Markov processes

Andreas Eberle

March 15, 2015

Contents

Chapter 0

Introduction

0.1 Stochastic processes

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 version 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 $\mathbf{X}$**, 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 $(\mathcal{F}_\mathbf{t})$ **Markov process** iff (X_t) is adapted w.r.t. the filtration $(\mathcal{F}_t)_{t \in I}$, and

$$P[X_t \in B | \mathcal{F}_s] = P[X_t \in B | X_s] \quad P\text{-a.s. for any } B \in \mathcal{B} \text{ and } s, t \in I \text{ with } s \leq t. \tag{0.1.1}$$

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 (0.1.1) implies

$$P[X_t \in B | \mathcal{F}_s] = p_{s,t}(X_s, B) \quad P\text{-a.s. for } B \in \mathcal{B} \text{ and } s \leq t \text{ and} \tag{0.1.2}$$

$$E[f(X_t) | \mathcal{F}_s] = (p_{s,t} f)(X_s) \quad P\text{-a.s. for any measurable function } f : S \to \mathbb{R}_+ \text{ and } s \leq t. \tag{0.1.3}$$

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

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

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

$$p_{s,u}(X_s, B) = \int p_{s,t}(X_s, dy) p_{t,u}(y, B) \tag{0.1.4}$$

P-almost surely for any $B \in \mathcal{B}$ and $s \leq t \leq u$, i.e.,

$$p_{s,u} f = p_{s,t} p_{t,u} f \quad P \circ X_s^{-1}\text{-almost surely for any } 0 \leq s \leq t \leq u. \tag{0.1.5}$$

Exercise. Show that the consistency conditions (0.1.4) and (0.1.5) follow from the defining property (0.1.2) of the kernels $p_{s,t}$.

0.2 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 $f \in \mathcal{F}_b(S)$, and 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 (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 $p(x, S) = 1$ ($p(x, S) \leq 1$ 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 $s \leq t$) of sub-probability kernels on $(S, \mathcal{B})$ satisfying*

$$p_{t,t}(x, \cdot) = \delta_x \quad \text{for any } x \in S \text{ and } t \in I, \text{ and} \tag{0.2.1}$$

$$p_{s,t}p_{t,u} = p_{s,u} \quad \text{for any } s \leq t \leq u, \tag{0.2.2}$$

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 (0.2.2) are called the **Chapman-Kolmogorov equations**. They correspond to the consistency conditions in (0.1.4). 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.

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 $\sum_{y \in S} p(x, y) \leq 1$. 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

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

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, cf. XXX below.

Definition (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 $(\Omega,\mathfrak{A},P)$ is called an $(\mathcal{F}_\mathbf{t})$ **Markov process with transition function** $(p_{s,t})$ 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*

$$p_{s,t} = p_{t-s} \quad \textit{for any } s \le t.$$

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

$$p_{s+t} = p_s p_t \quad \text{for any } s,t \in I. \tag{0.2.3}$$

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 $s \in I$ the time-space process $\hat{X}_t = (s+t, X_{s+t})$ is a time-homogeneous Markov process with state space $\mathbb{R}_+ \times S$ and transition function

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

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,

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

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

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

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

Examples.

1) **Random dynamical systems:** A stochastic process on a probability space $(\Omega, \mathfrak{A}, P)$ defined recursively by

$$X_{n+1} = \Phi_{n+1}(X_n, W_{n+1}) \quad \text{for } n \in \mathbb{Z}_+ \tag{0.2.4}$$

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 $t \ge s$ denotes the solution of the recurrence relation (0.2.4) 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 any $n \in \mathbb{N}$.

2) **Continuous time Markov chains:** If $(Y_n)_{n\in\mathbb{Z}_+}$ is a **time-homogeneous Markov chain** on a polish space $(\Omega, \mathfrak{A}, P)$, and $(N_t)_{t\ge 0}$ 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. [10]. 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 3.1 below.

3) **Diffusion processes on $\mathbb{R}^n$:** 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

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dB_t, \ X_0 = x_0, \tag{0.2.5}$$

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.

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 0.1 (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 $(\Omega_{can}, \mathfrak{A}_{can})$ 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

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

for any $0 = t_0 \leq t_1 \leq \cdots \leq t_n$, the proof of Theorem 0.1 is a consequence of Kolmogorov's extension theorem (which follows from Carathéodory's extension theorem), cf. XXX. Thus Theorem 0.1 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 [32].

0.3 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 $t \geq 0$,

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

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

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

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

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

$$(\mathcal{L}_t f)(x) = \lim_{h\downarrow 0} \frac{(p_{t,t+h} f)(x) - f(x)}{h} = \lim_{h\downarrow 0} \frac{1}{h} E[f(X_{t+h}) - f(X_t)|X_t = x]. \tag{0.3.1}$$

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

$$\text{(FE)} \qquad \frac{d}{dt} p_{s,t} f = \frac{d}{dh} \left(p_{s,t} p_{t,t+h} f\right)|_{h=0} = p_{s,t}\mathcal{L}_t f, \quad \text{and} \tag{0.3.2}$$

$$\text{(BE)} \qquad -\frac{d}{ds} p_{s,t} f = -\frac{d}{dh} \left(p_{s,s+h} p_{s+h,t} f\right)|_{h=0} + p_{s,s}\mathcal{L}_s p_{s,t} f = \mathcal{L}_s p_{s,t} f. \tag{0.3.3}$$

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 (0.3.1) 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:

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 (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 $f \in \mathcal{A}$.

In the discrete time case, a process $((X_t), P)$ is a solution to the martingale problem w.r.t. the operator $\mathcal{L}_t = \pi_t - I$ with domain $\mathcal{A} = \mathcal{F}_b(S)$ if and only if it is a Markov chain with one-step transition kernels π_t. 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 3.1 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 (0.2.5) 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. [7]/[9].

0.4 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.,

$$\text{Law}(X_t) \to \mu \quad \text{as } t \to \infty \tag{0.4.1}$$

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.,

$$\frac{1}{t}\sum_{n=0}^{t-1} F(X_n, X_{n+1}, \dots) \to \int_{S^{\mathbb{Z}_+}} F dP_\mu, \tag{0.4.2}$$

$$\frac{1}{t}\int_0^t F(X_{s:\infty})ds \to \int_{\mathcal{D}(\mathbb{R}_+,S)} F dP_\mu \quad \text{respectively} \tag{0.4.3}$$

w.r.t. appropriate notions of convergence.

Statements as in (0.4.2) and (0.4.3) 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 (0.4.1), (0.4.2) and (0.4.3) 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 XXX 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.

We start in Sections 1.1 - 1.4 by applying martingale theory to Markov chains in discrete time. 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 Section 1.5 - 1.7 and 2.5 for discrete time chains, and the L^2/Dirichlet form approach considered in Chapter 4. Chapter 2 focuses on ergodic theorems and bounds for ergodic averages as in (0.4.2) and (0.4.3), and in Chapter 3 we introduce basic concepts and examples for Markov processes in continuous time and the relation to their generator. The concluding Chapter **??** studies a few selected applications to interacting particle systems. Other jump processes with infinite jump intensities (e.g. general Lévy processes) as well as jump diffusions will be constructed and analyzed in the stochastic analysis course.

Chapter 1

Markov chains & stochastic stability

1.1 Transition probabilities and Markov chains

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 1.1 (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. □

1.1.1 Markov chains

Now suppose that we are given a probability measure μ on $(S, \mathcal{B})$ and a sequence $p_1, p_2, \ldots$ of stochastic kernels on $(S, \mathcal{B})$. Recall that a stochastic process $X_n : \Omega \to S$ $(n \in \mathbb{Z}_+)$ defined on a probability space $(\Omega, \mathfrak{A}, P)$ is called an $(\mathcal{F}_\mathbf{n})$ **Markov chain** with initial distribution μ and transition kernels p_n iff (X_n) is adapted to the filtration $(\mathcal{F}_n)$, $X_0 \sim \mu$, and $p_{n+1}(X_n, \cdot)$ is a version of the conditional distribution of X_{n+1} given $\mathcal{F}_n$ for any $n \in \mathbb{Z}_+$. By iteratively applying Lemma 1.1, we see that w.r.t. the measure

$$P = \mu \otimes p_1 \otimes p_2 \otimes \cdots \otimes p_n \quad \text{ on } S^{\{0,1,\ldots,n\}}$$

the canonical process $X_k(\omega_0, \omega_1, \ldots, \omega_n) = \omega_k$ $(k = 0, 1, \ldots, n)$ is a Markov chain with initial distribution μ and transition kernels $p_1, \ldots, p_n$ (e.g. w.r.t. the filtration generated by the process). More generally, there exists a unique probability measure P_μ on

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

endowed with the product σ-algebra $\mathfrak{A}_{\text{can}}$ generated by the maps $X_n(\omega) = \omega_n$ $(n \in \mathbb{Z}_+)$ such that w.r.t. P_μ, the canonical process $(X_n)_{n \in \mathbb{Z}_+}$ is a Markov chain with initial distribution μ and transition kernels p_n. The probability measure P_μ can be viewed as the infinite product

$$\begin{aligned} P_\mu &= \mu \otimes p_1 \otimes p_2 \otimes p_3 \otimes \ldots \quad \text{,i.e.,} \\ P_\mu(dx_{0:\infty}) &= \mu(dx_0) p_1(x_0, dx_1) p_2(x_1, dx_2) p_3(x_2, dx_3) \ldots . \end{aligned}$$

We denote by $P_x^{(n)}$ the canonical measure for the Markov chain with initial distribution δ_x and transition kernels $p_{n+1}, p_{n+2}, p_{n+3}, \ldots$.

Theorem 1.2 (Markov properties). *Let $(X_n)_{n\in\mathbb{Z}_+}$ be a stochastic process with state space $(S,\mathcal{B})$ defined on a probability space $(\Omega,\mathfrak{A},P)$. Then the following statements are equivalent:*

(i) (X_n,P) *is a Markov chain with initial distribution* μ *and transition kernels* $p_1,p_2,\dots$.

(ii) $X_{0:n}\sim\mu\otimes p_1\otimes p_2\otimes\cdots\otimes p_n$ *w.r.t.* P *for any* $n\geq 0$.

(iii) $X_{0:\infty}\sim P_\mu$.

(iv) For any $n\in\mathbb{Z}_+$, $P^{(n)}_{X_n}$ *is a version of the conditional distribution of* $X_{n:\infty}$ *given* $X_{0:n}$*, i.e.,*

$$E[F(X_n,X_{n+1},\dots)|X_{0:n}]=E^{(n)}_{X_n}[F]\quad P\text{-a.s.}$$

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

*In the time homogeneous case, the properties (i)-(iv) are also equivalent to the **strong Markov property:***

(v) For any $(\mathcal{F}^X_n)$ *stopping time* $T:\Omega\to\mathbb{Z}_+\cup\{\infty\}$,

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

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

The proofs can be found in the lectures notes of Stochastic processes [10], Sections 2.2 and 2.3.

On a Polish state space S, any 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 $[XXX]$. Often such representations arise naturally:

Example. 1) **Random Walk on $\mathbb{R}^d$.** 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 $W_1,W_2,W_3,\dots:\Omega\to\mathbb{R}^d$ 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

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

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.

1.1.2 Markov chains with absorption

Given an arbitrary Markov chain and a possibly time-dependent absorption rate on the state space we can define another Markov chain that follows the same dynamics until it is eventually absorbed with the given rate. To this end we add an extra point Δ to the state space S where the Markov chain stays after absorption. Let $(X_n)_{n \in \mathbb{Z}_+}$ be the original Markov chain with state space

S and transition probabilities p_n, and suppose that the absorption rates are given by measurable functions $w_n : S \times S \to [0, \infty]$, i.e., the survival (non-absorption) 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 the absorbed chains with state space $S\dot{\cup}\Delta$ recursively by $X_0^w = X_0$,

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

Example (Absorption at 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 is absorbed completely when exiting the domain D for the first time.

Lemma 1.3 (Absorbed Markov chain). *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*

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

Proof. For any Borel subset B of S,

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

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. □

1.2 Generators and martingales

Let (X_n, P_x) be a time-homogeneous Markov chain with transition probability p and initial distribution $X_0 = x$ P_x-almost surely for any $x \in S$.

1.2.1 Generator

The average change of $f(X_n)$ in one transition step of the Markov chain starting at x is given by

$$(\mathcal{L}f)(x) = E_x[f(X_1) - f(X_0)] = \int p(x, dy)(f(y) - f(x)). \tag{1.2.1}$$

Definition (Generator of a time-homogeneous Markov chain).
The linear operator $\mathcal{L} : \mathcal{F}_b(S) \to \mathcal{F}_b(S)$ defined by (1.2.1) *is called the **generator** of the Markov chain (X_n, P_x).*

Examples. 1) **Simple random walk on $\mathbb{Z}$.** Here $p(x, \cdot) = \frac{1}{2}\delta_{x+1} + \frac{1}{2}\delta_{x-1}$. Hence the generator is given by

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

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 absorption.** Suppose that $\mathcal{L}$ is the generator of a time-homogeneous Markov chain with state space S. Then the generator of the corresponding Markov chain on $S\dot{\cup}\{\Delta\}$ with absorption rate $w(x, y)$ is given by

$$\begin{aligned}(\mathcal{L}^w f)(x) &= (p^w f)(x) - f(x) = p\left(e^{-w(x,\cdot)} f\right) - f(x)\\ &= \mathcal{L}\left(e^{-w(x,\cdot)} f\right)(x) + \left(e^{-w(x,x)} - 1\right) f(x)\end{aligned}$$

for any bounded function $f : S \cup \{\Delta\} \to \mathbb{R}$ with $f(0) = 0$, and for any $x \in S$.

1.2.2 Martingale problem

The generator can be used to identify martingales associated to a Markov chain. Indeed if (X_n, P) is an $(\mathcal{F}_n)$ Markov chain with transition kernel p then for $f \in \mathcal{F}_b(S)$,

$$E\left[f(X_{k+1}) - f(X_k)|\mathcal{F}_k\right] = E_{X_k}\left[f(X_1) - f(X_0)\right] = (\mathcal{L}f)(X_k) \quad P\text{-a.s. } \forall k \geq 0.$$

Hence the process $M^{[f]}$ defined by

$$M_n^{[f]} = f(X_n) - \sum_{k=0}^{n-1} (\mathcal{L}f)(X_k), \quad n \in \mathbb{Z}_+, \tag{1.2.2}$$

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

Theorem 1.4 (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 kernel p if and only if $M^{[f]}$, defined by* (1.2.2) *is an $(\mathcal{F}_n)$ martingale for any function $f \in \mathcal{F}_b(S)$.*

The proof is left as an exercise.

The result in Theorem 1.4 can be extended to the time-inhomogeneous case. Indeed, if (X_n, P) is an inhomogeneous Markov chain with state space S and transition kernels $p_n, \ n \in \mathbb{N}$, then the time-space process $\hat{X}_n := (n, X_n)$ is a time-homogeneous Markov chains with state space $\mathbb{Z}_+ \times S$. Let

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

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

Corollary 1.5 (Time-dependent martingale problem). *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 $p_1, p_2, \ldots$ 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 bounded 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 p_n if and only if the time-space process $((n, X_n), P)$ is a time-homogeneous Markov chain with transition kernel

$$\hat{p}((n,x),\cdot) = \delta_{n+1} \otimes p_{n+1}(x,\cdot).$$

The assertion now follows from Theorem 1.4. □

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 (Non-negative supermartingale). *A real-valued stochastic process (M_n, P) is called a* ***non-negative supermartingale*** *w.r.t. a filtration $(\mathcal{F}_n)$ if and only if for any $n \in \mathbb{Z}_+$,*

(i) $M_n \geq 0 \quad P$*-almost surely,*

(ii) M_n *is* $\mathcal{F}_n$*-measurable, and*

(iii) $E[M_{n+1}|\mathcal{F}_n] \leq M_n \quad P$*-almost surely.*

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

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

holds for an **arbitrary** $(\mathcal{F}_n)$ stopping time $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$. Similarly, the limit

$$M_\infty = \lim_{n\to\infty} M_n$$

exists almost surely in $[0, \infty)$.

1.2.3 Potential theory for Markov chains

Let (X_n, P_x) be a canonical time-homogeneous Markov chain with state space $(S, \mathcal{B})$ and generator

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

By Theorem 1.4,

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

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 $\mathcal{L}f \leq -c$ 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

$$\partial D = \bigcup_{x \in D} \text{supp}\; p(x, \cdot) \setminus D$$

where the support supp(μ) 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.

Examples. (1). For the simple random walk on $\mathbb{Z}^d$, the exterior boundary of a subset $D \subset \mathbb{Z}^d$ is given by

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

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

$$p(x, \cdot) = \text{Unif}\left(B(x, r)\right),$$

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

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

Let

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

denote the first exit time from D. Then

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

Our aim is to compute or bound expectations of the form

$$u(x) = E_x\left[e^{-\sum\limits_{n=0}^{T-1} w(X_n)} f(X_T); T < \infty\right] + E_x\left[\sum_{n=0}^{T-1} e^{-\sum\limits_{i=0}^{n-1} w(X_i)} c(X_n)\right] \tag{1.2.3}$$

for given non-negative measurable functions $f : \partial D \to \mathbb{R}_+, c, w : D \to \mathbb{R}_+$. The general expression (1.2.3) combines a number of important probabilities and expectations related to the Markov chain:

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

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

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

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

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

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

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

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

$$\begin{aligned} u(x) &= G_D(x,B) \quad \text{where} \\ G_D(x,B) &= E_x\left[\sum_{n=0}^{T-1} 1_B(X_n)\right] = \sum_{n=0}^{\infty} P_x[X_n \in B, n < T]. \end{aligned}$$

G_D is called the **potential kernel** or **Green kernel** of the domain D, it is a kernel of positive measure.

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

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

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

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

The next fundamental theorem shows that supersolutions to an associated boundary value problem provide upper bounds for expectations of the form (1.2.3). This observation is crucial for studying stability properties of Markov chains.

Theorem 1.6 (Maximum principle). *Suppose $v \in \mathcal{F}_+(S)$ is a non-negative function satisfying*

$$\begin{aligned} \mathcal{L}v &\leq (e^w - 1)v - e^w c \quad &\text{on } D, \\ v &\geq f &\text{on } \partial D. \end{aligned} \tag{1.2.4}$$

Then $u \leq v$.

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

Theorem 1.7 (Dirichlet problem, Poisson equation, Feynman-Kac formula). *The function u is the **minimal non-negative solution** of the boundary value problem*

$$\begin{aligned} \mathcal{L}v &= (e^w - 1)v - e^w c \quad \text{on } D, \\ v &= f \qquad\qquad \text{on } \partial D. \end{aligned} \tag{1.2.5}$$

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 (1.2.5). We first prove both theorems in the case $w \equiv 0$. The extension to the general case will be discussed afterwards.

Proof of Theorem 1.6 for $w \equiv 0$: Let $v \in \mathcal{F}_+(S)$ such that $\mathcal{L}v \leq -c$ on D. Then the process

$$M_n = v(X_n) + \sum_{i=0}^{n-1} c(X_i)$$

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

$$M_T \geq f(X_T) 1_{\{T<\infty\}} + \sum_{i=0}^{T-1} c(X_i). \tag{1.2.6}$$

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

$$u(x) \leq E_x[M_T] \leq E_x[M_0] = v(x) \tag{1.2.7}$$

□

Proof of Theorem 1.7 for $w \equiv 0$: By Theorem 1.6, all non-negative solutions v of (1.2.5) 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 (1.2.6) and (1.2.7) are equalities. Thus if a non-negative solution of (1.2.5) exists then it coincides with u, i.e., uniqueness holds.

It remains to verify that u satisfies (1.2.4). 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

$$\begin{aligned}
&E_x\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T} c(X_n)|X_1\right] \\
&= c(x) + E_x\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T-1} c(X_{n+1})|X_1\right] \\
&= c(x) + E_{X_1}\left[f(X_T)1_{\{T<\infty\}} + \sum_{n<T} c(X_n)\right] \\
&= c(x) + u(X_1) \quad P_x\text{-almost surely,} \\
&\text{and hence} \\
&u(x) = E_x\left[c(x) + u(X_1)\right] = c(x) + (pu)(x), \\
&\text{i.e.,} \quad \mathcal{L}u(x) = -c(x).
\end{aligned}$$

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 (1.2.5) as a corresponding expectation with $w \equiv 0$ for an absorbed Markov chain:

Reduction of general case to $w \equiv 0$: We consider the Markov chain (X_n^w) with absorption 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 $\mathrm{Exp}(1)$ 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 Theorem 1.6 and 1.7 with $w \equiv 0$ to the Markov chain (X_n^w), we see that u is the minimal non-negative solution of

$$\mathcal{L}^w u = -c \quad \text{on } D, \quad u = f \quad \text{on } \partial D, \tag{1.2.8}$$

and any non-negative supersolution v of (1.2.8) dominates u from above. Moreover, the boundary value problem (1.2.8) is equivalent to (1.2.5) since

$$\begin{aligned} &\mathcal{L}^w u = e^{-w} pu - u = e^{-w}\mathcal{L}u + (e^{-w} - 1)u = -c \\ &\text{if and only if} \quad \mathcal{L}u = (e^w - 1)u - e^w c. \end{aligned}$$

This proves Theorem 1.6 and the main part of Theorem 1.7 in the case $w \not\equiv 0$. The proof of the last assertion of Theorem 1.7 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., $|W| \leq r$ for some constant $r \in (0, \infty)$. Our goal is to derive tail estimates for passage times.

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

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 $\lambda \leq 0$, the function $u(x) = e^{\lambda x}$ satisfies

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

By applying Theorem 1.6 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

$$E_x\left[Z(\lambda)^{-T_a} e^{\lambda(a+r)}; T < \infty\right] \leq e^{\lambda x} \quad \forall x \in \mathbb{R} \tag{1.2.9}$$

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, (1.2.9) yields the exponential moment bound

$$E_x\left[\left(\frac{1}{Z(\lambda)}\right)^{T_a}\right] \leq e^{-\lambda(a+r-x)} \tag{1.2.10}$$

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:

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

(ii) $E[W] = 0$: In this case, we may have $Z(\lambda) \geq 1$ 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. [Eberle:Stochastic processes] [10]. 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

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

where $C(\lambda) := E[cos(\lambda W)]$, and $cos(\lambda x) \geq cos\,(\lambda(a + r)) > 0$ 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

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

1.3 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 sections. As before we consider a time-homogeneous Markov chain (X_n, P_x) with generator $\mathcal{L} = p - I$ on a Polish state space S endowed with the Borel σ-algebra $\mathcal{B}$. We start with the following simple observation:

Lemma 1.8. ***[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 some constant $c \geq 0$. Then the process

$$M_n = V(X_{n\wedge T_A}) + c \cdot (n \wedge T_A) \tag{1.3.1}$$

is a non-negative supermartingale.

