of the second (1994) edition. Cambridge University Press, Cambridge, 2000, pp. xiv+480. ISBN: 0-521-77593-0. DOI: 10.1017/CBO9781107590120. URL: http://dx.doi.org/10.1017/CBO9781107590120.

[49] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge Mathematical Library. Cambridge University Press, 2000. ISBN: 9780521775946. URL: http://books.google.de/books?id=eJp33OpIvQ4C.

[50] G. Royer. *An initiation to logarithmic Sobolev inequalities*. American Mathematical Society, 2007.

[51] D.W. Stroock and S.R.S. Varadhan. *Multidimensional Diffussion Processes*. Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Springer, 1979. ISBN: 9783540903536. URL: http://books.google.de/books?id=DuDsmoyqCy4C.

[52] Daniel W. Stroock. *c*. 2nd. New York, NY, USA: Cambridge University Press, 2010. ISBN: 0521132509, 9780521132503.

[53] S. R. S. Varadhan. *Probability Theory*. Courant Lecture Notes. Courant Institute of Mathematical Sciences, 2001. ISBN: 9780821828526. URL: http://books.google.de/books?id=zAZJdAcDrKsC.

[54] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer, 2008. ISBN: 9783540710509. URL: http://books.google.es/books?id=hV8o5R7%5C_5tkC.

[55] Cédric Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003. ISBN: 9780821833124. URL: http://books.google.co.uk/books?id=GqRXYFxe0l0C.

[56] K. Yosida. *Functional Analysis*. Classics in Mathematics. Cambridge University Press, 1995. ISBN: 9783540586548. URL: http://books.google.de/books?id=QqNpbTQwKXMC.