The elementary proof is left as an exercise.

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

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

*It is called **positive recurrent** iff*

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

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: Markov Chains and Stochastic Stability] [23]. 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

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

(LR) There exists a function $V \in F_+(S)$ such that

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

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

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

Theorem 1.9. ***(Foster-Lyapunov conditions for non-recurrence, Harris recurrence and positive recurrence)***

(1). If (LT) holds then

$$P_y[T_A < \infty] \leq V(y)/\inf_A V < 1.$$

(2). If (LR) holds then

$$P_x[T_A < \infty] = 1 \quad \textit{for any } x \in S.$$

In particular, the set A is Harris recurrent.

(3). If (LP) holds then

$$E_x[T_A] \leq V(x) < \infty \quad \textit{for any } x \in A^c, \textit{ and}$$
$$E_x[T_A^+] \leq (pV)(x) < \infty \quad \textit{for any } x \in A.$$

In particular, the set A is positive recurrent.

Proof: (1). If $\mathcal{L}V \leq 0$ on A^c then by Lemma 1.8 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 $P_y[T_A < \infty] < 1$.

(2). 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.

(3). Finally, suppose that $\mathcal{L}V \leq -1$ on A^c. Then by Lemma 1.8,

$$M_n = V(X_{n\wedge T_A}) + n \wedge T_A$$

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 $V \geq 0$,

$$E_x[T_A] \leq E_x[M_{T_A}] \leq E_x[M_0] = V(x) \quad \text{for any } x \in S. \tag{1.3.2}$$

Moreover, we can also estimate the first return time by conditioning on the first step. Indeed, for $x \in A$ we obtain by (1.3.2):

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

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

$$x \mapsto x + b(x) + W$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable vector field and $W : \Omega \to \mathbb{R}^d$ is a square-integrable random vector with $E[W] = 0$ and $Cov(W^i, W^j) = \delta_{ij}$. As a Lyapunov function we try

$$V(x) = |x|/\varepsilon \quad \text{for some constant } \varepsilon > 0.$$

A simple calculation shows that

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

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

$$2x \cdot b(x) + |b(x)|^2 + d \leq -\varepsilon.$$

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

$$\limsup_{|x| \to \infty} \left(2x \cdot b(x) + |b(x)|^2\right) < -d. \tag{1.3.3}$$

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 $(1 - \delta)b_r(x) \leq -\frac{d}{2|x|}$ for some $\delta > 0$, and $|b(x)|^2/r \leq -\delta \cdot b_r(x)$.

Exercise. Derive a sufficient condition similar to (1.3.3) for positive recurrence of state space models with transition step

$$x \mapsto x + b(x) + \sigma(x)W$$

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 $p(x,y) = \frac{1}{2d}$ if $|x-y| = 1$ and $p(x,y) = 0$ 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 $f \in C^4(\mathbb{R}^d)$,

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

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

$$\Delta_{\mathbb{Z}^d} f(x) = \Delta f(x) + R(x), \text{ where} \tag{1.3.4}$$

$$|R(x)| \le \frac{d}{12}\sup_{B(x,1)} \|\partial^4 f\|. \tag{1.3.5}$$

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

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

By the expression for the Laplacian in polar coordinates,

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

where $r = |x|$. In particular, V is superharmonic on $\mathbb{R}^d$ 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

$$\|\partial^4 V(x)\| \leq C \cdot |x|^{p-4} \quad \text{(Exercise).}$$

This bound shows that the approximation of the discrete Laplacian by the Laplacian on $\mathbb{R}^d$ improves if $|x|$ is large. Indeed by (1.3.4) and (1.3.5) we obtain

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

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 $\mathcal{L}V \leq 0$ 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 1.9,

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

Theorem 1.10 below shows that this implies that any 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 $\mathcal{L}V \leq 0$ 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

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

Therefore, by Theorem 1.9, 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 1.10 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 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). A continuous-time stochastic process $\big((B_t)_{t\in[0,\infty)}, P_x\big)$ taking values in $\mathbb{R}^d$ is called a *Brownian motion starting at x* if the sample paths $t \mapsto B_t(\omega)$ are continuous, $B_0 = x$ P_x-a.s., and for every $f \in C_b^2(\mathbb{R}^d)$, the process

$$M_t^{[f]} = f(B_t) - \frac{1}{2}\int_0^t \Delta f(B_s)ds$$

is a martingale w.r.t. the filtration $\mathcal{F}_t^B = \sigma(B_s : s \in [0,t])$. Let $T_a = \inf\{t \geq 0 : |B_t| = a\}$.

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

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

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

You may assume the optional stopping theorem and the martingale convergence theorem in continuous time without proof. You may also assume that the Laplacian applied to a rotationally symmetric function $g(x) = \gamma(|x|)$ *is given by*

$$\Delta g(x) = r^{1-d}\frac{d}{dr}\left(r^{d-1}\frac{d}{dr}\gamma\right)(r) = \frac{d^2}{dr^2}\gamma(r) + \frac{d-1}{r}\frac{d}{dr}\gamma(r) \quad \textit{where } r = |x|.$$

(How can you derive this expression rapidly if you do not remember it?)

1.3.2 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 any 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, [23].

a) Countable state space

Suppose that $p(x,y) = p(x,\{y\})$ 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 (Irreducibility on countable state spaces). *The transition matrix p and the Markov chain (X_n, P_x) are called* ***irreducible*** *if and only if*

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

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

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 1.10 (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 $A \subset S$.*
- *(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", [10]. 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 but not positive recurrent. For $d \geq 3$ it is transient.
As a consequence of Theorem 1.10, we obtain Lyapunov conditions for transience, recurrence and positive recurrence on a countable state space that are both necessary and sufficient:

Corollary 1.11 (Foster-Lyapunov conditions for 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 $A \subset S$ 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 $A \subset S$ and a function $V \in \mathcal{F}_+(S)$ such that*

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

3) *The Markov chain is positive recurrent if and only if there exists a finite set $A \subset S$ and a function $V \in F_+(S)$ such that (LP) holds.*

Proof: Sufficiency of the Lyapunov conditions follows directly by Theorems 1.9 and 1.10: If (LT) holds then by 1.9 there exists $y \in S$ such that $P_y[T_A < \infty]$, and hence the Markov chain is transient by 1.10. Similarly, if (LP) holds then A is positive recurrent by 1.9, and hence global positive recurrence holds by 1.10. Finally, if (LR') holds and the state space is not finite, then for any $c \in \mathbb{R}_+$, the set $\{V \leq c\}$ is not empty. Therefore, (LR) holds by irreducibility, and the recurrence follows again from 1.9 and 1.10. If S is finite then any 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 $A \subset S$ such that the function $V(x) = P_x[T_A < \infty]$ satisfies

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

By Theorem 1.7, 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 $A \subset S$ and an arbitrary decreasing sequence of sets $D_n \subset S$ such that $A \subset D_1^c$, D_n^c is finite for any 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 $n \to \infty$,

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

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

$$V(x) := \sum_{n=0}^{\infty} V_{n_k}(x) < \infty \quad \text{for any } x \in S.$$

By Theorem 1.7, the functions V_n and V are harmonic on $S \setminus A$. Moreover, $V \geq k$ 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 $A \subset S$, the function $V(x) = E_x[T_A]$ is finite and satisfies $\mathcal{L}V = -1$ on A^c. Since

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

for any x, condition (LP) is satisfied.

□

b) Extension to locally compact state spaces

Extensions of Corollary 1.11 to general state spaces are not trivial. Suppose for example that S is **locally compact**, i.e., there exists a sequence of compact sets $K_n \subset S$ such that $S = \bigcup_{n\in\mathbb{N}} K_n$. Let p 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 the Lebesgue measure.

Definition (**λ-irreducibility and Feller property).**

*1) The transition kernel p 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 $p^n(x, A) > 0$.*

*2) p is called **Feller** iff*

(F) $\quad pf \in C_b(S) \quad$ *for any* $f \in C_b(S)$

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: Markov Chains and Stochastic Stability] [23]:

Theorem 1.12 (Necessary and sufficient conditions for Harris recurrence on a locally compact state space). *Suppose that p is a λ-irreducible Feller transition kernel on $(S, \mathcal{B})$. 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

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

(ii) There exists a compact set $K \subset S$ such that K is Harris recurrent.

(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 idea of the proof is to show at first that if p is λ-irreducible and Feller then for any 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

$$\sum_{n=0}^{\infty} a_n p^n(x, \cdot) \geq \varepsilon\nu \qquad (1.3.6)$$

holds for any $x \in K$. In the theory of Markov chain 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 start it visits a given set A satisfying $\lambda(A) > 0$ with a fixed strictly positive probability.

Example (Recurrence of Markov chains on $\mathbb{R}$).

1.4 The space of probability measures

Our next goal is to study convergence of Markov chains to stationary distributions. To this end we 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 $\mathcal{B}$. We study and apply weak convergence of probability measures in this section, and we consider Wasserstein and total variation metrics in the next two sections. A useful additional reference for this section is [Billingsley:Convergence of probability measures] [2]. 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 (Invariant measures, stationary distribution). *A measure $\mu \in \mathcal{M}_+(S)$ is called* ***invariant*** *w.r.t. a transition kernel p on $(S, \mathcal{B})$ iff $\mu p = \mu$, i.e., iff*

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

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

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

1.4.1 Weak topology

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. [Stroock:Probability Theory: An Analytic View] [34], Theorem 3.1.5, or [Billingsley:Convergence of probability measures] [2]. 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

$$h : \begin{array}{ccc} S & \to & [0,1]^{\mathbb{N}} \\ x & \to & \left(\frac{\varrho(x,x_n)}{1+\varrho(x,x_n)}\right)_{n\in\mathbb{N}} \end{array} \tag{1.4.1}$$

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, $h(S)$ 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,

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

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: 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

$$\Lambda(f) := \lim_{l\to\infty}\mu_{k_l}(f) \tag{1.4.2}$$

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 (1.4.2), the sequence (μ_{k_l}) converges weakly to μ. □

Remark (A metric for weak convergence). Choosing the function 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 any $n\in\mathbb{N}$. Thus weak convergence in $\mathcal{P}(S)$ is equivalent to convergence w.r.t. the metric

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

1.4.2 Prokhorov's theorem

We now consider the case where S is a non-compact polish space. By identifying S with the image $h(S)$ under the map h defined by (1.4.1), we can still view S as a measurable subset of the compact space $\hat{S}$:

$$S\subset\hat{S}\subset[0,1]^{\mathbb{N}}.$$

Hence $\mathcal{P}(S)$ can be viewed as a subset of the compact space $\mathcal{P}(\hat{S})$:

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

If μ_k $(k\in\mathbb{N})$ and μ are probability measures on S (that trivially extend to $\hat{S}$) then:

$$\begin{aligned} &\mu_k\to\mu \text{ weakly in } \mathcal{P}(S) \qquad (1.4.3)\\ \Leftrightarrow\ &\mu_k(f)\to\mu(f) \text{ for any uniformly continuous } f\in C_b(S)\\ \Leftrightarrow\ &\mu_k(f)\to\mu(f) \text{ for any } f\in C(\hat{S})\\ \Leftrightarrow\ &\mu_k\to\mu \text{ weakly in } \mathcal{P}(\hat{S}). \end{aligned}$$

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 $\hat{S}$ s.t. $\mu(S) < 1$. To exclude this possibility, the following tightness condition is required:

Definition (Tightness of collections of probability measures). *Let $\mathcal{R} \subset \mathcal{P}(S)$ 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 $K \subset S$ such that*

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

*Thus tightness means that the measures in the set $\mathcal{R}$ are concentrated uniformly on a compact set up to an arbitrary small positive amount of mass. A set $\mathcal{R} \subset \mathcal{P}(S)$ is called **relatively compact** iff every sequence in $\mathcal{R}$ has a subsequence that converges weakly in $\mathcal{P}(S)$.*

Theorem 1.14 (Prokhorov). *Suppose that S is polish, and let $\mathcal{R} \subset \mathcal{P}(S)$. Then*

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

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:Convergence of probability measures] [2].

Proof of "$\Leftarrow$": Let $(\mu_k)_{k \in \mathbb{N}}$ be a sequence in $\mathcal{R}$. We have to show that (μ_k) has a weakly convergent subsequence in $\mathcal{P}(S)$. Since $\mathcal{P}(\hat{S})$ is compact by Theorem 1.13, 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 Portmanteau Theorem,

$$\mu(K) \geq \limsup_{l \to \infty} \mu_{k_l}(K) \geq 1 - \varepsilon,$$

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 $\mu \in \mathcal{P}(S)$ and $\mu_{k_l} \to \mu$ weakly in $\mathcal{P}(S)$ by (1.4.3). $\square$

1.4.3 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 $p(x, dy)$ on $(S, \mathcal{B})$.

Definition (Feller property). *The stochastic kernel p is called **(weakly) Feller** iff pf is continuous for any $f \in C_b(S)$.*

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

$$x_n \to x \Rightarrow (pf)(x_n) \to (pf)(x) \quad \forall f \in C_b(S).$$

A topological space is said to be σ**-compact** iff it is the union of countably many compact subsets. For example, $\mathbb{R}^d$ is σ-compact whereas an infinite dimensional Hilbert space is not σ-compact.

Theorem 1.15 (Foguel, Krylov-Bogolionbov). *Suppose that p is a Feller transition kernel on the Polish space S, and let $\overline{p}_n := \frac{1}{n}\sum_{i=0}^{n-1} p^i$. Then there exists an invariant probability measure μ of p if one of the following conditions is satisfied for some $x \in S$:*

(i) The sequence $\{\overline{p}_n(x, \cdot) : n \in \mathbb{N}\}$ is tight, or

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

$$\liminf_{n\to\infty} \overline{p}_n(x, K) > 0.$$

Remark. If (X_n, P_x) is a canonical Markov chain with transition kernel p then

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

is the average proportion of time spent by the chain in the set K during the first n steps. Conditions (i) and (ii) say that

(i) $\forall \varepsilon > 0 \, \exists K \subset S$ compact: $\overline{p}_n(x, K) \geq 1 - \varepsilon$ for all $n \in \mathbb{N}$,

(ii) $\exists \varepsilon > 0, K \subset S$ compact, $n_k \nearrow \infty$: $\overline{p}_{n_k}(x, K) \geq \varepsilon$ for all $k \in \mathbb{N}$.

Clearly, the second condition is weaker than the first one in several respects.

Proof of Theorem 1.15: (i) Suppose that the sequence $\nu_n := \overline{p}_n(x, \cdot)$ is tight for some $x \in S$. Then 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 $\mu p = \mu$. Indeed for, $f \in C_b(S)$ we have $pf \in C_b(S)$ by the Feller property. Therefore,

$$\begin{aligned}(\mu p)(f) = \mu(pf) &= \lim_{k\to\infty} \nu_{n_k}(pf) = \lim_{k\to\infty} (\nu_{n_k} p)(f) \\ &= \lim_{k\to\infty} \nu_{n_k}(f) = \mu(f) \quad \text{for any } f \in C_b(S),\end{aligned}$$

where the second last equality holds since

$$\nu_{n_k} p = \frac{1}{n_k} \sum_{i=0}^{n_k - 1} p^{i+1}(x, \cdot) = \nu_{n_k} - \frac{1}{n_k}\delta_x + \frac{1}{n_k} p^{n_k}(x, \cdot).$$

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

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

Note that weak convergence in $\hat{S}$ does not imply weak convergence in S. However $\nu_{n_k}(f) \to \mu(f)$ for any compactly supported function $f \in C(S)$, and

$$\hat{\mu}(S) \geq \hat{\mu}(K) \geq \limsup \nu_{n_k}(K) \geq \varepsilon.$$

Therefore, it can be verified similarly as above that the conditioned measure

$$\mu(B) = \frac{\hat{\mu}(B \cap S)}{\hat{\mu}(S)} = \hat{\mu}(B|S), \quad B \in \mathcal{B}(S),$$

is an invariant probability measure for p.

□

In practice, the assumptions in Theorem 1.15 can be verified via appropriate Lyapunov functions:

Corollary 1.16 (Lyapunov condition for the existence of an invariant probability measure). *Suppose that p is a Feller transition kernel and S is σ-compact. Then an invariant probability measure for p exists if the following Lyapunov condition is satisfied:*

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

$$\mathcal{L}V \leq c 1_K - \varepsilon.$$

Proof: By (LI),

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

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

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

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

$$\liminf_{n\to\infty} \overline{p}_n(x, K) \geq \varepsilon \qquad \text{for any } x \in S.$$

The assertion now follows by Theorem 1.15. □

Example. 1) **Countable state space:** If S is countable and p is irreducible then an invariant probability measure exists if and only if the Markov chain is positive recurrent. On the other hand, by Corollary 1.11, 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) $\mathbf{S = \mathbb{R}^d}$: On $\mathbb{R}^d$, Condition (LI) is satisfied in particular if $\mathcal{L}V$ is continuous and

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

1.5 Couplings and transportation metrics

Additional reference: [Villani:Optional transport-old and new] [37].

Let S be a Polish space endowed with its Borel σ-algebra $\mathcal{B}$. An invariant probability measure is a fixed point of the map $\mu \mapsto \mu p$ 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.

1.5.1 Wasserstein distances

We fix a metric $d : S \times S \to [0,\infty)$ on the state space S. 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 (Coupling of probability measures). *A **coupling** of measures $\mu, \nu \in \mathcal{P}^p(S)$ 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 (Wasserstein distance, Kantorovich distance). *For $p \in [1, \infty)$, the* **Lᵖ** ***Wasserstein distance** of probability measures $\mu, \nu \in \mathcal{P}(S)$ is defined by*

$$\mathcal{W}^p(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \left(\int d(x, y)^p \gamma(dxdy) \right)^{\frac{1}{p}} = \inf_{\substack{X \sim \mu \\ Y \sim \nu}} E\left[d(X, Y)^p\right]^{\frac{1}{p}}, \tag{1.5.1}$$

*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 (1.5.1) is a particular case of an optimal transport problem. Given a cost function $c : S \times S \to [0, \infty]$, one is either looking for a map $T : S \to S$ minimizing the average cost

$$\int c(x, T(x))\mu(dx)$$

under the constraint $\nu = \mu \circ T^{-1}$ (**Monge problem**, 8^{th} century), or, less restrictively, for a coupling $\gamma \in \Pi(\mu, \nu)$ minimizing

$$\int c(x, y)\gamma(dxdy)$$

(**Kantorovich problem**, around 1940).

Note that the definition of the $\mathcal{W}^p$ distance depends in an essential way on the distance d considered on S. In particular, we can create different distances on probability measures by modifying the underlying metric. For example, if $f : [0, \infty) \to [0, \infty)$ 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 $U \sim \mathrm{Unif}(0,1)$ 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

$$X = F_\mu^{-1}(U) \text{ and } Y = F_\nu^{-1}(U), \quad \text{where } U \sim \mathrm{Unif}(0,1).$$

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 $\mathcal{W}^p$ distance** for any $p \geq 1$, i.e.,

$$\mathcal{W}^p(\mu,\nu) = E\left[|X-Y|^p\right]^{\frac{1}{p}} = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^p(0,1)},$$

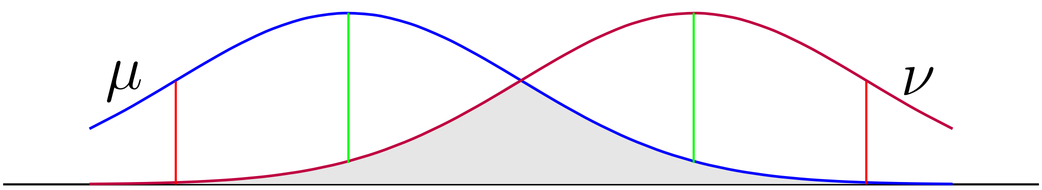
cf. e.g. [Rachev&Rueschendorf] [25]. On the other hand, the coupling by monotone rearrangement is **not optimal w.r.t. $\mathcal{W}_f$ if f is strictly concave**. Indeed, consider for example $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ and $\nu = \frac{1}{2}(\delta_0 + \delta_{-1})$. Then the coupling above 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 any concave 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 1.6. 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

$$\begin{aligned} \mu &= (\mu \wedge \nu) + (\mu - \nu)^+ = p\alpha + (1-p)\beta, \\ \nu &= (\mu \wedge \nu) + (\mu - \nu)^- = p\alpha + (1-p)\gamma \end{aligned}$$

of probability measures α, β and γ. Hence a coupling (X,Y) of μ and ν as described above is given by setting

$$(X,Y) = \begin{cases} (F_\alpha^{-1}(U), F_\alpha^{-1}(U)) & \text{if } B = 1, \\ \big(F_\beta^{-1}(U), F_\gamma^{-1}(1-U)\big) & \text{if } B = 0, \end{cases}$$

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 any concave f, cf. [McCann:Exact solution to the transportation problem on the line] [22].

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 1.17 (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) 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

$$\mu(S \setminus K) < \frac{\varepsilon}{2}, \quad \nu(S \setminus K) < \frac{\varepsilon}{2},$$

and hence for any $n \in \mathbb{N}$,

$$\begin{aligned}\gamma_n\left((x,y) \notin K \times K\right) &\leq \gamma_n(x \notin K) + \gamma_n(y \notin K) \\ &= \mu(S \setminus K) + \nu(S \setminus K) < \varepsilon.\end{aligned}$$

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,

$$I(\gamma) = \int d(x,y)^p \gamma(dxdy) \leq \liminf_{k\to\infty} \int d(x,y)^p \gamma_{n_k}(dxdy) = \mathcal{W}^p(\mu,\nu)^p$$

by the portemanteau Theorem. □

Lemma 1.18 (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

$$\mathcal{W}^p(\mu,\varrho) \leq \mathcal{W}^p(\mu,\nu) + \mathcal{W}^p(\nu,\varrho). \tag{1.5.2}$$

The other properties of a metric can be verified easily. To prove (1.5.2) let γ and $\widetilde{\gamma}$ be couplings of μ and ν, ν and ϱ respectively. We show

$$\mathcal{W}^p(\mu,\varrho) \leq \left(\int d(x,y)^p \gamma(dxdy)\right)^{\frac{1}{p}} + \left(\int d(y,z)^p \widetilde{\gamma}(dydz)\right)^{\frac{1}{p}}. \tag{1.5.3}$$

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

$$\gamma(dxdy) = \mu(dx)p(x,dy) \text{ and } \widetilde{\gamma}(dydz) = \nu(dy)\overline{p}(y,dz)$$

where p and $\overline{p}$ are regular versions of conditional distributions of the first component w.r.t. $\gamma, \widetilde{\gamma}$ given the second component. The disintegration enables us to "glue" the couplings γ and $\widetilde{\gamma}$ to a joint coupling

$$\hat{\gamma}(dxdydz) := \mu(dx)p(x,dy)\overline{p}(y,dz)$$

of the measures μ, ν and ϱ such that under $\hat{\gamma}$,

$$(x,y) \sim \gamma \quad \text{and} \quad (y,z) \sim \widetilde{\gamma}.$$

Therefore, by the triangle inequality for the L^p norm, we obtain

$$\begin{aligned}\mathcal{W}^p(\mu,\varrho) &\leq \left(\int d(x,z)^p \hat{\gamma}(dxdydz)\right)^{\frac{1}{p}} \\ &\leq \left(\int d(x,y)^p \hat{\gamma}(dxdydz)\right)^{\frac{1}{p}} + \left(\int d(y,z)^p \hat{\gamma}(dxdydz)\right)^{\frac{1}{p}} \\ &= \left(\int d(x,y)^p \gamma(dxdy)\right)^{\frac{1}{p}} + \left(\int d(y,z)^p \widetilde{\gamma}(dydz)\right)^{\frac{1}{p}}.\end{aligned}$$

□

Exercise (Couplings in $\mathbb{R}^d$). Let $W : \Omega \to \mathbb{R}^d$ be a random variable on $(\Omega, \mathcal{A}, P)$ with $W \sim -W$, 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) Let $\widetilde{Y} = \widetilde{W} + b$ where $\widetilde{W} \equiv W - 2e \cdot W\, e$ with $e = \frac{a-b}{|a-b|}$. Prove that $(X, \widetilde{Y})$ is also a coupling of μ_a and μ_b, and if $|W| \leq \frac{|a-b|}{2}$ a.s. 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 $f(0) = 0$.

1.5.2 Kantorovich-Rubinstein duality

The **Lipschitz 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 estimate differences of integrals of Lipschitz continuous functions w.r.t. different probability measures. Indeed, one even has:

Theorem 1.19 (Kantorovich-Rubinstein duality). *For any $\mu, \nu \in \mathcal{P}(S)$,*

$$\mathcal{W}^1(\mu, \nu) = \sup_{\|g\|_{Lip} \leq 1} \left(\int g d\mu - \int g d\nu \right). \tag{1.5.4}$$

Remark. There is a corresponding dual description of $\mathcal{W}^p$ for $p > 1$ but it takes a more complicated form, cf. [Villani:OT-old&new] [37].

Proof: We only prove the easy "$\geq$" part. For different proofs of the converse inequality see Rachev and Rueschendorf [25], Villani1 [38], Villani2 [37] and Mufa Chen [4]. For instance one can approximate μ and ν by finite convex combinations of Dirac measures for which (1.5.4) is a

consequence of the standard duality principle of linear programming, cf. Chen [4].
To prove "$\geq$" let $\mu, \nu \in \mathcal{P}(S)$ and $g \in C(S)$. If γ is a coupling of μ and ν then

$$\int g d\mu - \int g d\nu = \int (g(x) - g(y))\gamma(dxdy)$$
$$\leq \|g\|_{\text{Lip}} \int d(x,y)\gamma(dxdy).$$

Hence, by taking the infimum over $\gamma \in \Pi(\mu, \nu)$, we obtain

$$\int g d\mu - \int g d\nu \leq \|g\|_{\text{Lip}} \mathcal{W}^1(\mu, \nu).$$

□

As a consequence of the "$\geq$" part of (1.5.4), 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 any Lipschitz continuous function $g : S \to \mathbb{R}$, and hence $\mu_n \to \mu$ weakly. The following more general statement connects convergence in Wasserstein distances and weak convergence:

Theorem 1.20 (**$\mathcal{W}^p$ 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*

$$\int g d\mu_n \to \int g d\mu \quad \text{for any } g \in C(S) \text{ satisfying } g(x) \leq C \cdot (1 + d(x, x_o)^p)$$
$$\text{for a finite constant } C \text{ and some } x_0 \in S.$$

Among other things, the proof relies on Prokhorov's Theorem - we refer to [Villani:OT-old&new] [37].

1.5.3 Contraction coefficients

Let $p(x, dy)$ be a transition kernel on $(S, \mathcal{B})$ and fix $q \in [1, \infty)$. We will be mainly interested in the case $q = 1$.

Definition (Wasserstein contraction coefficient of a transition kernel). *The global* ***contraction coefficient*** *of p w.r.t. the distance $\mathcal{W}^q$ is defined as*

$$\alpha_q(p) = \sup \left\{ \frac{\mathcal{W}^q(\mu p, \nu p)}{\mathcal{W}^q(\mu, \nu)} : \mu, \nu \in \mathcal{P}^q(S) s.t. \mu \neq \nu \right\}.$$

In other words, $\alpha_q(p)$ is the Lipschitz norm of the map $\mu \mapsto \mu p$ w.r.t. the $\mathcal{W}^q$ distance. By applying the Banach fixed point theorem, we obtain:

Theorem 1.21 (Geometric ergodicity for Wasserstein contractions). *If $\alpha_q(p) < 1$ then there exists a unique invariant probability measure μ of p in $\mathcal{P}^q(S)$. Moreover, for any initial distribution $\nu \in \mathcal{P}^q(S)$, νp^n converges to μ with a geometric rate:*

$$\mathcal{W}^q(\nu p^n, \mu) \leq \alpha_q(p)^n W^q(\nu, \mu).$$

Proof: The Banach fixed point theorem can be applied by Theorem 1.20. □

The assumption $\alpha_q(p) < 1$ 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 $\mu \mapsto \mu p$ is a contraction w.r.t. the $\mathcal{W}^1$ distance based on the modified metric $f \circ d$.

The next lemma is crucial for bounding $\alpha_q(p)$ in applications:

Lemma 1.22 (Bounds for contraction coefficients, Path coupling). *1) Suppose that the transition kernel $p(x, dy)$ is Feller. Then*

$$\alpha(p) = \sup_{x \neq y} \frac{\mathcal{W}^q\left(p(x,\cdot), p(y,\cdot)\right)}{d(x,y)}. \tag{1.5.5}$$

*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 $x_0 = x, x_1, x_2, \dots, x_{n-1}, x_n = y$ from x to y such that $\{x_{i-1}, x_i\} \in E$ for $i = 1, \dots, n$ and $d(x,y) = \sum\limits_{i=1}^{n} d(x_{i-1}, x_i)$. Then*

$$\alpha(p) = \sup_{\{x,y\} \in E} \frac{\mathcal{W}\left(p(x,\cdot), p(y,\cdot)\right)}{d(x,y)}. \tag{1.5.6}$$

The application of the second assertion of the lemma to prove upper bounds for $\alpha_q(p)$ is known as the **path coupling method** of Bubley and Dyer.

Proof: 1) Let $\beta := \sup\limits_{x \neq y} \frac{\mathcal{W}^q(p(x,\cdot),p(y,\cdot))}{d(x,y)}$. We have to show that

$$\mathcal{W}^q(\mu p, \nu p) \leq \beta \mathcal{W}^q(\mu, \nu) \tag{1.5.7}$$

holds for arbitrary probability measures $\mu, \nu \in \mathcal{P}(S)$. By definition of β and since $\mathcal{W}^q(\delta_x, \delta_y) = d(x, y)$, (1.5.7) 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 $\delta_x p$ and $\delta_y p$ such that

$$\left(\int d(x', y')^q \gamma_{xy}(dx'dy')\right)^{\frac{1}{q}} = \mathcal{W}^q(\delta_x p, \delta_y p) \le \beta d(x, y). \tag{1.5.8}$$

Let $\xi(dxdy)$ be an arbitrary coupling of μ and ν. Then a coupling $\gamma(dx'dy')$ of μp and νp is given by

$$\gamma := \int \gamma_{xy}\xi(dxdy),$$

and therefore, by (1.5.8),

$$\begin{aligned}\mathcal{W}^q(\mu p, \nu p) &\le \left(\int d(x', y')^q \gamma(dx'dy')\right)^{\frac{1}{q}} \\ &= \left(\int\int d(x', y')^q \gamma_{xy}(dx'dy')\xi(dxdy)\right)^{\frac{1}{q}} \le \beta \left(\int d(x, y)^q \xi(dxdy)\right)^{\frac{1}{q}}.\end{aligned}$$

By taking the infimum over all couplings $\xi \in \Pi(\mu, \nu)$, we see that μ and ν satisfy (1.5.7).

Finally, to show that (1.5.7) 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 $\mathcal{W}^q$. Hence μ and ν are $\mathcal{W}^q$ limits of corresponding convex combinations μ_n and ν_n $(n \in \mathbb{N})$. By the Feller property, the sequence $\mu_n p$ and $\nu_n p$ converge weakly to $\mu p, \nu p$ respectively. Hence

$$\begin{aligned}\mathcal{W}^q(\mu p, \nu p) &\le \liminf \mathcal{W}^q(\mu_n p, \nu_n p) \\ &\le \beta \liminf \mathcal{W}^q(\mu_n, \nu_n) = \beta \mathcal{W}^q(\mu, \nu).\end{aligned}$$

2) Let $\widetilde{\beta} := \sup\limits_{(x,y)\in E} \frac{\mathcal{W}^q(p(x,\cdot), p(y,\cdot))}{d(x,y)}$. We show that

$$\mathcal{W}^q(p(x, \cdot), p(y, \cdot)) \le \widetilde{\beta} d(x, y)$$

holds for arbitrary $x, y \in S$. Indeed, let $x_0 = x, x_1, x_2, \dots, x_n = y$ be a geodesic from x to y such that $(x_{i-1}, x_i) \in E$ for $i = 1, \dots, n$. Then by the triangle inequality for the $\mathcal{W}^q$ distance,

$$\begin{aligned}\mathcal{W}^q(p(x,\cdot), p(y,\cdot)) &\leq \sum_{i=1}^n \mathcal{W}^q(p(x_{i-1},\cdot), p(x_i,\cdot)) \\ &\leq \widetilde{\beta} \sum_{i=1}^n d(x_{i-1}, x_i) = \widetilde{\beta} d(x,y),\end{aligned}$$

where we have used in the last equality that $x_0, \dots, x_n$ is a geodesic.

□

Exercise. Let $\overline{p}$ be a transition kernel on $S \times S$ such that $\overline{p}((x,y), dx'dy')$ is a coupling of $p(x, dx')$ and $p(y, dy')$ for any $x, y \in S$. Prove that if there exists a distance function $d : S \times S \to [0, \infty)$ and a constant $\alpha \in (0,1)$ such that

$$\overline{p} d \leq \alpha d,$$

then there is a unique invariant probability measure μ of p, and

$$\mathcal{W}_d^1(\nu p^n, \mu) \leq \alpha^n \mathcal{W}_d^1(\nu, \mu) \quad \text{for any } \nu \in \mathcal{P}^1(S).$$

1.5.4 Glauber dynamics, Gibbs sampler

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, colors 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 distribution w.r.t. $\mu(dy)$ of $\eta(x)$ given that $\eta(y) = \xi(y)$ for any $y \neq x$.

To make this precise, we fix a regular version $\mu(d\eta | \eta = \xi \text{ on } V \setminus \{x\})$ of the conditional probability given $(\eta(y))_{y \in V \setminus \{x\}}$, and we define the transition kernel p by

$$p = \frac{1}{|V|} \sum_{x \in V} p_x, \quad \text{where}$$

$p_x(\xi, d\xi') = \mu\left(d\xi' | \xi' = \xi \text{ on } V \setminus \{x\}\right).$

Definition. *A time-homogeneous Markov chain with transition kernel p is called **Glauber dynamics** or **random scan Gibbs sampler** with stationary distribution μ.*

That μ is indeed invariant w.r.t. p is shown in the next lemma:

Lemma 1.23. *The transition kernels p_x $(x \in V)$ and p satisfy the detailed balance conditions*

$$\mu(d\xi)p_x(\xi, d\xi') = \mu(d\xi')p_x(\xi', d\xi),$$
$$\mu(d\xi)p(\xi, d\xi') = \mu(d\xi')p(\xi', d\xi).$$

In particular, μ is a stationary distribution for p.

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

$$\begin{aligned}
\mu\left(d\xi\right) p_x\left(\xi, d\xi'\right) &= \hat{\mu}_x\left(d\hat{\xi}(x)\right) \mu_x\left(d\xi(x)|\hat{\xi}(x)\right) \delta_{\hat{\xi}(x)}\left(d\hat{\xi}'(x)\right) \mu_x\left(d\xi'(x)|\hat{\xi}(x)\right) \\
&= \hat{\mu}_x\left(d\hat{\xi}'(x)\right) \mu_x\left(d\xi(x)|\hat{\xi}'(x)\right) \delta_{\hat{\xi}'(x)}\left(d\hat{\xi}(x)\right) \mu_x\left(d\xi'(x)|\hat{\xi}'(x)\right) \\
&= \mu\left(d\xi'\right) p_x\left(\xi', d\xi\right).
\end{aligned}$$

Hence the detailed balance condition is satisfied w.r.t. p_x for any $x \in V$, and, by averaging over x, also w.r.t. p. □

Examples. 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:

$$\mu = \text{Unif}\left(\{\eta \in T^V : \eta(x) \neq \eta(y)\ \forall (x, y) \in E\}\right).$$

The Gibbs sampler selects in each step 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

$$\mu_\lambda(\eta) = \frac{1}{Z_\lambda}\lambda^{\sum\limits_{x \in V}\eta(v)} \qquad \text{if } \eta(x)\eta(y) = 0 \text{ for any } (x, y) \in E,$$

and $\mu_\lambda(\eta) = 0$ otherwise, where Z_λ is a finite normalization constant. The Gibbs sampler updates in each step $\xi(x)$ for a randomly chosen vertex x according to

$$\xi'(x) = 0 \text{ if } \xi(y) = 1 \text{ for some } y \sim x,$$
$$\xi'(x) \sim \text{Bernoulli}\left(\frac{\lambda}{1+\lambda}\right) \text{ otherwise}.$$

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

$$\mu_\beta(\eta) = \tfrac{1}{Z_\beta} e^{-\beta H(\eta)} \qquad \text{for any } \eta \in \{-1, +1\}^V,$$

where Z_β is again a normalizing constant, and the Ising Hamiltonian H is given by

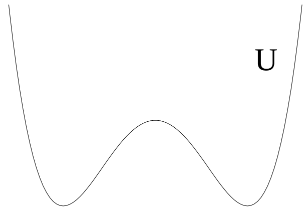
$$H(\eta) = \frac{1}{2} \sum_{\{x,y\} \in E} |\eta(x) - \eta(y)|^2 = - \sum_{\{x,y\} \in E} \eta(x)\eta(y) + |E|.$$

Thus μ_β favours configurations where neighbouring spins coincide, and this preference gets stronger as the temperature $\frac{1}{\beta}$ decreases. The heat bath dynamics 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 **meanfield 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 "meanfield" $\frac{1}{n} \sum_{y \in V} \eta(y)$.

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 massles **Gaussian free field over V**. If U is a double-well potential then μ_β is a continuous version of the Ising model.



5) **Bayesian posterior distributions.** Gibbs samplers are applied frequently to sample from posterior distributions in Bayesian statistical models. For instance in a typical hierarchical Bayes model one assumes that the data are realizations of conditionally independent random variables Y_{ij} $(i = 1, \ldots, k, j = 1, \ldots, m_i)$ with conditional laws

$$Y_{ij}|(\theta_1, \ldots, \theta_k, \lambda_e) \sim \mathcal{N}(\theta_i, \lambda_e^{-1}).$$

The parameters $\theta_1, \ldots, \theta_k$ and λ_e are again assumed to be conditionally independent random variables with

$$\theta_i|(\mu, \lambda_\theta) \sim \mathcal{N}(\mu, \lambda_\theta^{-1}) \text{ and } \lambda_e|(\mu, \lambda_\theta) \sim \Gamma(a_2, b_2).$$

Finally, μ and λ_θ are independent with

$$\mu \sim \mathcal{N}(m, v) \text{ and } \lambda_\theta \sim \Gamma(a_1, b_1)$$

where $a_1, b_1, a_2, b_2 \in \mathbb{R}_+$ and $v \in \mathbb{R}$ are given constants, cf. [Jones] [13]. The posterior distribution μ of $(\theta_1, \ldots, \theta_k, \mu, \lambda_e, \lambda_\theta)$ on $\mathbb{R}^{k+3}$ given observations $Y_{ij} = y_{ij}$ is then given by Bayes' formula. Although the density is explicitly 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 distributions w.r.t. μ of one of the parameters $\theta_1, \ldots, \theta_k, \mu, \lambda_e, \lambda_\theta$ 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 w.r.t. μ on a computer. If this 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 longtime 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, cf. [Roberts&Rosenthal:Markov chains and MCMC algorithms] [30].

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.,

$$d(\xi, \eta) = \sum_{x \in V} \varrho\left(\xi(x), \eta(x)\right), \quad \xi, \eta \in T^V.$$

A frequent choice is $\varrho(s, t) = 1_{s \neq t}$. In this case,

$$d(\xi, \eta) = |\{x \in V : \xi(x) \neq \eta(x)\}|$$

is called the **Hamming distance** of ξ and η.

Lemma 1.24. *Let $n = |V|$. Then for the Gibbs sampler,*

$$\mathcal{W}_d^1\left(p(\xi,\cdot),p(\eta,\cdot)\right) \le \left(1-\frac{1}{n}\right)d(\xi,\eta) + \frac{1}{n}\sum_{x\in V}\mathcal{W}_\varrho^1\left(\mu_x(\cdot|\xi),\mu_x(\cdot|\eta)\right)$$

for any $\xi,\eta \in T^V$.

Proof: Let γ_x for $x \in V$ be optimal couplings w.r.t. $\mathcal{W}_\varrho^1$ of the conditional measures $\mu_x(\cdot|\xi)$ and $\mu_x(\cdot|\eta)$. Then we can construct a coupling of $p(\xi,d\xi')$ and $p(\eta,d\eta')$ in the following way:

- Draw $U \sim$ Unif(V).
- Given U, choose $(\xi'(U),\eta'(U)) \sim \gamma_U$, and set $\xi'(x) = \xi(x)$ and $\eta'(x) = \eta(x)$ for any $x \neq U$.

For this coupling we obtain:

$$\begin{aligned} E[d(\xi',\eta')] &= \sum_{x\in V} E[\varrho(\xi'(x),\eta'(x))] \\ &= d(\xi,\eta) + E\left[\varrho(\xi'(U),\eta'(U)) - \varrho(\xi(U),\eta(U))\right] \\ &= d(\xi,\eta) + \frac{1}{n}\sum_{x\in V}\left(\int \varrho(s,t)\gamma_x(dsdt) - \varrho(\xi(x),\eta(x))\right) \\ &= \left(1-\frac{1}{n}\right)d(\xi,\eta) + \frac{1}{n}\sum_{x\in V}\mathcal{W}_\varrho^1\left(\mu_x(\cdot|\xi),\mu_x(\cdot|\eta)\right). \end{aligned}$$

Here we have used in the last step the optimality of the coupling γ_x. The claim follows since $\mathcal{W}_d^1(p(x,\cdot),p(y,\cdot)) \le E[d(\xi',\eta')]$. □

The lemma shows that we obtain contractivity w.r.t. $\mathcal{W}_d^1$ if the conditional distributions at $x \in V$ do not depend too strongly on the values of the configuration at other vertices:

Theorem 1.25 (Geometric ergodicity of the Gibbs sampler for weak interactions).

1) *Suppose that there exists a constant $c \in (0,1)$ such that*

$$\sum_{x\in V}\mathcal{W}_\varrho^1\left(\mu_x(\cdot|\xi),\mu_x(\cdot|\eta)\right) \le cd(\xi,\eta) \quad \textit{for any } \xi,\eta\in T^V. \tag{1.5.9}$$

Then

$$\mathcal{W}_d^1(\nu p^t,\mu) \le \alpha(p)^t\mathcal{W}_d^1(\nu,\mu) \quad \textit{for any } \nu\in\mathcal{P}(T^V) \textit{ and } t\in\mathbb{Z}_+, \tag{1.5.10}$$

where $\alpha(p) \le \exp\left(-\frac{1-c}{n}\right)$.

2) *If T is a graph and ϱ is geodesic then it suffices to verify* (1.5.9) *for neighbouring configurations $\xi, \eta \in T^V$ such that $\xi = \eta$ on $V \setminus \{x\}$ for some $x \in V$ and $\xi(x) \sim \eta(x)$.*

Proof: 1) If (1.5.9) holds then by Lemma 1.24,

$$\mathcal{W}_d\left(p(\xi,\cdot),p(\eta,\cdot)\right) \leq \left(1-\frac{1-c}{n}\right) d(\xi,\eta) \quad \text{for any } \xi,\eta \in T^V.$$

Hence (1.5.10) holds with $\alpha(p) = 1 - \frac{1-c}{n} \leq \exp\left(-\frac{1-c}{n}\right)$.

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 from the path coupling lemma 1.22.

□

The results in Theorem 1.25 can be applied to many basic models including random colourings, hardcore models and meanfield 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

$$\mu_x(\cdot|\xi) = \mu_x(\cdot|\eta) \quad \text{for } x = y \text{ and for any } x \not\sim y.$$

Moreover, for $x \sim y$ and $\varrho(s,t) = 1_{s \neq t}$ we have

$$\mathcal{W}_\varrho^1\left(\mu_x(\cdot|\xi),\mu_x(\cdot|\eta)\right) \leq \frac{1}{|T|-\Delta}$$

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

$$\sum_{x \in V} \mathcal{W}_\varrho^1\left(\mu_x(\cdot|\xi),\mu_x(\cdot|\eta)\right) \leq \frac{\Delta}{|T|-\Delta} d(\xi,\eta),$$

and therefore, (1.5.10) holds with

$$\alpha(p) \leq \exp\left(-\left(1-\frac{\Delta}{|T|-\Delta}\right)\cdot\frac{1}{n}\right), \quad \text{and hence}$$
$$\alpha(p)^t \leq \exp\left(-\frac{|T|-2\Delta}{|T|-\Delta}\cdot\frac{t}{n}\right).$$

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:

1.6 Geometric and subgeometric convergence to equilibrium

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. Moreover, bounds on the rate of convergence to equilibrium are derived by coupling methods. We assume again that S is a polish space with Borel σ-algebra $\mathcal{B}$.

1.6.1 Total variation norm

The variation $|\eta|(B)$ of an additive set-function $\eta : \mathcal{B} \to \mathbb{R}$ on a set $B \in \mathcal{B}$ is defined by

$$|\eta|(B) := \sup\left\{\sum_{i=1}^n |\eta(A_i)| : n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{B} \text{ disjoint with } \bigcup_{i=1}^n A_i \subset B\right\}.$$

The **total variation norm** of η is

$$\|\eta\|_{\mathrm{TV}} = \frac{1}{2}|\eta|(S).$$

Note that this definition 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 1.26 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

$$S = S^+ \dot\cup S^- \text{ with } S^+ = \{\varrho \geq 0\}, S^- = \{\varrho < 0\},$$
$$\eta = \eta^+ - \eta^- \text{ with } d\eta^+ = \varrho^+ d\lambda, d\eta^- = \varrho^- d\lambda.$$

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, the total variation norm of η is the L^1 norm of ϱ:

$$\|\eta\|_{\mathrm{TV}} = \int |\varrho| dx = \|\varrho\|_{L^1(\lambda)}. \tag{1.6.1}$$

Lemma 1.26 (Equivalent descriptions of the total variation norm). *Let $\mu, \nu \in \mathcal{P}(S)$ and $\lambda \in M_+(S)$ such that μ and ν are both absolutely continuous w.r.t. λ. Then the following identities hold:*

$$\begin{aligned}\|\mu - \nu\|_{TV} &= (\mu-\nu)^+(S) = (\mu-\nu)^-(S) = 1 - (\mu \wedge \nu)(S) \\ &= \left\| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right\|_{L^1(\lambda)} \\ &= \frac{1}{2} \sup \{ |\mu(f) - \nu(f)| : f \in \mathcal{F}_b(S) \text{ s.t. } \|f\|_{\sup} \leq 1 \} && (1.6.2) \\ &= \inf \{ P[X \neq Y] : X \sim \mu, Y \sim \nu \} && (1.6.3)\end{aligned}$$

In particular, $\|\mu - \nu\|_{TV} \in [0,1]$.

Remarks. 1) The last identity shows that the total variation distance of μ and ν is the Kantorovich distance $\mathcal{W}^1_d(\mu,\nu)$ based on the trivial metric $d(x,y) = 1_{x \neq y}$ on S.

2) The assumption $\mu, \nu << \lambda$ 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,

$$(\mu - \nu)(S) = \mu(S) - \nu(S) = 0.$$

Hence $(\mu-\nu)^+(S) = (\mu-\nu)^-(S)$, 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 $\|\mu - \nu\|_{\mathrm{TV}} = \left\| \frac{d\mu}{d\lambda} - \frac{d\nu}{d\lambda} \right\|_{L^1(\lambda)}$ holds by (1.6.1). Moreover, for $f \in \mathcal{F}_b(S)$ with $\|f\|_{\sup} \leq 1$,

$$\begin{aligned}|\mu(f) - \nu(f)| &\leq |(\mu-\nu)^+(f)| + |(\mu-\nu)^-(f)| \\ &\leq (\mu-\nu)^+(S) + (\mu-\nu)^-(S) = 2\|\mu - \nu\|_{\mathrm{TV}}\end{aligned}$$

with identity for $f = 1_{S^+} - 1_{S^-}$. This proves the representation (1.6.2) of $\|\mu - \nu\|_{\mathrm{TV}}$.
Finally, to prove (1.6.3) 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 f with $\|f\|_{\sup} \leq 1$. Hence by (1.6.2),

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \inf_{\substack{X \sim \mu \\ Y \sim \nu}} P[X \neq Y].$$

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

$$\mu = (\mu \wedge \nu) + (\mu - \nu)^+ = p\alpha + (1-p)\beta, \tag{1.6.4}$$

$$\nu = (\mu \wedge \nu) + (\mu - \nu)^- = p\alpha + (1-p)\gamma \tag{1.6.5}$$

with $p = (\mu \wedge \nu)(S)$ and probability measures $\alpha, \beta, \gamma \in \mathcal{P}(S)$. We choose independent random variables $U \sim \alpha, V \sim \beta, W \sim \gamma$ and $Z \sim$ 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 (1.6.4) and (1.6.5), $(X, Y) \in \Pi(\mu, \nu)$ and

$$P[X \neq Y] \leq P[Z = 0] = 1 - p = 1 - (\mu \wedge \nu)(S) = \|\mu - \nu\|_{\mathrm{TV}}.$$

□

Remark. The last equation can also be seen as a special case of the Kantorovich-Rubinstein duality formula.

1.6.2 Geometric ergodicity

Let p be a transition kernel on $(S, \mathcal{B})$. We define the **local contraction coefficient** $\alpha(p, K)$ of p on a set $K \subset S$ w.r.t. the total variation distance by

$$\alpha(p, K) = \sup_{x, y \in K} \|p(x, \cdot) - p(y, \cdot)\|_{\mathrm{TV}} = \sup_{\substack{x, y \in K \\ x \neq y}} \frac{\|\delta_x p - \delta_y p\|_{\mathrm{TV}}}{\|\delta_x - \delta_y\|_{\mathrm{TV}}}. \tag{1.6.6}$$

Note that in contrast to more general Wasserstein contraction coefficients, we always have

$$\alpha(p, K) \leq 1.$$

Moreover, $\alpha(p, K) \leq 1 - \varepsilon$ holds for $\varepsilon > 0$ if p satisfies the following **local minorization condition**: There exists a probability measure ν on S such that

$$p(x, B) \geq \varepsilon \nu(B) \quad \text{for any } x \in K \text{ and } B \in \mathcal{B}. \tag{1.6.7}$$

Doeblin's classical theorem states that if $\alpha(p^n, S) < 1$ for some $n \in \mathbb{N}$ then there exists a unique stationary distribution μ of p, and uniform ergodicity holds in the following sense:

$$\sup_{x \in S} \|p^t(x, \cdot) - \mu\|_{\text{TV}} \to 0 \quad \text{as } t \to \infty. \tag{1.6.8}$$

Exercise (Doeblin's Theorem). Prove that (1.6.8) holds if $\alpha(p^n, S) < 1$ for some $n \in \mathbb{N}$.

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:

Example (Autoregressive process AR$(\mathbf{1})$). Suppose that

$$X_{n+1} = \alpha X_n + W_{n+1},\ X_0 = x$$

with $\alpha \in (-1, 1), x \in \mathbb{R}$, and i.i.d. random variables $W_n : \Omega \to \mathbb{R}$. 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

$$p^n(x, \cdot) = N\left(\alpha^n x, \frac{1-\alpha^{2n}}{1-\alpha^2}\right),\ x \in S.$$

As $n \to \infty$, $p^n(x, \cdot) \to \mu$ in total variation, where

$$\mu = N\left(0, \frac{1}{1-\alpha^2}\right)$$

is the unique stationary distribution. However, the convergence is not uniform in x, since

$$\sup_{x \in \mathbb{R}} \|p^n(x, \cdot) - \mu\|_{\text{TV}} = 1 \quad \text{for any } n \in \mathbb{N}.$$

The example 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 (Geometric ergodicity). *A time-homogeneous Markov chain (X_n, P_x) with transition kernel p is called **geometrically ergodic with stationary distribution** μ iff there exist $\gamma \in (0, 1)$ and a non-negative function $M : S \to \mathbb{R}$ such that*

$$\|p^n(x, \cdot) - \mu\|_{TV} \leq M(x)\gamma^n \quad \textit{for } \mu\textit{-almost every } x \in S.$$

Harris' Theorem states that geometric ergodicity is a consequence of a **local minorization condition** and 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

$$\mathcal{L}V(x) \leq C - \lambda V(x) \quad \text{for any } x \in S. \tag{1.6.9}$$

In terms of the transition kernel the condition (LG) states that

$$pV(x) \leq C + \gamma V(x) \tag{1.6.10}$$

where $\gamma = 1 - \lambda < 1$.

Below, we follow the approach of M. Hairer and J. Mattingly to give a simple proof of a quantitative version of the Harris Theorem, cf. [Hairer:Convergence of Markov processes,Webpage M.Hairer] [12]. 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 $\|\mu - \nu\|_{\mathrm{TV}} \leq \mathcal{W}_\beta(\mu,\nu)$ with equality for $\beta = 0$.

Theorem 1.27 (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*

$$\alpha(p, \{V \leq r\}) < 1 \quad \textit{for some } r > 2C/\lambda. \tag{1.6.11}$$

Then there exists a constant $\beta \in \mathbb{R}_+$ such that $\alpha_\beta(p) < 1$. In particular, there is a unique stationary distribution μ of p satisfying $\int V\,d\mu < \infty$, and geometric ergodicity holds:

$$\|p^n(x,\cdot) - \mu\|_{TV} \leq \mathcal{W}_\beta\left(p^n(x,\cdot),\mu\right) \leq \left(1 + \beta V(x) + \beta\int V\,d\mu\right)\alpha_\beta(p)^n$$

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

Remark. There are explicit expressions for the constants β and $\alpha_\beta(p)$.

Proof: Fix $x, y \in S$ with $x \neq y$, and let (X,Y) be a **maximal** coupling of $p(x,\cdot)$ and $p(y,\cdot)$ w.r.t. the total variation distance, i.e.,

$$P[X \neq Y] = \|p(x,\cdot) - p(y,\cdot)\|_{\mathrm{TV}}.$$

Then for $\beta \geq 0$,

$$\begin{aligned}
\mathcal{W}_\beta(p(x,\cdot),p(y,\cdot)) &\leq E[d_\beta(X,Y)] \\
&\leq P[X \neq Y] + \beta E[V(X)] + \beta E[V(Y)] \\
&= \|p(x,\cdot)-p(y,\cdot)\|_{\mathrm{TV}} + \beta(pV)(x) + \beta(pV)(y) \\
&\leq \|p(x,\cdot)-p(y,\cdot)\|_{\mathrm{TV}} + 2C\beta + (1-\lambda)\beta(V(x)+V(y)), \qquad (1.6.12)
\end{aligned}$$

where we have used (1.6.10) in the last step. We now fix r as in (1.6.11), and distinguish cases:

(i) If $V(x)+V(y) \geq r$, then the Lyapunov condition ensures contractivity. Indeed, by (1.6.12),

$$\mathcal{W}_\beta(p(x,\cdot),p(y,\cdot)) \leq d_\beta(x,y) + 2C\beta - \lambda\beta\cdot(V(x)+V(y)). \qquad (1.6.13)$$

Since $d_\beta(x,y) = 1+\beta V(x)+\beta V(y)$, the expression on the right hand side in (1.6.13) is bounded from above by $(1-\delta)d_\beta(x,y)$ for some constant $\delta > 0$ provided $2C\beta + \delta \leq (\lambda-\delta)\beta r$. This condition is satisfied if we choose

$$\delta := \frac{\lambda - \frac{2C}{r}}{1+\frac{1}{\beta r}} = \frac{\lambda r - 2C}{1+\beta r}\beta,$$

which is positive since $r > 2C/\lambda$.

(ii) If $V(x)+V(y) < r$ then contractivity follows from (1.6.11). Indeed, (1.6.12) implies that for $\varepsilon := \min\left(\frac{1-\alpha(p,\{V\leq r\})}{2},\lambda\right)$,

$$\begin{aligned}
\mathcal{W}_\beta(p(x,\cdot),p(y,\cdot)) &\leq \alpha(p,\{V\leq r\}) + 2C\beta + (1-\lambda)\beta(V(x)+V(y)) \\
&\leq (1-\varepsilon)d_\beta(x,y)
\end{aligned}$$

provided $\beta \leq \frac{1-\alpha(p,\{V\leq r\})}{4C}$.

Choosing $\delta,\varepsilon,\beta > 0$ as in (i) and (ii), we obtain

$$\mathcal{W}_\beta(p(x,\cdot),p(y,\cdot)) \leq (1-\min(\delta,\varepsilon))d_\beta(x,y) \quad \text{for any } x,y\in S,$$

i.e., the **global** contraction coefficient $\alpha_\beta(p)$ w.r.t. $\mathcal{W}_\beta$ is strictly smaller than one. Hence there exists a unique stationary distribution

$$\mu \in \mathcal{P}^1_\beta(S) = \left\{\mu\in\mathcal{P}(S) : \int V\,d\mu < \infty\right\}, \quad \text{and}$$

$$\begin{aligned}
\mathcal{W}_\beta(p^n(x,\cdot),\mu) &= \mathcal{W}_\beta\left(\delta_x p^n, \mu p^n\right) \leq \alpha_\beta(p)^n \mathcal{W}_\beta(\delta_x,\mu) \\
&= \alpha_\beta(p)^n\left(1+\beta V(x) + \beta\int V\,d\mu\right).
\end{aligned}$$

□

Remark (**Doeblin's Theorem**). If $\alpha(p, S) < 1$ then by choosing $V \equiv 0$, we recover Doeblin's Theorem:

$$\|p^n(x, \cdot) - \mu\|_{\text{TV}} \leq \alpha(p, S)^n \to 0 \quad \text{uniformly in } x \in S.$$

Example (**State space model in $\mathbb{R}^{\text{d}}$**). Consider the Markov chain with state space $\mathbb{R}^d$ and transition step

$$x \mapsto x + b(x) + \sigma(x)W,$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are measurable functions, and $W : \Omega \to \mathbb{R}^d$ is a random vector with $E[W] = 0$ and $\text{Cov}(W_i, W_j) = \delta_{ij}$. Choosing $V(x) = |x|^2$, we obtain

$$\mathcal{L}V(x) = 2x \cdot b(x) + |b(x)|^2 + \text{tr}(\sigma^T\sigma)(x) \leq C - \lambda V(x)$$

for some $C, \lambda \in (0, \infty)$ provided

$$\limsup_{|x| \to \infty} \frac{x \cdot b(x) + |b(x)|^2 + \text{tr}(\sigma^T\sigma)(x)}{|x|^2} < 0.$$

Since

$$\alpha(p, \{V \leq r\}) = \sup_{|x| \leq \sqrt{r}} \sup_{|y| \leq \sqrt{r}} \left\| N\left(x + b(x), (\sigma\sigma^T)(x)\right) - N\left(y + b(y), (\sigma\sigma^T)(y)\right) \right\|_{\text{TV}} < 1$$

for any $r \in (0, \infty)$, 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 1.27 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, cf. e.g. [Roberts&Rosenthal:Markov chains & MCMC algorithms] [30].

1.6.3 Couplings of Markov chains 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. There are two possible approaches for proving convergence to equilibrium at subgeometric rates:

a) The Harris' Theorem can be extended to the subgeometric case provided a Lyapunov condition of the form

$$\mathcal{L}V \leq C - \varphi \circ V$$

holds with a concave increasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\varphi(0) = 0$, cf. [Hairer] [12] and [Meyn&Tweedie] [23].

b) Alternatively, couplings of Markov chains 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.

Definition (Couplings of stochastic processes).

1) *A **coupling** of two stochastic processes (X_n, P) and (Y_n, Q) with state space S and T is given by a process $\left((\widetilde{X}, \widetilde{Y}), \widetilde{P}\right)$ with state space $S \times T$ such that*

$$(\widetilde{X}_n)_{n\geq 0} \sim (X_n)_{n\geq 0} \quad \textit{and} \quad (\widetilde{Y}_n)_{n\geq 0} \sim (Y_n)_{n\geq 0}.$$

2) *The coupling is called **Markovian** iff the process $(\widetilde{X}_n, \widetilde{Y}_n)_{n\geq 0}$ is a Markov chain on the product space $S \times T$.*

Example (Construction of Markovian couplings). A Markovian coupling of two time homogeneous Markov chains can be constructed from a coupling of the transition functions. Suppose that p and q are transition kernels on measurable spaces $(S, \mathcal{B})$ and $(T, \mathcal{C})$, and $\overline{p}$ is a transition kernel on $(S \times T, \mathcal{B} \otimes \mathcal{C})$ such that $\overline{p}\left((x, y), dx'dy'\right)$ is a coupling of the measures $p(x, dx')$ and $p(y, dy')$ for any $x \in S$ and $y \in T$. Then for any $x \in S$ and $y \in T$, the canonical Markov chain $((X_n, Y_n), P_{xy})$ with 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, $((X_n, Y_n), P_\gamma)$ is a coupling of Markov chains with transition kernels p, q and initial distributions μ, ν provided γ is a coupling of μ and ν.

Theorem 1.28 (Coupling lemma). *Suppose that $((X_n, Y_n)_{n\geq 0}, P)$ is a Markovian coupling of Markov chains with transition kernel p and initial distributions μ and ν. Then*

$$\|\mu p^n - \nu p^n\|_{TV} \leq \|\mathit{Law}(X_{n:\infty}) - \mathit{Law}(Y_{n:\infty})\|_{TV} \leq P[T > n],$$

*where T is the **coupling time** defined by*

$$T = \min\{n \geq 0 : X_n = Y_n\}.$$

In particular, if $T < \infty$ almost surely then

$$\lim_{n\to\infty} \|\mathrm{Law}(X_{n:\infty}) - \mathrm{Law}(Y_{n:\infty})\|_{\mathrm{TV}} = 0.$$

Proof: 1) We first show that we may assume without loss of generality that $X_n = Y_n$ for any $n \geq T$. Indeed, if this is not the case then we can define a modified coupling $(X_n, \widetilde{Y}_n)$ with **the same coupling time** by setting

$$\widetilde{Y}_n := \begin{cases} Y_n & \text{for } n < T, \\ X_n & \text{for } n \geq T. \end{cases}$$

The fact that $(X_n, \widetilde{Y}_n)$ is again a coupling of the same Markov chains follows from the strong Markov property: T is a stopping time w.r.t. the filtration $(\mathcal{F}_n)$ generated by the process (X_n, Y_n), and hence on $\{T < \infty\}$ and under the conditional law given $\mathcal{F}_T$, $X_{T:\infty}$ is a Markov chain with transition kernel p and initial value Y_T. Therefore, the conditional law of

$$\widetilde{Y}_{0:\infty} = (Y_1, \dots, Y_{T-1}, X_T, X_{T+1}, \dots)$$

given $\mathcal{F}_T$ coincides with the conditional law of

$$Y_{0:\infty} = (Y_1, \dots, Y_{T-1}, Y_T, Y_{T+1}, \dots)$$

given $\mathcal{F}_T$, and hence the unconditioned law of $(\widetilde{Y}_n)$ and (Y_n) coincides as well.

2) Now suppose that $X_n = Y_n$ for $n \geq T$. Then also $X_{n:\infty} = Y_{n:\infty}$ for $n \geq T$, and thus we obtain

$$\|\text{Law}\,(X_{n:\infty}) - \text{Law}\,(Y_{n:\infty})\|_{\text{TV}} \leq P\,[X_{n:\infty} \neq Y_{n:\infty}] \leq P[T > n].$$

□

If μ is a stationary distribution for p then $\mu p^n = \mu$ and Law$(X_{n:\infty}) = P_\mu$ for any $n \geq 0$. Hence the coupling lemma provides upper bounds for the total variation distance to stationarity. As an immediate consequence we note:

Corollary 1.29 (Convergence rates by coupling). *Let T be the coupling time for a Markovian coupling of time-homogeneous Markov chains with transition kernel p and initial distributions μ and ν. Suppose that*

$$E[\psi(T)] < \infty$$

for some increasing function $\psi : \mathbb{Z}_+ \to \mathbb{R}_+$ with $\lim_{n\to\infty} \psi(n) = \infty$. Then

$$\|\mu p^n - \nu p^n\|_{TV} = O\left(\frac{1}{\psi(n)}\right), \qquad \textit{and even} \tag{1.6.14}$$

$$\sum_{n=0}^{\infty} (\psi(n+1) - \psi(n))\,\|\mu p^n - \nu p^n\|_{TV} < \infty. \tag{1.6.15}$$

Proof: By the coupling lemma and Markov's inequality,

$$\|\mu p^n - \nu p^n\|_{\mathrm{TV}} \le P[T > n] \le \frac{1}{\psi(n)} E[\psi(T)] \quad \text{for any } n \in \mathbb{N}.$$

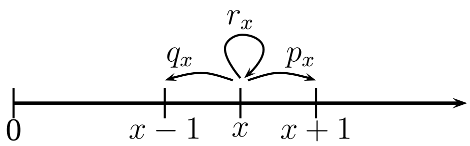
Furthermore, by Fubini's Theorem,

$$\begin{aligned}\sum_{n=0}^{\infty} (\psi(n+1) - \psi(n)) \, \|\mu p^n - \nu p^n\|_{\mathrm{TV}} &\le \sum_{n=0}^{\infty} (\psi(n+1) - \psi(n)) \, P[T > n] \\ &= \sum_{i=1}^{\infty} P[T = i] \, (\psi(n) - \psi(0)) \le E[\psi(T)].\end{aligned}$$

□

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 $p(x, x+1) = p_x, p(x, x-1) = q_x$ and $p(x, x) = r_x$. 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 any x (i.e., the Markov chain is "**lazy**"). For $f \in \mathcal{F}_b(\mathbb{Z}_+)$, the generator is given by

$$(\mathcal{L}f)(x) = p_x \, (f(x+1) - f(x)) + q_x \, (f(x-1) - f(x)) \qquad \forall x \in \mathbb{Z}_+.$$

By solving the system of equations $\mu\mathcal{L} = \mu - \mu p = 0$ explicitly, one shows that there is a two-parameter family of invariant measures given by

$$\mu(x) = a + b \cdot \frac{p_0 p_1 \cdots p_{x-1}}{q_1 q_2 \cdots q_x} \quad (a, b \in \mathbb{R}).$$

In particular, a stationary distribution 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

$$p_x \le \left(1 - \frac{1+\varepsilon}{x}\right) q_{x+1} \quad \text{for large } x.$$

Now suppose that a stationary distribution $\overline{\mu}$ exists. To obtain an upper bound on the rate of convergence to $\overline{\mu}$, we consider the straightforward Markovian coupling $((X_n, Y_n), P_{xy})$ of two chains with transition kernel p determined by the transition step

$$(x,y) \to \begin{cases} (x+1,y) & \text{with probability } p_x, \\ (x-1,y) & \text{with probability } q_x, \\ (x,y+1) & \text{with probability } p_x, \\ (x,y-1) & \text{with probability } q_x. \end{cases}$$

Since at each transition step only one of the chains (X_n) and (Y_n) is moving one unite, 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 a stationary distribution exists and the chain is irreducible, all states are positive recurrent. Hence

$$E[T] \leq E[T_0^X] < \infty.$$

Therefore by Corollary 1.29, the total variation distance from equilibrium is **always** decaying at least of order $O(n^{-1})$:

$$\|p^n(x,\cdot) - \overline{\mu}\|_{\mathrm{TV}} = O(n^{-1}), \quad \sum_{n=1}^{\infty} \|p^n(x,\cdot) - \overline{\mu}\|_{\mathrm{TV}} < \infty.$$

To prove a stronger decay, one can construct appropriate Lyapunov functions for bounding higher moments of T. For instance suppose that

$$p_x - q_x \sim -a x^{\gamma} \quad \text{as} \quad x \to \infty$$

for some $a > 0$ and $\gamma \in (-1, 0]$.

(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 1.6 or Theorem 1.9 that

$$E[T^k] \leq E[(T_0^X)^k] < \infty \quad \text{for any } k < \frac{n}{1-\gamma}.$$

Since n can be chosen arbitrarily large, we see that the convergence rate is faster than any polynomial rate:

$$\|p^n(x,\cdot) - \overline{\mu}\|_{\mathrm{TV}} = O(n^{-k}) \quad \text{for any } k \in \mathbb{N}.$$

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 1.29 with $\psi(n) = e^{cn}$.

1.7 Mixing times

Let p be a transition kernel on $(S, \mathcal{B})$ with stationary distribution μ. For $K \in \mathcal{B}$ and $t \geq 0$ let

$$d_{\mathrm{TV}}(t,K) = \sup_{x \in K} \|p^t(x,\cdot) - \mu\|_{\mathrm{TV}}$$

denote the maximal total variation distance from equilibrium after t steps of the Markov chain with transition kernel p and initial distribution concentrated on K.

Definition (Mixing time). *For $\varepsilon > 0$, the ε-**mixing time** of the chain with initial value in K is defined by*

$$t_{mix}(\varepsilon, K) = \min\{t \geq 0 : d(t,K) \leq \varepsilon\}.$$

Moreover, we denote by $t_{mix}(\varepsilon)$ the global mixing time $t_{mix}(\varepsilon, S)$.

Exercise (Decay of TV-distance to equilibrium).
Prove that for any initial distribution $\nu \in \mathcal{P}(S)$, the total variation distance $\|\nu p^t - \mu\|_{\mathrm{TV}}$ is a decreasing function in t. Hence conclude that

$$d_{\mathrm{TV}}(t,K) \leq \varepsilon \quad \text{for any } t \geq t_{\mathrm{mix}}(\varepsilon, K).$$

An important problem is the dependence of mixing times on parameters such as the dimension of the underlying state space. In particular, the distinction between "**slow**" and "**rapid**" **mixing**, i.e., exponential vs. polynomial increase of the mixing time as a parameter goes to infinity, is often related to phase transitions.

1.7.1 Upper bounds in terms of contraction coefficients

To quantify mixing times note that by the triangle inequality for the TV-distances,

$$d_{\mathrm{TV}}(t,S) \le \alpha(p^t) \le 2 d_{\mathrm{TV}}(t,S),$$

where α denotes the global TV-contraction coefficient.

Example (Random colourings). For the random colouring chain with state space T^V, we have shown in the example below Theorem 1.25 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

$$\alpha_d(p^t) \le \alpha_d(p)^t \le \exp\left(-\frac{|T|-2\Delta}{|T|-\Delta}\cdot\frac{t}{n}\right). \tag{1.7.1}$$

Here Δ denotes the degree of the regular graph V and $n = |V|$. Since

$$1_{\xi\neq\eta} \le d(\xi,\eta) \le n \cdot 1_{\xi\neq\eta} \quad \text{for any } \xi,\eta \in T^V,$$

we also have

$$\|\nu-\mu\|_{\mathrm{TV}} \le \mathcal{W}_d^1(\nu,\mu) \le n\|\nu-\mu\|_{\mathrm{TV}} \quad \text{for any } \nu \in \mathcal{P}(S).$$

Therefore, by (1.7.1), we obtain

$$\|p^t(\xi,\cdot)-\mu\|_{\mathrm{TV}} \le n\alpha_d(p^t) \le n\exp\left(-\frac{|T|-2\Delta}{|T|-\Delta}\cdot\frac{t}{n}\right)$$

for any $\xi \in T^V$ and $t \ge 0$. The right-hand side is smaller than ε for $t \ge \frac{|T|-\Delta}{|T|-2\Delta} n\log(n/\varepsilon)$. Thus we have shown that

$$t_{\mathrm{mix}}(\varepsilon) = O\left(n\log n + n\log\varepsilon^{-1}\right) \quad \text{for } |T| > 2\Delta.$$

This is a typical example of **rapid mixing** with a total variation **cast-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 $O(n)$.

Example (Harris Theorem). In the situation of Theorem 1.27, the global distance $d_{\mathrm{TV}}(t,S)$ to equilibrium does not go to 0 in general. However, on the level sets of the Lyapunov function V,

$$d_{\mathrm{TV}}(t,\{V\le r\}) \le \left(1+\beta r+\beta\int V\,d\mu\right)\alpha_\beta(p)^t$$

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

$$t_{\mathrm{mix}}(\varepsilon,\{V\le r\}) \le \frac{\log\left(1+\beta r+\beta\int V\,d\mu\right)+\log(\varepsilon^{-1})}{\log(\alpha_\beta(p)^{-1})}.$$

1.7.2 Upper bounds by Coupling

We can also apply the coupling lemma to derive upper bounds for mixing times in the following way:

Corollary 1.30 (Coupling times and mixing times). *Suppose that* $((X_n, Y_n), P_{x,y})$ *is a Markovian coupling of the Markov chains with initial value* $x, y \in S$ *and transition kernel* p *for any* $x, y \in S$, *and let* $T = \inf\{n \in \mathbb{Z}_+ : X_n = Y_n\}$. *Then:*

1) $\|p^n(x,\cdot) - p^n(y,\cdot)\|_{\mathrm{TV}} \le P_{x,y}[T > n]$ *for any* $x, y \in S$ *and* $n \in \mathbb{N}$.

2) $\alpha(p^n, K) \le \sup\limits_{x,y\in K} P_{x,y}[T > n]$.

Example (Lazy Random Walks). A lazy random walk on a graph is a random walk that stays in 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 a simple coupling argument we obtain bounds for total variation distances and mixing times on different graphs:

1) $\underline{S = \mathbb{Z}}$: Here the transition probabilities of the lazy simple random walk are $p(x, x+1) = p(x, x-1) = 1/4, p(x, x) = 1/2$, and $p(x, y) = 0$ 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 $1/2$ each. Hence only one of two copies is moving during each step so that the two random walks X_n and Y_n can not cross each other without meeting at the same position. The coupling time T is the hitting time of 0 for the process $X_n - Y_n$ which is a simple random walk on $\mathbb{Z}$. Hence $T < \infty$ $P_{x,y}$-almost surely, and

$$\lim_{n\to\infty} \|p^n(x,\cdot) - p^n(y,\cdot)\|_{\mathrm{TV}} = 0 \quad \text{for any } x, y \in S.$$

Nevertheless, a stationary distribution does not exist.

2) $\underline{S = \mathbb{Z}/(m\mathbb{Z})}$: On a discrete circle with m points we can use the analogue coupling for the lazy random walk. Again, $X_n - Y_n$ is a simple random walk on S, and T is the hitting time of 0. Hence by the Poisson equation,

$$E_{x,y}[T] = E^{\mathrm{RW}(\mathbb{Z})}_{|x-y|}[T_{\{1,2,\dots,m-1\}^c}] = |x-y| \cdot (m - |x-y|) \le \frac{1}{n} m^2.$$

Corollary 1.30 and Markov's inequality now implies that the TV-distance to the uniform distribution after n steps is bounded from above by

$$d_{\mathrm{TV}}(n, S) \le \alpha(p^n) \le \sup_{x,y} P_{x,y}[T > n] \le \frac{m^2}{4n}.$$

Hence $t_{\mathrm{mix}}(1/4) \le m^2$ which is a rather sharp upper bound.

3) $\underline{S = \{0,1\}^d}$: The lazy random walk on the hypercube $\{0,1\}^d$ coincides with the Gibbs sampler for the uniform distribution. Constructing a coupling similarly as before, 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 $c \geq 0$,

$$\begin{aligned} d_{\text{TV}}(d\log d + cd) &\leq P[T > d\log d + cd] \\ &\leq \sum_{k=1}^{d} \left(1 - \frac{1}{d}\right)^{\lceil d\log d + cd\rceil} \leq de^{-\frac{d\log d + cd}{d}} \leq e^{-c}, \end{aligned}$$

and hence

$$t_{\text{mix}}(\varepsilon) \leq d\log d + \log(\varepsilon^{-1})d.$$

Conversely the coupon collecting problem also shows that this upper bound is again almost sharp.

1.7.3 Conductance lower bounds

A simple and powerful way to derive lower bounds for mixing times due to constraints by bottlenecks is the conductance.

Exercise (Conductance and lower bounds for mixing times). Let p be a transition kernel on $(S, \mathcal{B})$ with stationary distribution μ. For sets $A, B \in \mathcal{B}$ with $\mu(A) > 0$, the **equilibrium flow** $Q(A, B)$ from A to B is defined by

$$Q(A, B) \;=\; (\mu \otimes p)(A \times B) \;=\; \int_A \mu(dx)\, p(x, B),$$

and the **conductance** of A is given by

$$\Phi(A) \;=\; \frac{Q(A, A^C)}{\mu(A)}\,.$$

The **bottleneck ratio (isoperimetric constant)** Φ_* is defined as

$$\Phi_* = \min_{A:\mu(A)\leq 1/2} \Phi(A).$$

Let $\mu_A(B) = \mu(B|A)$ denote the conditioned measure on A.

a) Show that for any $A \in \mathcal{B}$ with $\mu(A) > 0$,

$$\|\mu_A p - \mu_A\|_{TV} \;=\; (\mu_A p)(A^C) \;=\; \Phi(A).$$

Hint: Prove first that

(i) $(\mu_A p)(B) - \mu_A(B) \leq 0$ *for any measurable* $B \subseteq A$*, and*

(ii) $(\mu_A p)(B) - \mu_A(B) = (\mu_A p)(B) \geq 0$ *for any measurable* $B \subseteq A^C$*.*

b) Conclude that

$$\|\mu_A - \mu\|_{TV} \leq t\Phi(A) + \|\mu_A p^t - \mu\|_{TV} \qquad \text{for any } t \in \mathbb{Z}_+.$$

c) Hence prove the lower bound

$$t_{mix}\left(\frac{1}{4}\right) \geq \frac{1}{4\Phi_*}.$$

Chapter 2

Ergodic averages

Suppose that (X_n, P_x) is a canonical time-homogeneous Markov chain with transition kernel p. 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

$$\mu = \mu p.$$

A probability measure μ with this property is called a **stationary (initial) distribution** or an **invariant probability measure for the transition kernel $\mathbf{p}$**. 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, and we will briefly discuss the validity of ergodic theorems for Markov chains that are not started in stationary. Moreover, we will study the fluctuations of ergodic averages around their limit both asymptotically and non-asymptotically. 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 Breimann $[XXX]$, Durrett $[XXX]$ and Varadhan $[XXX]$.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.

2.1.1 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 (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

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

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 relations 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

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

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

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

and Θ invariance of F implies

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

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.

2.1.2 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 p satisfying $\mu = \mu p$. Note that by stationarity, the map $f \mapsto pf$ is a contraction on $\mathcal{L}^2(\mu)$. Indeed, by the Cauchy-Schwarz inequality,

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

In particular,

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

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

Theorem 2.1 (Characterizations of ergodicity for 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 $p1_B = 1_B$ 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

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

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 $p1_B = 1_B$ 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 $ph = h$, 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 $x \in S$,

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

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 $\{0, 1\}$.

$\square$

The third condition in Theorem 2.1 is reminiscent of the definition of irreducibility. However, there is an important difference as the following example shows:

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 $p1_B = 1_B$ μ-almost surely such that

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

Example (Ergodicity and irreducibility). Consider the constant Markov chain on $S = \{0, 1\}$ with transition probabilities $p(0, 0) = p(1, 1) = 1$. Obviously, any probability measure on S is a stationary distribution for p. The matrix p is not irreducible, for instance $p1_{\{1\}} = 1_{\{1\}}$. Nevertheless, condition 3) is satisfied and P_μ is ergodic if (and only if) μ is a Dirac measure.

2.1.3 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 2.2 (Birkhoff). *Suppose that $P = P \circ \Theta^{-1}$ and let $p \in [1, \infty)$. Then as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F|\mathcal{J}] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P) \tag{2.1.1}$$

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

$$\frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P). \tag{2.1.2}$$

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 $n \to \infty$,

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) = \frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[f(X_0)|\mathcal{J}] \qquad (2.1.3)$$

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 (2.1.3) 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,

$$\frac{1}{n}\sum_{i=0}^{n-1} F(X_i, X_{i+1}, \dots) = \frac{1}{n}\sum_{i=0}^{n-1} F \circ \Theta^i \to E[F|\mathcal{J}] \qquad (2.1.4)$$

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 $F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P)$,

$$\frac{1}{n}\sum_{i=0}^{n-1} U^i F \to \pi F \qquad \text{in } L^p(\Omega, \mathfrak{A}, P) \text{ as } n \to \infty, \text{ where } \pi F := E[F|\mathcal{J}]. \qquad (2.1.5)$$

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

$$H_0 = L^2(\Omega, \mathcal{J}, P) = \left\{F \in L^2(\Omega, \mathfrak{A}, P) : UF = F\right\} \qquad (2.1.6)$$

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.,

$$L^2(\Omega, \mathfrak{A}, P) = H_0 \oplus \overline{(U - I)(L^2)}. \tag{2.1.7}$$

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

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

The L^2 convergence in (2.1.5) 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 (2.1.5) holds for $p = 2$ and any $F \in \mathcal{L}^2(\Omega, \mathfrak{A}, P)$.

Notation (Averaging operator). *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

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

Proof of Theorem 2.2. 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 (2.1.1) 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 $F \in H_0 + H_1$,

$$A_nF - \pi F \to 0 \quad \text{in } L^\infty(\Omega, \mathfrak{A}, P). \tag{2.1.8}$$

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

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

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,

$$A_nF \to \pi F \quad \text{in } L^2(\Omega, \mathfrak{A}, P) \tag{2.1.9}$$

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 (2.1.9) 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 (2.1.7),

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

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

$$A_nF \to \pi F \quad \text{in } L^p(\Omega, \mathfrak{A}, P) \tag{2.1.10}$$

by (2.1.9) and the dominated convergence theorem. Since A_n and π are contractions on each L^p space, the convergence in (2.1.10) extends to all $F \in L^p(\Omega, \mathfrak{A}, P)$ by an $\varepsilon/3$ argument.

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

$$A_nF \to \pi F \quad P\text{-almost surely} \tag{2.1.11}$$

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 $\varepsilon > 0$,

$$\begin{aligned} P[\limsup A_n F - \liminf A_n F \geq \varepsilon] &\leq P[\sup_n |A_n F - A_n F_k| \geq \varepsilon/2] \\ &= P[\sup_n |A_n(F - F_k)| \geq \varepsilon/2]. \end{aligned} \tag{2.1.12}$$

Hence we are done if we can show for any $\varepsilon > 0$ that the right hand side in (2.1.12) converges to 0 as $k \to \infty$. Since $E[|F - F_k|] \to 0$, the proof is now completed by Lemma 2.3 below.

□

Lemma 2.3 (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) $E[F; \max_{1\leq i\leq n} A_i F \geq 0] \geq 0$ *for any* $n \in \mathbb{N}$,

2) $P[\sup_{n\in\mathbb{N}} |A_n F| \geq c] \leq \frac{1}{c} E[|F|]$ *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 $M_n = \max_{1\leq i\leq n}(F + F\circ\Theta + \cdots + F\circ\Theta^{i-1})$, and let $B = \{M_n \geq 0\} = \{\max_{1\leq i\leq n} A_i F \geq 0\}$. Then $M_n = F + M_{n-1}^+ \circ \Theta$, and hence

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

Taking expectations we obtain

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

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 $F \geq 0$ and $c \in (0, \infty)$,

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

by 1). Therefore,

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

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

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

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

□

2.1.4 Application to 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 p. Then Θ is measure-preserving w.r.t. P_μ if and only if μ is a stationary distribution for p. Furthermore, by Theorem 2.1, the measure P_μ is ergodic if and only if any set $B \in \mathcal{B}$ such that $p1_B = 1_B$ μ-almost surely has measure $\mu(B) \in \{0, 1\}$. In this case, Birkhoff's theorem has the following consequences:

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

$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \to \int f d\mu \quad P_\mu\text{-almost surely as } n \to \infty. \tag{2.1.13}$$

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 $A, B \in \mathcal{B}$,

$$\frac{1}{n} \sum_{i=0}^{n-1} 1_{A \times B}(X_i, X_{i+1}) \to E[1_{A \times B}(X_0, X_1)] = \int_A \mu(dx) p(x, B) \tag{2.1.14}$$

P_μ-a.s. as $n \to \infty$. This is applied in statistics of Markov chains for estimating 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 these important questions later - in particular in Sections 2.4 and 2.5. For the moment we conclude with some preliminary observations concerning the second question:

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 νp^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 νp^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 p then the limits $E_\nu[F|\mathcal{J}]$ and $E_\mu[F|\mathcal{J}]$ of the ergodic averages $A_n F$ w.r.t. P_ν and P_μ respectively do *not coincide*.

Exercise (Ergodicity of stationary Markov chains). Suppose that μ is a stationary distribution for the transition kernel p 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 $B \in \mathcal{B}$,

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

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

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

(iv) Any $B \in \mathcal{B}$ such that $p1_B = 1_B$ μ-a.s. has measure $\mu(B) \in \{0,1\}$.

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.

2.2.1 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

$$\Theta_0 = \mathrm{id}_\Omega \quad \text{and} \quad \Theta_t \circ \Theta_s = \Theta_{t+s} \quad \text{for any } t,s \geq 0. \tag{2.2.1}$$

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 (2.2.1). 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 2.4 (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 $F \in \mathcal{L}^p(\Omega, \mathfrak{A}, P)$,*

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t F \circ \Theta_s \, ds = E[F|\mathcal{J}] \quad P\text{-almost surely and in } L^p(\Omega, \mathfrak{A}, P). \tag{2.2.2}$$

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 (2.2.1),

$$A_n F = \hat{A}_n \hat{F} \quad \text{for any } n \in \mathbb{N}, \text{ where } \quad \hat{A}_n \hat{F} := \frac{1}{n} \sum_{i=0}^{n-1} \hat{F} \circ \Theta_i$$

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,

$$\lim_{t \to \infty} A_t F = \lim_{n \to \infty} \hat{A}_n \hat{F} = E[\hat{F}|\hat{\mathcal{J}}] \quad P\text{-a.s. and in } L^p \text{ for any } p \in [1, \infty), \tag{2.2.3}$$

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 (2.2.3) coincides with the conditional expectation $E[F|\mathcal{J}]$ P-almost surely. To this end note that the limit superior of A_tF as $t \to \infty$ is $\mathcal{J}$-measurable, since

$$(A_tF) \circ \Theta_s = \frac{1}{t}\int_0^t F \circ \Theta_u \circ \Theta_s \, du = \frac{1}{t}\int_0^t F \circ \Theta_{u+s} \, du = \frac{1}{t}\int_s^{s+t} F \circ \Theta_u \, du$$

has the same limit superior as A_tF for any $s \in [0, \infty)$. Since L^1 convergence holds,

$$\lim_{t\to\infty} A_tF = E[\lim A_tF|\mathcal{J}] = \lim E[A_tF|\mathcal{J}] = \lim_{t\to\infty} \frac{1}{t}\int_0^t E[F \circ \Theta_s|\mathcal{J}] \, ds$$

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

$$\lim_{t\to\infty} A_tF = E[F|\mathcal{J}] \quad P\text{-almost surely}.$$

Step 3: Extension to general $\mathbf{F} \in \mathcal{L}^\mathbf{p}$. 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 (2.2.2) 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 2.3. For $t \geq 1$,

$$|A_tF| \leq \frac{1}{t}\int_0^{\lfloor t\rfloor+1} |F \circ \Theta_s| \, ds = \frac{\lfloor t\rfloor + 1}{t} \hat{A}_{\lfloor t\rfloor+1}|\hat{F}| \leq 2\hat{A}_{\lfloor t\rfloor+1}|\hat{F}|.$$

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 2.2 above.

□

The ergodic theorem implies the following alternative characterizations of ergodicity:

Corollary 2.5 (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.

2.2.2 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

$$\frac{d}{dt}\Theta_t(\omega) = b(\Theta_t(\omega)), \quad \Theta_0(\omega) = \omega \quad \text{for any } \omega \in \mathbb{R}^d. \tag{2.2.4}$$

For a smooth function $F : \mathbb{R}^d \to \mathbb{R}$ and $t \in \mathbb{R}$ let

$$(U_t F)(\omega) = F(\Theta_t(\omega)).$$

Then the flow equation (2.2.4) implies the **forward equation**

$$\frac{d}{dt}U_t F = \dot{\Theta}_t \cdot (\nabla F) \circ \Theta_t = (b \cdot \nabla F) \circ \Theta_t, \quad \text{i.e.,}$$

$$\text{(F)} \qquad \frac{d}{dt}U_t F = U_t \mathcal{L} F \qquad \text{where} \quad \mathcal{L}F = b \cdot \nabla F$$

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:

$$\frac{d}{dt}H(q,p) = \frac{\partial H}{\partial q}(q,p)\cdot\frac{dq}{dt} + \frac{\partial H}{\partial q}(q,p)\cdot\frac{dp}{dt} = (b\cdot\nabla H)(q,p) = 0 \tag{2.2.5}$$

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.

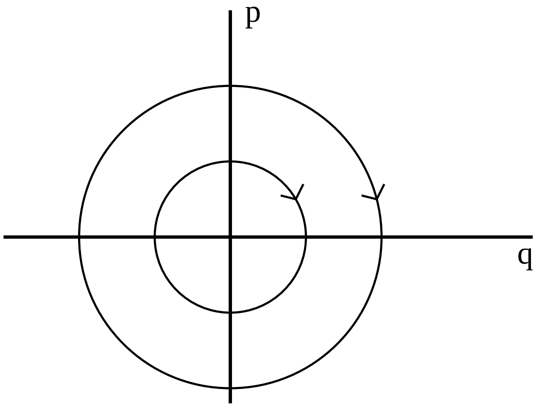


Figure 2.1: Trajectories of harmonic oscillator

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

$$\mathrm{div}(\varrho b) = b\cdot\nabla\varrho + \varrho\,\mathrm{div}(b) = (g'\circ H)\,(b\cdot\nabla H) + \varrho\left(\frac{\partial^2 H}{\partial q\partial p} - \frac{\partial^2 H}{\partial p\partial q}\right) = 0$$

by (2.2.5). 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 dynamic. 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 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,

$$X_t\circ\Theta_s = X_{t+s} \quad \text{for any } t,s\geq 0.$$

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

$$m(t) = E[X_t], \quad c(s,t) = \mathrm{Cov}(X_s, X_t), \quad s, t \geq 0.$$

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

$$\mathrm{Var}\left(\frac{1}{t}\int_0^t X_s\, ds\right) \to 0 \quad \text{as } t \to \infty.$$

On the other hand, by Fubini's theorem,

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{t}\int_0^t X_s\, ds\right) &= \mathrm{Cov}\left(\frac{1}{t}\int_0^t X_s\, ds, \frac{1}{t}\int_0^t X_u\, du\right) \\
&= \frac{1}{t^2}\int_0^t\int_0^t \mathrm{Cov}\left(X_s, X_u\right)\, duds = \frac{1}{2t^2}\int_0^t\int_0^s r(s-u)\, duds \\
&= \frac{1}{2t^2}\int_0^t (t-v) r(v)\, dv = \frac{1}{2t}\int_0^t \left(1 - \frac{v}{t}\right) r(v)\, dv \\
&\sim \frac{1}{2t}\int_0^t r(v) dv \quad \text{asymptotically as } t \to \infty.
\end{aligned}$$

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" [20].

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. Strook, "Probability Theory: An Analytic View". [34]. 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 to the study of stationary random fields. Here we just mention briefly two typical examples:

Example (Massless Gaussian free field on $\mathbb{Z}^\mathbf{d}$). 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

$$\text{“}\, P(d\omega) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{\substack{s,t\in\mathbb{Z}^d \\ |s-t|=1}} |\omega_t - \omega_s|^2 \right) \prod_{s\in\mathbb{Z}^d} d\omega_s \text{”}. \tag{2.2.6}$$

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 (2.2.6) 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 $G(s,t) = \sum\limits_{n=0}^{\infty} p^n(s,t)$ is the Green's function of the Random Walk on $\mathbb{Z}^d$. The connection to the informal expression in (2.2.6) is made by observing that the generator of the random walk is the discrete Laplacian $\Delta_{\mathbb{Z}^d}$, and the informal density in (2.2.6) takes the form

$$Z^{-1} \exp\left(-\frac{1}{2} \left(\omega, \Delta_{\mathbb{Z}^d}\omega\right)_{l^2(\mathbb{Z}^d)} \right).$$

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 correlation 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

$$p(\omega, x, y) = q\left(\Theta_x(\omega), y - x\right) \quad \text{for any } \omega \in \Omega \text{ and } x, y \in \mathbb{Z}^d.$$

For any fixed $\omega \in \Omega$, $p(\omega, \cdot)$ 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 $p(\omega, \cdot, \cdot)$ 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" [3].

2.2.3 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

$$\mu p_t = \mu \quad \text{for any } t \in [0,\infty).$$

The existence of stationary distributions can be shown similarly to the discrete time case:

Theorem 2.6 (Krylov-Bogolinbov). *Suppose that the family*

$$\nu \overline{p}_t = \frac{1}{t}\int_0^t \nu p_s\, ds, \quad t \geq 0,$$

of probability measures on S is tight for some $\nu \in \mathcal{P}(S)$. Then there exists a stationary distribution μ of $(p_t)_{t\geq 0}$.

The proof of this and of the next theorem are left as exercises.

Theorem 2.7 (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.

One way to verify ergodicity is the strong Feller property:

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

Corollary 2.8. *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 as connected support.*

Proof. Let $B \in \mathcal{B}$ such that

$$p_t 1_B = 1_B \quad \mu\text{-a.e. for any } t \geq 0. \tag{2.2.7}$$

By Theorem 2.7 it suffices to show $\mu(B) \in \{0, 1\}$. If p_t is strong Feller for some t then $p_t 1_B$ is a continuous function. Therefore, by (2.2.7) and since the support of μ is connected, either $p_t 1_B \equiv 0$ or $p_t 1_B \equiv 1$ on supp(μ). Hence

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

□

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

$$p_t(x, y) = \sum_{n\in\mathbb{Z}} p_t(x, y + n) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{|x-y-n|^2}{2t}} \qquad \text{for any } t > 0 \text{ and } x, y \in [0, 1).$$

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 2.8, 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 2.8 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 $(p_t)_{t\geq 0}$.

2.3.1 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 2.9 (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 exist 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

$$C_e = \{x \in C : x_1, x_2 \in C\backslash\{x\}, \alpha \in (0,1) : x = \alpha x_1 + (1-\alpha)x_2\}.$$

Theorem 2.10 (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 2.9 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" [36]. 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.

□

2.3.2 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 2.10 for Markov processes. It can be either deduced from Theorem 2.10 or proven independently.

Theorem 2.11 (Structure and extremals of $\mathcal{S}(\mathbf{p})$). *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 XXX 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 Quantitative bounds & CLT 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, $p_t = p^t$ where p 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, 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 $\mu(f) = \int f d\mu$ for any $f \in \mathcal{L}^1(\mu)$. In this section, we study the asymptotics of the functions of $A_t f$ around $\mu(f)$ as $t \to \infty$ for $f \in \mathcal{L}^2(\mu)$.

2.4.1 Bias and variance of stationary ergodic averages

Theorem 2.12 (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

$$\mathrm{Var}_{P_\mu}[A_t f] = \frac{1}{t}\,\mathrm{Var}_\mu(f) + \frac{2}{t}\sum_{k=1}^{t}\left(1-\frac{k}{t}\right)\mathrm{Cov}_\mu(f, p^k f) \quad \textit{in discrete time,}$$

$$\mathrm{Var}_{P_\mu}[A_t f] = \frac{2}{t}\int_0^t \left(1-\frac{r}{t}\right)\mathrm{Cov}_\mu(f, p_r f)dr \quad \textit{in continuous time, respectively.}$$

3) Suppose that the series $Gf_0 = \sum_{k=0}^{\infty} p^k f_0$ 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

$$\lim_{t\to\infty} t \cdot \mathrm{Var}_{P_\mu}[A_t f] = \sigma_f^2, \quad \textit{where}$$

$$\sigma_f^2 = \mathrm{Var}_\mu(f) + 2\sum_{k=1}^{\infty}\mathrm{Cov}_\mu(f, p^k f) = 2(f_0, Gf_0)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)}$$

in the discrete time case, and

$$\sigma_f^2 = \int_0^\infty \mathrm{Cov}_\mu(f, p_s f)ds = 2(f_0, Gf_0)_{L^2(\mu)}$$

in the continuous time case, respectively.

Remark. 1) The asymptotic variance equals

$$\sigma_f^2 = \mathrm{Var}_{P_\mu}[f(X_0)] + 2\sum_{k=1}^{\infty}\mathrm{Cov}_{P_\mu}[f(X_0), f(X_k)],$$

$$\sigma_f^2 = \int_0^\infty \mathrm{Cov}_{P_\mu}[f(X_0), f(X_s)]ds \quad \text{respectively.}$$

If Gf_0 exists then the variance of the ergodic averages behaves asymptotically as σ_f^2/t.

2) The statement 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 2.5 below.

Proof of Theorem 2.12:

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,

$$\begin{aligned}\mathrm{Var}_{P_\mu}\left[A_tf\right]&=\mathrm{Cov}_{P_\mu}\left[\frac{1}{t}\int_0^t f(X_s)\,ds,\frac{1}{t}\int_0^t f(X_u)\,du\right]\\&=\frac{2}{t^2}\int_0^t\int_0^u\mathrm{Cov}_{P_\mu}\left[f(X_s),f(X_u)\right]\,dsdu\\&=\frac{2}{t^2}\int_0^t\int_0^u\mathrm{Cov}_\mu(f,p_{u-s}f)\,dsdu\\&=\frac{2}{t^2}\int_0^t(t-r)\,\mathrm{Cov}_\mu(f,p_rf)\,dr.\end{aligned}$$

3) Note that by stationarity, $\mu(p_rf)=\mu(f)$, and hence

$$\mathrm{Cov}_\mu(f,p_rf)=\int f_0\,p_rf_0\;d\mu\qquad\text{for any } r\geq 0.$$

Therefore, by 2) and Fubini's theorem,

$$\begin{aligned}t\cdot\mathrm{Var}_{P_\mu}\left[A_tf\right]&=2\int_0^t\left(1-\frac{r}{t}\right)\int f_0\;p_rf_0\,d\mu dr\\&=2\left(f_0,\int_0^t\left(1-\frac{r}{t}\right)p_rf_0\,dr\right)_{L^2(\mu)}\\&\to 2\left(f_0,\int_0^\infty p_rf_0\,dr\right)_{L^2(\mu)}\qquad\text{as } t\to\infty\end{aligned}$$

provided the integral $\int_0^\infty p_rf_0\,dr$ converges in $L^2(\mu)$. Here the last conclusion holds since $L^2(\mu)$-convergence of $\int_0^t p_rf_0\,dr$ as $t\to\infty$ implies that

$$\int_0^t\frac{r}{t}\,p_rf_0\,dr=\frac{1}{t}\int_0^t\int_0^r p_rf_0\,dsdr=\frac{1}{t}\int_0^t\int_s^t p_rf_0\,drds\to 0\text{ in }L^2(\mu)\text{ as }t\to\infty.$$

□

Remark (**Potential operator, existence of asymptotic variance**). The theorem states that the asymptotic variance of $\sqrt{t}A_tf$ 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 in the exercise 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,

$$-\mathcal{L}Gf_0 = (I - p)\sum_{n=0}^{\infty} p^n f_0 = f_0$$

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 3.11 below.

Exercise (Sufficient conditions for existence of the asymptotic variance). Prove that in the continuous time case, $Gf_0 = \int_0^\infty p_t f_0$ 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) **$\mathbf{L^2}$ 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 $|\operatorname{Cov}_{P_\mu}[f(X_0), f(X_t)]|$ or the $L^2(\mu)$ norms $\|p_t f_0\|_{L^2(\mu)}$ are bounded by an integrable function $r(t)$.

2.4.2 Central limit theorem for Markov chains

We now restrict ourselves to the discrete time case. Let $f \in \mathcal{L}^2(\mu)$, and suppose that the asymptotic variance

$$\sigma_f^2 = \lim_{n\to\infty} n \operatorname{Var}_{P_\mu}[A_n f]$$

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

$$\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(X_i) \xrightarrow{\mathcal{D}} N(0, \sigma_f^2) \tag{2.4.1}$$

where "$\xrightarrow{\mathcal{D}}$" stands for convergence in distribution. The key idea is to use the martingale problem in order to reduce (2.4.1) 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 $n \geq 0$, and hence

$$g(X_n) - g(X_0) = M_n + \sum_{k=0}^{n-1} (\mathcal{L}g)(X_k) \tag{2.4.2}$$

where (M_n) is a square-integrable $(\mathcal{F}_n^X)$ martingale with $M_0 = 0$ w.r.t. P_μ, and $\mathcal{L}g = pg - g$. 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 $Gf = \sum\limits_{n=0}^{\infty} p^n f$ converges in $L^2(\mu)$. Then by (2.4.2),

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} f(X_k) = \frac{M_n}{\sqrt{n}} + \frac{g(X_0) - g(X_n)}{\sqrt{n}}. \tag{2.4.3}$$

As $n \to \infty$, the second summand converges to 0 in $L^2(P_\mu)$. Therefore, (2.4.1) is equivalent to a central limit theorem for the martingale (M_n). Explicitly,

$$M_n = \sum_{i=1}^{n} Y_i \quad \text{for any } n \geq 0,$$

where the martingale increments Y_i are given by

$$\begin{aligned} Y_i = M_i - M_{i-1} &= g(X_i) - g(X_{i-1}) - (\mathcal{L}g)(X_{i-1}) \\ &= g(X_i) - (pg)(X_{i-1}). \end{aligned}$$

These increments form a stationary sequence w.r.t. P_μ. Thus we can apply the following theorem:

Theorem 2.13 (CLT for martingales with stationary increments). *Let $(\mathcal{F}_n)$ be a filtration on a probability space $(\Omega, \mathfrak{A}, P)$. Suppose that $M_n = \sum\limits_{i=1}^{n} Y_i$ 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*

$$\frac{1}{n} \sum_{i=1}^{n} Y_i^2 \to \sigma^2 \quad \text{in } L^1(P) \text{ as } n \to \infty \tag{2.4.4}$$

then

$$\frac{1}{\sqrt{n}} M_n \xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad \text{w.r.t. } P. \tag{2.4.5}$$

The proof of Theorem 2.13 will be given at the end of this section. Note that by the ergodic theorem, the condition (2.4.4) is satisfied with $\sigma^2 = E[Y_i^2]$ if the process (Y_i, P) is ergodic. As consequence of Theorem 2.13 and the considerations above, we obtain:

Corollary 2.14 (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 p, and let $f \in \mathcal{L}^2(\mu)$. Suppose that there exists a function $g \in \mathcal{L}^2(\mu)$ such that*

$$-\mathcal{L}g = f - \mu(f). \tag{2.4.6}$$

Then as $n \to \infty$,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}(f(X_k) - \mu(f)) \xrightarrow{\mathcal{D}} N(0, \sigma_f^2), \qquad \textit{where}\\ &\sigma_f^2 = 2\operatorname{Cov}_\mu(f, g) - \operatorname{Var}_\mu(f).\end{aligned}$$

Remark. Recall that (2.4.6) is satisfied with $g = G(f - \mu(f))$ if it exists.

Proof. Let $Y_i = g(X_i) - (pg)(X_{i-1})$. 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:

$$\begin{aligned}E_\mu[Y_1^2] &= E_\mu[(g(X_1) - (pg)(X_0))^2]\\ &= \int \mu(dx) E_x[g(X_1)^2 - 2g(X_1)(pg)(X_0) + (pg)(X_0)^2]\\ &= \int (pg^2 - 2(pg)^2 + (pg)^2)d\mu = \int g^2\, d\mu - \int (pg)^2 d\mu\\ &= (g - pg, g + pg)_{L^2(\mu)} = 2(f_0, g)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)} = \sigma_f^2.\end{aligned}$$

Here $f_0 := f - \mu(f) = -\mathcal{L}g = g - pg$ by assumption. The martingale CLT 2.13 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 $L^2(P_\mu)$. ☐

Some explicit bounds on σ_f^2 are given in the next section. We conclude this section with a proof of the CLT for martingales with stationary increments:

2.4.3 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

$$E[Y_i|\mathcal{F}_{i-1}] = 0 \quad P\text{-a.s. for any } i \in \mathbb{N} \tag{2.4.7}$$

w.r.t. a filtration $(\mathcal{F}_n)$. We now prove the central limit Theorem 2.13, i.e.,

$$\frac{1}{n}\sum_{i=1}^n Y_i^2 \to \sigma^2 \text{ in } L^1(P) \Rightarrow \frac{1}{\sqrt{n}}M_n \stackrel{\mathcal{D}}{\to} N(0,\sigma^2). \tag{2.4.8}$$

Proof of Theorem 2.13. 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 $p \in \mathbb{R}$,

$$\begin{aligned} &E\left[e^{ipM_n/\sqrt{n}}\right] \to \varphi(p) \quad \text{ as } n \to \infty, \text{ or, equivalently,} \\ &E\left[e^{ipM_n/\sqrt{n}+\sigma^2p^2/2} - 1\right] \to 0 \quad \text{ as } n \to \infty. \end{aligned} \tag{2.4.9}$$

Let

$$Z_{n,k} := \exp\left(i\frac{p}{\sqrt{n}}M_k + \frac{\sigma^2p^2}{2}\frac{k}{n}\right), \quad k = 0,1,\ldots,n.$$

Then the left-hand side in (2.4.9) is given by

$$\begin{aligned} E[Z_{n,n} - Z_{n,0}] &= \sum_{k=1}^n E[Z_{n,k} - Z_{n,k-1}] \\ &= \sum_{k=1}^n E\left[Z_{n,k-1}\cdot E\left[\exp\left(\frac{ip}{\sqrt{n}}Y_k + \frac{\sigma^2p^2}{2n}\right) - 1|\mathcal{F}_{k-1}\right]\right]. \end{aligned} \tag{2.4.10}$$

The random variables $Z_{n,k-1}$ are uniformly bounded independently of n and k, and by a Taylor approximation and (2.4.7),

$$\begin{aligned} E\left[\exp\left(\frac{ip}{\sqrt{n}}Y_k + \frac{\sigma^2p^2}{2n}\right) - 1|\mathcal{F}_{k-1}\right] &= E\left[\frac{ip}{\sqrt{n}}Y_k - \frac{p^2}{2n}\left(Y_k^2 - \sigma^2\right)|\mathcal{F}_{k-1}\right] + R_{n,k} \\ &= -\frac{p^2}{2n}E[Y_k^2 - \sigma^2|\mathcal{F}_{k-1}] + R_{n,k} \end{aligned}$$

with a remainder $R_{n,k}$ of order $o(1/n)$. Hence by (2.4.10),

$$E\left[e^{ipM_n/\sqrt{n}+\sigma^2p^2/2} - 1\right] = -\frac{p^2}{2n}\sum_{k=1}^n E\left[Z_{n,k-1}\cdot(Y_k^2 - \sigma^2)\right] + r_n$$

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 (2.4.4) after replacing $Z_{n,k-1}$ by $Z_{n,jl}$ on the j-th block. We first estimate

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n} E[Z_{n,k-1}(Y_k^2-\sigma^2)]\right| \\
&\le \frac{1}{n}\sum_{j=0}^{\lfloor n/l \rfloor}\left(\left|E\left[Z_{n,jl}\sum_{\substack{jl\le k<(j+1)l\\ k<n}}(Y_k^2-\sigma^2)\right]\right| + \sup_{\substack{jl\le k<(j+1)l\\ k<n}} E[|Z_{n,k-1}-Z_{n,jl}|\cdot|Y_k^2-\sigma^2|]\right) \\
&\le c_1\cdot E\left[\left|\frac{1}{l}\sum_{k=1}^{l}(Y_k^2-\sigma^2)\right|\right] + \frac{c_2}{n} + c_3 \sup_{1\le k<l} E\left[|Z_{n,k-1}-1|\cdot\left|Y_k^2-\sigma^2\right|\right]. \qquad (2.4.11)
\end{aligned}$$

Here we have used that the random variables $Z_{n,k}$ are uniformly bounded, the sequence (Y_k) is stationary, and

$$|Z_{n,k-1} - Z_{n,jl}| \le |Z_{n,jl}| \cdot \left|\exp\left(ip(M_k - M_{jl}) + \frac{\sigma^2 p^2}{2}\frac{k-jl}{n}\right) - 1\right|$$

where the exponential has the same law as $Z_{n,k-jl}$ by stationarity. By the assumption (2.4.4), the first term on the right-hand side of (2.4.11) can be made arbitrary small by choosing l sufficiently large. Moreover, for any fixed $l \in \mathbb{N}$, the two other summands converge to 0 as $n \to \infty$ by dominated convergence. Hence the left-hand side in (2.4.11) also converges to 0 as $n \to \infty$, and thus (2.4.4) holds. $\square$

2.5 Asymptotic stationarity & MCMC integral estimation

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 p satisfying $\mu = \mu p$, and $b, n \in \mathbb{N}$ are sufficiently large integers. 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. A typical example of a Markov chain used in MCMC methods is the Gibbs sampler that has been introduced in Section 1.5.4 above. The second important class of Markov chains applied in MCMC are Metropolis-Hastings chains.

Example (Metropolis-Hastings method). 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 p with stationary distribution μ can be constructed by proposing moves according to an absolutely continuous proposal kernel

$$q(x, dy) = q(x, y)\,\lambda(dy)$$

with strictly positive density $q(x, y)$, and accepting a proposed move from x to y with probability

$$\alpha(x, y) = \min\left(1, \frac{\mu(y)q(y, x)}{\mu(x)q(x, y)}\right).$$

If a proposed move is not accepted then the Markov chain stays at its current position x. The transition kernel is hence given by

$$p(x, dy) = \alpha(x, y)q(x, dy) + r(x)\delta_x(dy)$$

where $r(x) = 1 - \int \alpha(x, y)q(x, dy)$ is the rejection probability for the next move from x. Typical examples of Metropolis-Hastings methods are Random Walk Metropolis algorithms where q is the transition kernel of a random walk. Note that if q is symmetric then the acceptance probability simplifies to

$$\alpha(x, y) = \min\left(1, \mu(y)/\mu(x)\right).$$

Lemma 2.15 (Detailed balance). *The transition kernel p of a Metropolis-Hastings chain satisfies the detailed balance condition*

$$\mu(dx)p(x, dy) = \mu(dy)p(y, dx). \tag{2.5.1}$$

In particular, μ is a stationary distribution for p.

Proof. On $\{(x, y) \in S \times S : x \neq y\}$, the measure $\mu(dx)p(x, dy)$ is absolutely continuous w.r.t. $\lambda \otimes \lambda$ with density

$$\mu(dx)\alpha(x, y)q(x, y) = \min\big(\mu(x)q(x, y), \mu(y)q(y, x)\big).$$

The detailed balance condition (2.5.1) follows, since this expression is a symmetric function of x and y. □

A central problem in the mathematical study of MCMC methods for the estimation of integrals w.r.t. μ is the derivation of bounds for the approximation error

$$A_{b,n}f - \mu(f) = A_{b,n}f_0,$$

where $f_0 = f - \mu(f)$. 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 both asymptotic and non-asymptotic bounds for ergodic averages of non-stationary Markov chains.

2.5.1 Asymptotic bounds for ergodic averages

As above, we assume that (X_n, P) is a time-homogeneous Markov chain with transition kernel p, stationary distribution μ, and initial distribution ν.

Theorem 2.16 (Ergodic theorem and CLT for non-stationary Markov chains). *Let $b, n \in \mathbb{N}$.*

1) The bias of the estimator $A_{b,n}f$ is bounded by

$$|E[A_{b,n}f] - \mu(f)| \leq \|\nu p^b - \mu\|_{TV} \|f_0\|_{sup}.$$

2) If $\|\nu p^n - \mu\|_{TV} \to 0$ as $n \to \infty$ then

$$A_{b,n}f \to \mu(f) \quad P\text{-a.s. for any } f \in \mathcal{L}^1(\mu), \text{ and}$$

$$\sqrt{n}\,(A_{b,n}f - \mu(f)) \xrightarrow{\mathcal{D}} N(0, \sigma_f^2) \text{ for any } f \in \mathcal{L}^2(\mu) \text{ s.t. } Gf_0 = \sum_{n=0}^{\infty} p^n f_0 \text{ converges in } L^2(\mu),$$

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. 1) Since $E[A_{b,n}f] = \frac{1}{n} \sum_{i=b}^{b+n-1} (\nu p^i)(f)$, the bias is bounded by

$$\begin{aligned} |E[A_{b,n}f] - \mu(f)| &= |E[A_{b,n}f_0] - \mu(f_0)| \\ &\leq \frac{1}{n} \sum_{i=b}^{b+n-1} |(\nu p^i)(f_0) - \mu(f_0)| \leq \frac{1}{n} \sum_{i=b}^{b+n-1} \|\nu p^i - \mu\|_{\mathrm{TV}} \cdot \|f_0\|_{\mathrm{sup}}. \end{aligned}$$

The assertion follows since the total variation distance $\|\nu p^i - \mu\|_{\mathrm{TV}}$ from the stationary distribution μ is a decreasing function of i.

2) If $\|\nu p^n - \mu\|_{\mathrm{TV}} \to 0$ then one can show that there is a coupling (X_n, Y_n) of the Markov chains with transition kernel p and initial distributions ν and μ such that the coupling time

$$T = \inf\{n \geq 0 : X_n = Y_n \text{ for } n \geq T\}$$

is almost surely finite (Exercise). We can then approximate $A_{b,n}f$ by ergodic averages for the stationary Markov chain (Y_n):

$$A_{b,n}f = \frac{1}{n}\sum_{i=b}^{b+n-1} f(Y_i) + \frac{1}{n}\sum_{i=b}^{b+n-1} (f(X_i) - f(Y_i))1_{\{i<T\}}.$$

The second sum is constant for $b+n \geq T$, so $\frac{1}{n}$ times the sum converges almost surely to zero, whereas the ergodic theorem and the central limit theorem apply to the first term on the right hand side. This proves the assertion.

□

To apply the theorem in practice, bounds for the asymptotic variance are required. One possibility for deriving such bounds is to estimate the contraction coefficient of the transition kernels on the orthogonal complement

$$L_0^2(\mu) = \{f \in L^2(\mu) : \mu(f) = 0\}$$

of the constants in the Hilbert space $L^2(\mu)$. Indeed, let

$$\gamma(p) = \|p\|_{L_0^2(\mu)\to L_0^2(\mu)} = \sup_{f\perp 1} \frac{\|pf\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}}$$

denote the operator norm of p on $L_0^2(\mu)$. If

$$c := \sum_{n=0}^{\infty} \gamma(p^n) < \infty \tag{2.5.2}$$

then $Gf_0 = \sum\limits_{n=0}^{\infty} p^n f_0$ converges for any $f \in \mathcal{L}^2(\mu)$, i.e., the asymptotic variances σ_f^2 exist, and

$$\begin{aligned} \sigma_f^2 &= 2(f_0, Gf_0)_{L^2(\mu)} - (f_0, f_0)_{L^2(\mu)} \\ &\leq (2c-1)\|f_0\|_{L^2(\mu)}^2 = (2c-1)\operatorname{Var}_\mu(f). \end{aligned} \tag{2.5.3}$$

A sufficient condition for (2.5.2) to hold is $\gamma(p) < 1$; in that case

$$c \leq \sum_{n=0}^{\infty} \gamma(p)^n = \frac{1}{1-\gamma(p)} < \infty \tag{2.5.4}$$

by multiplicativity of the operator norm.

Remark (Relation to spectral gap). By definition,

$$\gamma(p) = \sup_{f\perp 1} \frac{(pf,pf)_{L^2(\mu)}^{1/2}}{(f,f)_{L^2(\mu)}^{1/2}} = \sup_{f\perp 1} \frac{(f,p^*pf)_{L^2(\mu)}^{1/2}}{(f,f)_{L^2(\mu)}^{1/2}} = \varrho(p^*p|_{L_0^2(\mu)})^{1/2},$$

i.e., $\gamma(p)$ is the **spectral radius** of the linear operator p^*p restricted to $L^2_0(\mu)$. Now suppose that p satisfies the detailed balance condition w.r.t. μ. As remarked above, this is the case for Metropolis-Hastings chains and random scan Gibbs samplers. Then p is a self-adjoint linear operator on the Hilbert space $L^2_0(\mu)$. Therefore,

$$\gamma(p) = \varrho(p^*p|_{L^2_0(\mu)})^{1/2} = \varrho(p|_{L^2_0(\mu)}) = \sup_{f\perp 1} \frac{(f,pf)_{L^2(\mu)}}{(f,f)_{L^2(\mu)}}, \quad \text{and}$$
$$1-\gamma(p) = \inf_{f\perp 1} \frac{(f,f-pf)_{L^2(\mu)}}{(f,f)_{L^2(\mu)}} = \mathrm{Gap}(\mathcal{L}),$$

where the **spectral gap** $\mathrm{Gap}(\mathcal{L})$ of the generator $\mathcal{L} = p - I$ is defined by

$$\mathrm{Gap}(\mathcal{L}) = \inf_{f\perp 1} \frac{(f,-\mathcal{L}f)_{L^2(\mu)}}{(f,f)_{L^2(\mu)}} = \inf \mathrm{spec}(-\mathcal{L}|_{L^2_0(\mu)}).$$

$\mathrm{Gap}(\mathcal{L})$ 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 (2.5.2) and (2.5.3), $2\,\mathrm{Gap}(\mathcal{L}) - 1$ provides upper bound for the asymptotic variances in the symmetric case.

2.5.2 Non-asymptotic bounds for ergodic averages

For deriving non-asymptotic error bounds for estimates by ergodic averages 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) $\mathcal{W}^1_d(\nu p, \widetilde{\nu} p) \le \alpha \mathcal{W}^1_d(\nu, \widetilde{\nu})$ for any $\nu, \widetilde{\nu} \in \mathcal{P}(S)$, and

(A2) $\mathrm{Var}_{p(x,\cdot)}(f) \le \overline{\sigma}^2 \|f\|^2_{\mathrm{Lip}(d)}$ for any $x \in S$ and any Lipschitz continuous function $f : S \to \mathbb{R}$.

Suppose that (X_n, P_x) is a Markov chain with transition kernel p.

Lemma 2.17 (Decay of correlations). *If (A1) and (A2) hold, then the following non-asymptotic bounds hold for any $n, k \in \mathbb{N}$ and any Lipschitz continuous function $f : S \to \mathbb{R}$:*

$$\mathrm{Var}_{P_x}[f(X_n)] \le \sum_{k=0}^{n-1} \alpha^{2k} \overline{\sigma}^2 \|f\|^2_{Lip(d)}, \quad \textit{and} \tag{2.5.5}$$
$$|\mathrm{Cov}_{P_x}[f(X_n), f(X_{n+k})]| \le \frac{\alpha^k}{1-\alpha^2} \overline{\sigma}^2 \|f\|^2_{Lip(d)}. \tag{2.5.6}$$

Proof. The inequality (2.5.5) follows by induction on n. It holds true for $n = 0$, and if (2.5.5) holds for some $n \geq 0$ then

$$\begin{aligned}
\mathrm{Var}_{P_x}[f(X_{n+1})] &= E_x\big[\,\mathrm{Var}_{P_x}[f(X_{n+1})|\mathcal{F}_n^X]\big] + \mathrm{Var}_{P_x}\,\big[E_x[f(X_{n+1})|\mathcal{F}_n^X]\big] \\
&= E_x\big[\,\mathrm{Var}_{p(X_n,\cdot)}(f)\big] + \mathrm{Var}_{P_x}\,\big[(pf)(X_n)\big] \\
&\leq \overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)} + \sum_{k=0}^{n-1}\alpha^{2k}\overline{\sigma}^2\|pf\|^2_{\mathrm{Lip}(d)} \\
&\leq \sum_{k=0}^{n}\alpha^{2k}\overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)}
\end{aligned}$$

by the Markov property and the assumptions (A1) and (A2). Noting that

$$\sum_{k=0}^{n-1}\alpha^{2k} \leq \frac{1}{1-\alpha^2} \qquad \text{for any } n \in \mathbb{N},$$

the bound (2.5.6) for the correlations follows from (2.5.5) since

$$\begin{aligned}
\big|\,\mathrm{Cov}_{P_x}\,[f(X_n), f(X_{n+k})]\,\big| &= \big|\,\mathrm{Cov}_{P_x}\,\big[f(X_n), (p^k f)(X_n)\big]\,\big| \\
&\leq \mathrm{Var}_{P_x}\,[f(X_n)]^{1/2}\ \mathrm{Var}_{P_x}\,\big[(p^k f)(X_n)\big]^{1/2} \\
&\leq \frac{1}{1-\alpha^2}\overline{\sigma}^2\|f\|_{\mathrm{Lip}(d)}\|p^k f\|_{\mathrm{Lip}(d)} \\
&\leq \frac{\alpha^k}{1-\alpha^2}\overline{\sigma}^2\|f\|^2_{\mathrm{Lip}(d)}
\end{aligned}$$

by Assumption (A1). □

As a consequence of Lemma 2.17 we obtain a non-asymptotic upper bound for variances of ergodic averages.

Theorem 2.18 (**Quantitative bounds for bias and variance of ergodic averages of non stationary Markov chains**). *Suppose that (A1) and (A2) hold. Then the following upper bounds hold for any* $b, n \in \mathbb{N}$*, any initial distribution* $\nu \in \mathcal{P}(S)$*, and any Lipschitz continuous function* $f : S \to \mathbb{R}$*:*

$$\big|E_\nu\,[A_{b,n}f] - \mu(f)\big| \leq \frac{1}{n}\frac{\alpha^b}{1-\alpha}\mathcal{W}_d^1(\nu,\mu)\|f\|_{Lip(d)}, \tag{2.5.7}$$

$$\mathrm{Var}_{P_\nu}\,[A_{b,n}f] \leq \frac{1}{n}\|f\|^2_{Lip(d)} \cdot \frac{1}{(1-\alpha)^2}\left(\overline{\sigma}^2 + \frac{\alpha^{2b}}{n}\,\mathrm{Var}(\nu)\right) \tag{2.5.8}$$

where μ *is a stationary distribution for the transition kernel* p*, and*

$$\mathrm{Var}(\nu) := \frac{1}{2}\int\int d(x,y)^2\nu(dx)\nu(dy).$$

Proof. 1) By definition of the averaging operator,

$$\begin{aligned} E_\nu[A_{b,n}f] &= \frac{1}{n}\sum_{i=b}^{b+n-1}(\nu p^i)(f), \qquad \text{and thus} \\ |E_\nu[A_{b,n}f]-\mu(f)| &\le \frac{1}{n}\sum_{i=b}^{b+n-1}|(\nu p^i)(f)-\mu(f)| \\ \le \frac{1}{n}\sum_{i=b}^{b+n-1}\mathcal{W}_d^1(\nu p^i,\mu)\,\|f\|_{\mathrm{Lip}(d)} &\le \frac{1}{n}\sum_{i=b}^{b+n-1}\alpha^i\,\mathcal{W}_d^1(\nu,\mu)\,\|f\|_{\mathrm{Lip}(d)}. \end{aligned}$$

2) By the correlation bound in Lemma 2.17,

$$\begin{aligned} \mathrm{Var}_{P_x}[A_{b,n}f] &= \frac{1}{n^2}\sum_{i,j=b}^{b+n-1}\mathrm{Cov}_{P_x}[f(X_i),f(X_j)] \le \frac{1}{n^2}\sum_{i,j=b}^{b+n-1}\frac{\alpha^{|i-j|}}{1-\alpha^2}\,\overline{\sigma}^2\,\|f\|^2_{\mathrm{Lip}(d)} \\ &\le \frac{1}{n}\,\frac{\overline{\sigma}^2}{(1-\alpha^2)}\left(1+2\sum_{k=1}^{\infty}\alpha^k\right)\|f\|^2_{\mathrm{Lip}(d)} = \frac{1}{n}\,\frac{\overline{\sigma}^2}{(1-\alpha)^2}\,\|f\|^2_{\mathrm{Lip}(d)}. \end{aligned}$$

Therefore, for an arbitrary initial distribution $\nu \in \mathcal{P}(S)$,

$$\begin{aligned} \mathrm{Var}_{P_\nu}[A_{b,n}f] &= E_\nu\left[\mathrm{Var}_{P_\nu}\left[A_{b,n}f|X_0\right]\right] + \mathrm{Var}_{P_\nu}\left[E_\nu\left[A_{b,n}f|X_0\right]\right] \\ &= \int \mathrm{Var}_{P_x}\left[A_{b,n}f\right]\nu(dx) + \mathrm{Var}_\nu\left[\frac{1}{n}\sum_{i=b}^{b+n-1}p^i f\right] \\ &\le \frac{1}{n}\frac{\overline{\sigma}^2}{(1-\alpha)^2}\|f\|^2_{\mathrm{Lip}(d)} + \left(\frac{1}{n}\sum_{i=b}^{b+n-1}\mathrm{Var}_\nu(p^i f)^{1/2}\right)^2. \end{aligned}$$

The assertion now follows since

$$\begin{aligned} \mathrm{Var}_\nu(p^i f) &\le \frac{1}{2}\|p^i f\|^2_{\mathrm{Lip}(d)}\iint d(x,y)^2\nu(dx)\nu(dy) \\ &\le \alpha^{2i}\,\|f\|^2_{\mathrm{Lip}(d)}\,\mathrm{Var}(\nu). \end{aligned}$$

□

Chapter 3

Continuous time Markov processes, generators and martingales

This chapter focuses on the connection between continuous-time Markov processes 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** $(\mathcal{L}_\mathbf{t}, \mathcal{A}), t \in \mathbb{R}_+$**, of linear operators** with domain $\mathcal{A} \subseteq \mathcal{F}_b(S)$ if and only if

$$M_t^{[f]} = f(X_t) - \int_0^t (\mathcal{L}_s f)(X_s)\, ds \tag{3.0.1}$$

is an $(\mathcal{F}_t)$ martingale for any 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

$$(\mathcal{L}_t f)(X_t) = \lim_{h\downarrow 0} E\left[\frac{f(X_{t+h}) - f(X_t)}{h} \middle| \mathcal{F}_t \right] \tag{3.0.2}$$

is the expected rate of change of $f(X_t)$ in the next instant of time given the previous information.

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, cf. Theorem 3.22. 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 3.4.

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 w.r.t. some probability measure. In Section 3.3 we show how to realize transition functions of time-homogeneous Markov processes as strongly continuous contraction semigroups on appropriate Banach space of functions, and we establish the relation between such semigroups and their generators. The connection to martingale problems is made in Section 3.4, and Section 3.5 indicates in a special situation how solutions of martingale problems can be constructed from their generators by exploiting stability of the martingale problem under weak convergence. Before turning to the general setup, Section 3.1 is devoted mainly to an explicit construction of jump processes with finite jump intensities from their jump rates (i.e. from their generators), and the derivation of forward and backward equations and the martingale problem in this more concrete context. Section 3.2 briefly discusses the application of Lyapunov function techniques in continuous time.

3.1 Jump processes and diffusions

3.1.1 Jump processes with finite jump intensity

Let $q_t \colon S \times \mathcal{S} \to [0, \infty]$ be a ***kernel of positive measure***, i.e. $x \mapsto q_t(x, A)$ is measurable and $A \mapsto q_t(x, A)$ is a positive measure.

Aim: Construct a pure jump process with instantaneous jump rates $q_t(x, dy)$, i.e.

$$P_{t,t+h}(x, B) = q_t(x, B) \cdot h + o(h) \quad \forall\, t \geq 0,\ x \in S,\ B \subseteq S \setminus \{x\} \text{ measurable}$$

$(X_t)_{t\geq 0} \leftrightarrow (Y_n, J_n)_{n\geq 0} \leftrightarrow (Y_n, S_n)$ with J_n ***holding times***, S_n ***jumping times*** of X_t.
$J_n = \sum\limits_{i=1}^{n} S_i \in (0,\infty]$ with jump time $\{J_n : n \in \mathbb{N}\}$ point process on $\mathbb{R}^+$, $\zeta = \sup J_n$ ***explosion time***.

Construction of a process with initial distribution $\mu \in \mathbf{M_1(S)}$:

$\lambda_t(x) := q_t(x, S \setminus \{x\})$ intensity, *total rate of jumping away from x*.
Assumption: $\lambda_t(x) < \infty \quad \forall x \in S$, *no instantaneous jumps*.
$\pi_t(x, A) := \frac{q_t(x,A)}{\lambda_t(x)}$ transition probability , *where jumps from x at time t go to*.

a) ***Time-homogeneous case:*** $q_t(x, dy) \equiv q(x, dy)$ independent of t, $\lambda_t(x) \equiv \lambda(x)$, $\pi_t(x, dy) \equiv \pi(x, dy)$.

Y_n $(n = 0, 1, 2, \ldots)$ Markov chain with transition kernel $\pi(x, dy)$ and initial distribution μ

$S_n := \dfrac{E_n}{\lambda(Y_{n-1})}$, $E_n \sim \mathrm{Exp}(1)$ independent and identically distributed random variables, independent of Y_n, i.e.

$$S_n|(Y_0, \ldots Y_{n-1}, E_1, \ldots E_{n-1}) \sim \mathrm{Exp}(\lambda(Y_{n-1})),$$

$$J_n = \sum_{i=1}^{n} S_i$$

$$X_t := \begin{cases} Y_n & \text{for} \quad t \in [J_n, J_{n+1}),\ n \geq 0 \\ \Delta & \text{for} \quad t \geq \zeta = \sup J_n \end{cases}$$

where Δ is an extra point, called the *cemetery*.

Example. 1) **Poisson process with intensity $\lambda > 0$**

$$S = \{0, 1, 2, \ldots\}, \quad q(x, y) = \lambda \cdot \delta_{x+1,y}, \quad \lambda(x) = \lambda\ \forall x, \quad \pi(x, x + 1) = 1$$

$S_i \sim \mathrm{Exp}(\lambda)$ independent and identically distributed random variables, $Y_n = n$

$$N_t = n \ :\Leftrightarrow\ J_n \leq t \leq J_{n+1},$$

$N_t = \#\{i \geq 1 \ :\ J_i \leq t\}$ *counting process of point process* $\{J_n \mid n \in \mathbb{N}\}$.

Distribution at time t:

$$P[N_t \geq n] = P[J_n \leq t] \overset{J_n=\sum_{i=1}^{n} S_i\sim\Gamma(\lambda,n)}{=} \int_0^t \frac{(\lambda s)^{n-1}}{(n-1)!}\lambda e^{-\lambda s}\, ds \overset{\text{differentiate r.h.s.}}{=} e^{-t\lambda}\sum_{k=n}^{\infty}\frac{(t\lambda)^k}{k!},$$

$$N_t \sim \mathrm{Poisson}(\lambda t)$$

2) **Continuization of discrete time chain**

Let $(Y_n)_{n\geq 0}$ be a time-homogeneous Markov chain on S with transition functions $p(x,dy)$,

$$X_t = Y_{N_t}, \quad N_t \text{ Poisson(1)-process independent of } (Y_n),$$
$$q(x,dy) = \pi(x,dy), \quad \lambda(x) = 1$$

e.g. ***compound Poisson process*** (continuous time random walk):

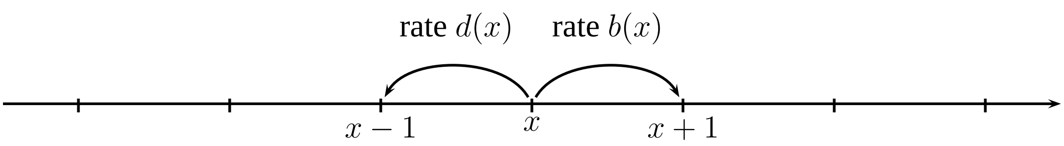
$$X_t = \sum_{i=1}^{N_t} Z_i,$$

$Z_i\colon \Omega \to \mathbb{R}^d$ independent and identically distributed random variables, independent of (N_t).

3) **Birth and death chains**

$S = \{0,1,2,\ldots\}$.

$$q(x,y) = \begin{cases} b(x) & \text{if } y = x+1 \quad \text{"birth rate"} \\ d(x) & \text{if } y = x-1 \quad \text{"death rate"} \\ 0 & \text{if } |y-x| \geq 2 \end{cases}$$



b) ***Time-inhomogeneous case:***

Remark (Survival times)**.** Suppose an event occurs in time interval $[t, t+h]$ with probability $\lambda_t \cdot h + o(h)$ provided it has not occurred before:

$$\begin{aligned}
& P[T \leq t+h | T > t] = \lambda_t \cdot h + o(h) \\
\Leftrightarrow \quad & \underbrace{\frac{P[T > t+h]}{P[T > t]}}_{\text{survival rate}} = P[T > t+h | T > t] = 1 - \lambda_t h + o(h) \\
\Leftrightarrow \quad & \frac{\log P[T > t+h] - \log P[T > t]}{h} = -\lambda_t + o(h) \\
\Leftrightarrow \quad & \frac{d}{dt} \log P[T > t] = -\lambda_t \\
\Leftrightarrow \quad & P[T > t] = \exp\left(- \int\limits_0^t \lambda_s \, ds \; t \right)
\end{aligned}$$

where the integral is the *accumulated hazard rate up to time* t,

$$f_T(t) = \lambda_t \exp\left(- \int\limits_0^t \lambda_s \, ds \right) \cdot I_{(0,\infty)}(t) \quad \text{the } \textit{survival distribution with hazard rate } \lambda_s$$

Simulation of T:

$E \sim \mathrm{Exp}(1), \; T := \inf\{t \geq 0 \; : \; \int\limits_0^t \lambda_s \, ds \geq E\}$

$$\Rightarrow \quad P[T > t] = P\left[\int\limits_0^t \lambda_s \, ds < E \right] = e^{-\int\limits_0^t \lambda_s \, ds}$$

Construction of time-inhomogeneous jump process:

Fix $t_0 \geq 0$ and $\mu \in M_1(S)$ (the initial distribution at time t_0).

Suppose that with respect to $P_{(t_0,\mu)}$,

$$J_0 := t_0, \qquad Y \sim \mu$$

and

$$P_{(t_0,\mu)}[J_1 > t \mid Y_0] = e^{-\int\limits_{t_0}^{t \vee t_0} \lambda_s(Y_0) \, ds}$$

for all $t \geq t_0$, and $(Y_{n-1}, J_n)_{n \in \mathbb{N}}$ is a *time-homogeneous* Markov chain on $S \times [0, \infty)$ with transition law

$$P_{(t_0,\mu)}[Y_n \in dy, \; J_{n+1} > t \mid Y_0, J_1, \ldots, Y_{n-1}, J_n] = \pi_{J_n}(Y_{n-1}, dy) \cdot e^{-\int\limits_{J_n}^{t \vee J_n} \lambda_s(y) \, ds}$$

i.e.

$$P_{(t_0,\mu)}[Y_n \in A,\ J_{n+1} > t \,|Y_0, J_1, \ldots, Y_{n-1}, J_n] = \int_A \pi_{J_n}(Y_{n-1}, dy) \cdot e^{-\int\limits_{J_n}^{t \vee J_n} \lambda_s(y)\, ds}$$

P-a.s. for all $A \in \mathcal{S},\ t \geq 0$.

Explicit (algorithmic) construction:

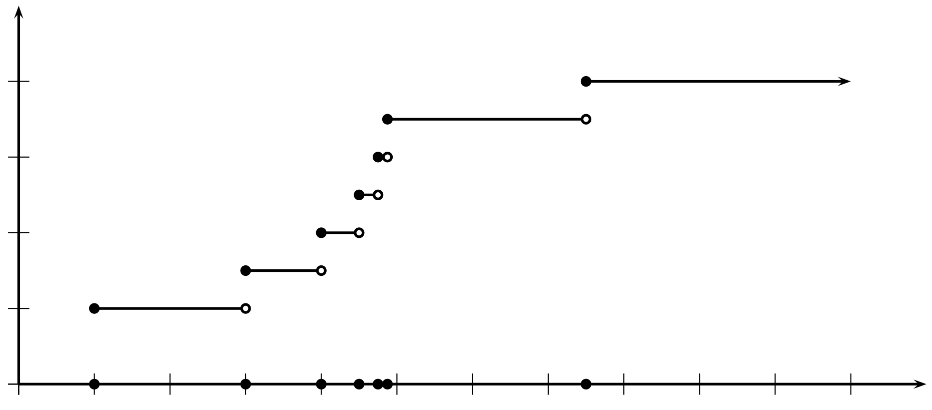
- $J_0 := t_0,\ Y_0 \sim \mu$

For $n = 1, 2, \ldots$ do

- $E_n \sim \mathrm{Exp}(1)$ independent of $Y_0, \ldots, Y_{n-1}, E_1, \ldots, E_{n-1}$
- $J_n := \inf \left\{ t \geq 0 \ :\ \int_{J_{n-1}}^t \lambda_s(Y_{n-1})\, ds \geq E_n \right\}$
- $Y_n | (Y_0, \ldots, Y_{n-1}, E_0, \ldots, E_n) \sim \pi_{J_n}(Y_{n-1}, \cdot)$ where $\pi_\infty(x, \cdot) = \delta_x$ (or other arbitrary definition)

Example (**Non-homogeneous Poisson process on** $\mathbb{R}_+$**).**

$S = \{0, 1, 2, \ldots\}$,	$q_t(x, y) = \lambda_t \cdot \delta_{x+1,y}$,
$Y_n = n$,	$J_{n+1} \| J_n \sim \lambda_t \cdot e^{-\int_{J_n}^t \lambda_s\, ds}\, dt$,
$N_t := \#\{n \geq 1 \ :\ J_n \leq t\}$	the *associated counting process*



high intensity *low intensity*

Claim:

(1). $f_{J_n}(t) = \frac{1}{(n-1)!} \left(\int_0^t \lambda_s\, ds \right)^{n-1} \lambda_t e^{-\int_0^t \lambda_s\, ds}$

(2). $N_t \sim \text{Poisson}\left(\int_0^t \lambda_s\, ds\right)$

Proof. (1). By induction:

$$\begin{aligned} f_{J_{n+1}}(t) &= \int\limits_0^t f_{J_{n-1}|J_n}(t|r) f_{J_n}(r)\, dr \\ &= \int\limits_0^t \lambda_t e^{-\int_r^t \lambda_s\, ds} \frac{1}{(n-1)!} \left(\int\limits_0^r \lambda_s\, ds \right)^{n-1} \lambda_r e^{-\int_0^r \lambda_s\, ds}\, dr \;= \dots \end{aligned}$$

(2). $P[N_t \geq n] = P[J_n \leq t]$

$\square$

Remark. In general, (Y_n) is not a Markov chain. However:

(1). (Y_{n-1}, J_n) is a Markov chain with respect to $\mathcal{G}_n = \sigma(Y_0, \dots, Y_{n-1}, E_1, \dots, E_n)$ with transition functions

$$p\left((x,s), dydt\right) = \pi_s(x, dy) \lambda_t(y) \cdot e^{-\int_s^t \lambda_r(y)\, dr} I_{(s,\infty)}(t)\, dt$$

(2). (J_n, Y_n) is a Markov chain with respect to $\widetilde{\mathcal{G}} = \sigma(Y_0, \dots, Y_n, E_1, \dots, E_n)$ with transition functions

$$\widetilde{p}\left((x,s), dtdy\right) = \lambda_t(x) \cdot e^{-\int_s^t \lambda_r(x)\, dr} I_{(s,\infty)}(t) \pi_t(x, dy)$$

Remark. (1). J_n strictly increasing.

(2). $J_n = \infty \; \forall\, n, m$ is possible $\rightsquigarrow$ X_t absorbed in state Y_{n-1}.

(3). $\sup J_n < \infty$ $\rightsquigarrow$ explosion in finite time

(4). $\{s < \zeta\} = \{X_s \neq \Delta\} \in \mathcal{F}_s \rightsquigarrow$ no explosion before time s.

3.1.2 Markov property

$K_s := \min\{n \,:\, J_n > s\}$ first jump after time s. Stopping time with respect to $\mathcal{G}_n = \sigma\left(E_1, \dots, E_n, Y_0, \dots, Y_{n-1}\right)$,

$$\{K_s < \infty\} = \{s < \zeta\}$$

Lemma (Memoryless property). *Let $s \geq t_0$. Then for all $t \geq s$,*

$$P_{(t_0,\mu)}\left[\{J_{K_s} > t\} \cap \{s < \zeta\} \mid \mathcal{F}_s\right] = e^{-\int_s^t \lambda_r(X_s)\,dr} \quad P\text{-a.s. on } \{s < \zeta\}$$

i.e.

$$P_{(t_0,\mu)}\left[\{J_{K_s} > t\} \cap \{s < \zeta\} \cap A\right] = E_{(t_0,\mu)}\left[e^{-\int_s^t \lambda_r(X_s)\,dr}\,;\ A \cap \{s < \zeta\}\right] \quad \forall A \in \mathcal{F}_s$$

Remark. The assertion is a restricted form of the Markov property in continuous time: The conditional distribution with respect to $P_{(t_0,\mu)}$ of J_{K_s} given $\mathcal{F}_s$ coincides with the distribution of J_1 with respect to $P_{(s,X_s)}$.

Proof.

$$A \in \mathcal{F}_s \overset{\text{(Ex.)}}{\Rightarrow} A \cap \{K_s = n\} \in \sigma\left(J_0, Y_0, \ldots, J_{n-1}, Y_{n-1}\right) = \widetilde{\mathcal{G}}_{n-1}$$

$$\Rightarrow \quad P\left[\{J_{K_s} > t\} \cap A \cap \{K_s = n\}\right] = E\left[P[J_n > t \mid \widetilde{\mathcal{G}}_{n-1}]\,;\ A \cap \{K_s = n\}\right]$$

where

$$P[J_n > t \mid \widetilde{\mathcal{G}}_{n-1}] = \exp\left(-\int_{J_{n-1}}^{t} \lambda_r(Y_{n-1})\,dr\right) = \exp\left(-\int_s^t \lambda_r(\underbrace{Y_{n-1}}_{=X_s})\,dr\right) \cdot P[J_n > s \mid \widetilde{\mathcal{G}}_{n-1}],$$

hence we get

$$P[J_n > t \mid \widetilde{\mathcal{G}}_{n-1}] = E\left[e^{-\int_s^t \lambda_r(X_s)\,dr}\,;\ A \cap \{K_s = n\} \cap \{J_n > s\}\right] \quad \forall\, n \in \mathbb{N}$$

where $A \cap \{K_s = n\} \cap \{J_n > s\} = A \cap \{K_s = n\}$.
Summing over n gives the assertion since

$$\{s < \zeta\} = \dot{\bigcup_{n \in \mathbb{N}}} \{K_s = n\}.$$

□

For $y_n \in S$, $t_n \in [0, \infty]$ strictly increasing define

$$x := \Phi\left((t_n, y_n)_{n=0,1,2,\ldots}\right) \in \mathrm{PC}\left([t_0, \infty),\ S \dot{\cup} \{\Delta\}\right)$$

by

$$x_t := \begin{cases} Y_n & \text{for} \quad t_n \leq t < t_{n+1}\ ,\ n \geq 0 \\ \Delta & \text{for} \quad t \geq \sup t_n \end{cases}$$

Let

$$\begin{aligned}(X_t)_{t\geq t_0} &:= \Phi\left((J_n, Y_n)_{n\geq 0}\right)\\ \mathcal{F}_t^X &:= \sigma\left(X_s \mid s\in[t_0,t]\right), \quad t\geq t_0\end{aligned}$$

Theorem 3.1 (Markov property). *Let $s \geq t_0$, $X_{s:\infty} := (X_t)_{t\geq s}$. Then*

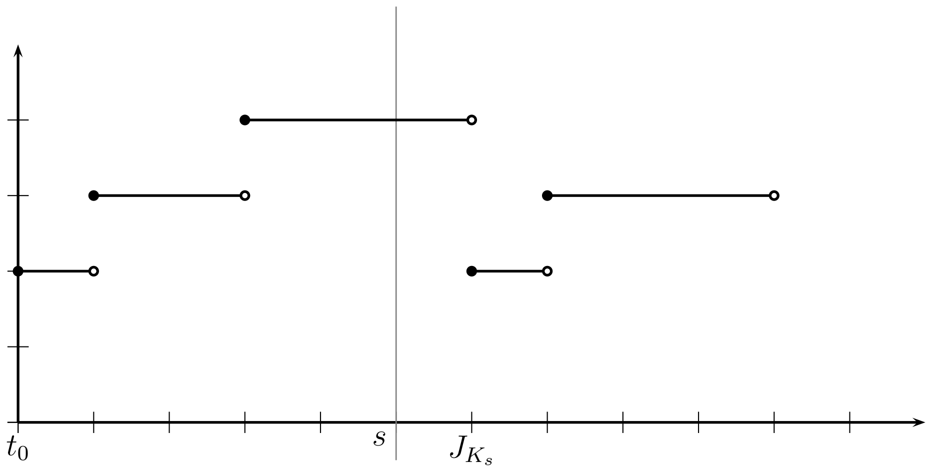
$$E_{(t_0,\mu)}\left[F(X_{s:\infty})\cdot I_{\{s<\zeta\}} \mid \mathcal{F}_s^X\right](\omega) = E_{(s,X_s(\omega))}\left[F(X_{s:\infty})\right] \quad P\text{-a.s. } \{s<\zeta\}$$

for all

$$F\colon \ \mathrm{PC}\left([s,\infty),\ S\cup\{\Delta\}\right)\to\mathbb{R}^+$$

measurable with respect to $\sigma\left(x\mapsto x_t \mid t\geq s\right)$.

Proof. $X_{s:\infty} = \Phi(s, Y_{K_s-1}, J_{K_s}, Y_{K_s}, J_{K_s+1},\ldots)$ on $\{s<\zeta\} = \{K_s<\infty\}$



i.e. the process after time s is constructed in the same way from $s, Y_{K_s-1}, J_{K_s},\ldots$ as the original process is constructed from $t_0, Y_0, J_1,\ldots$. By the Strong Markov property for the chain (Y_{n-1}, J_n),

$$\begin{aligned}&E_{(t_0,\mu)}\left[F(X_{s:\infty})\cdot I_{\{s<\zeta\}} \mid \mathcal{G}_{K_s}\right]\\ =&E_{(t_0,\mu)}\left[F\circ\Phi(s, Y_{K_s-1}, J_{K_s},\ldots)\cdot I_{\{K_s<\infty\}} \mid \mathcal{G}_{K_s}\right]\\ =&E^{\text{Markov chain}}_{(Y_{K_s-1},J_{K_s})}\left[F\circ\Phi(s,(Y_0,J_1),(Y_1,J_2),\ldots)\right] \quad \text{a.s. on } \{K_s<\infty\}=\{s<\zeta\}.\end{aligned}$$

Since $\mathcal{F}_s \subseteq \mathcal{G}_{K_s}$, we obtain by the projectivity of the conditional expectation,

$$\begin{aligned}
&E_{(t_0,\mu)}\left[F(X_{s:\infty})\cdot I_{\{s<\zeta\}} \mid \mathcal{F}_s\right] \\
=&E_{(t_0,\mu)}\left[E^{\text{Markov chain}}_{(X_s,J_{K_s})}\left[F\circ\Phi(s,(Y_0,J_1),\dots)\cdot I_{\{s<\zeta\}}\mid \mathcal{F}_s\right]\right]
\end{aligned}$$

taking into account that the conditional distribution given $\mathcal{G}_{K_s}$ is 0 on $\{s\geq\zeta\}$ and that $Y_{K_s-1}=X_s$.

Here the conditional distribution of J_{K_s} ist $k(X_s,\cdot)$, by Lemma 3.1.2

$$k(x,dt)=\lambda_t(x)\cdot e^{-\int_s^t \lambda_r(x)\,dr}\cdot I_{(s,\infty)}(t)\,dt$$

hence

$$E_{(t_0,\mu)}\left[F(X_{s:\infty})\cdot I_{\{s<\zeta\}} \mid \mathcal{F}_s\right]=E^{\text{Markov chain}}_{(X_s,k(X_s,\cdot))}\left[F\circ\Phi(\dots)\right] \quad \text{a.s. on } \{s<\zeta\}$$

Here $k(X_s,\cdot)$ is the distribution of J_1 with respect to P_{s,X_s}, hence we obtain

$$\begin{aligned}
E_{(t_0,\mu)}\left[F(X_{s:\infty})\cdot I_{\{s<\zeta\}} \mid \mathcal{F}_s\right]&=E_{(s,X_s)}\left[F(\Phi(s,Y_0,J_1,\dots))\right]\\
&=E_{(s,X_s)}\left[F(X_{s:\infty})\right]
\end{aligned}$$

□

Example (Non-homogeneous Poisson process). A non-homogeneous Poisson process $(N_t)_{t\geq 0}$ with intensity λ_t has independent increments with distribution

$$N_t-N_s\sim \text{Poisson}\left(\int_s^t \lambda_r\,dr\right)$$

Proof:

$$\begin{aligned}
&P\left[N_t-N_s\geq k \mid \mathcal{F}_s^N\right]\overset{\text{MP}}{=}P_{(s,N_s)}\left[N_t-N_s\geq k\right]=P_{(s,N_s)}\left[J_k\leq t\right]\\
&\overset{\text{as above}}{=}\text{Poisson}\left(\int_s^t \lambda_r\,dr\right)\ (\{k,k+1,\dots\})\,.
\end{aligned}$$

Hence N_t-N_s independent of $\mathcal{F}_s^N$ and $\text{Poisson}\left(\int_s^t\lambda_r\,dr\right)$ distributed. □

3.1.3 Generator and backward equation

Definition. *The* ***infinitesimal generator*** *(or* ***intensity matrix***, ***kernel****) of a Markov jump process at time t is defined by*

$$\mathcal{L}_t(x,dy) = q_t(x,dy) - \lambda_t(x)\delta_x(dy)$$

i.e.

$$\begin{aligned}(\mathcal{L}_t f)(x) &= (q_t f)(x) - \lambda_t(x) f(x) \\ &= \int q_t(x,dy)\cdot(f(y)-f(x))\end{aligned}$$

for all bounded and measurable $f\colon S\to\mathbb{R}$*.*

Remark. (1). $\mathcal{L}_t$ is a linear operator on functions $f\colon S\to\mathbb{R}$.

(2). If S is discrete, $\mathcal{L}_t$ is a matrix, $\mathcal{L}_t(x,y) = q_t(x,y) - \lambda_t(x)\delta(x,y)$. This matrix is called ***Q-Matrix***.

Theorem 3.2 (Integrated backward equation). *(1). Under* $P_{(t_0,\mu)}$, $(X_t)_{t\geq t_0}$ *is a Markov jump process with initial distribution* $X_{t_0}\sim\mu$ *and transition probabilities*

$$p_{s,t}\left(x,B\right) = P_{(s,x)}\left[X_t\in B\right] \quad (0\leq s\leq t,\ x\in S,\ B\in\mathcal{S})$$

satisfying the Chapman-Kolmogorov equations $p_{s,t}p_{t,u} = p_{s,u} \quad \forall\, 0\leq s\leq t\leq u$*.*

(2). The ***integrated backward equation***

$$p_{s,t}(x,B) = e^{-\int_s^t \lambda_r(x)\,dr}\delta_x(B) + \int_s^t e^{-\int_s^r \lambda_u(x)\,du}(q_r p_{r,t})(x,B)\,dr \tag{3.1.1}$$

holds for all $0\leq s\leq t$, $x\in S$ *and* $B\in\mathcal{S}$*.*

(3). If $t\mapsto\lambda_t(x)$ *is continuous for all* $x\in S$*, then*

$$(p_{s,s+h}f)(x) = (1-\lambda_s(x)\cdot h)f(x) + h\cdot(q_s f)(x) + o(h) \tag{3.1.2}$$

holds for all $s\geq 0$, $x\in S$ *and bounded functions* $f\colon\ \to\mathbb{R}$ *such that* $t\mapsto(q_t f)(x)$ *is continuous.*

Remark. (1). (3.1.2) shows that (X_t) is the continuous time Markov chain with intensities $\lambda_t(x)$ and transition rates $q_t(x,dy)$.

(2). If $\zeta = \sup J_n$ is finite with strictly positive probability, then there are other possible continuations of X_t after the explosion time ζ.

$\rightsquigarrow$ non-uniqueness.

The constructed process is called the ***minimal chain*** for the given jump rates, since its transition probabilities $p_t(x,B)$, $B \in \mathcal{S}$ are minimal for all continuations, cf. below.

(3). The integrated backward equation extends to bounded functions $f\colon S \to \mathbb{R}$

$$(p_{s,t}f)(x) = e^{-\int_s^t \lambda_r(x)\,dr} f(x) + \int\limits_s^t e^{-\int_s^r \lambda_u(x)\,du} (q_r p_{r,t} f)(x)\,dr \tag{3.1.3}$$

Proof. (1). By the Markov property,

$$P_{(t_0,\mu)}\left[X_t \in B | \mathcal{F}_s^X\right] = P_{(s,X_s)}\left[X_t \in B\right] = p_{s,t}(X_s, B) \quad \text{a.s.}$$

since $\{X_t \in B\} \subseteq \{t < \zeta\} \subseteq \{s < \zeta\}$ for all $B \in \mathcal{S}$ and $0 \le s \le t$.

Thus $\left((X_t)_{t\ge t_0}\ ,\ P_{(t_0,\mu)}\right)$ is a Markov jump process with transition kernels $p_{s,t}$. Since this holds for any initial condition, the Chapman-Kolmogorov equations

$$(p_{s,t}p_{t,u}f)(x) = (p_{s,u}f)(x)$$

are satisfied for all $x \in S$, $0 \le s \le t \le u$ and $f\colon S \to \mathbb{R}$.

(2). ***First step analysis***: Condition on $\widetilde{G_1} = \sigma(J_0, Y_0, J_1, Y_1)$:

Since $X_t = \Phi_t(J_0, Y_0, J_1, Y_1, J_2, Y_2, \ldots)$, the Markov property of (J_n, Y_n) implies

$$P_{(s,x)}\left[X_t \in B | \widetilde{G_1}\right](\omega) = P_{(J_1(\omega), Y_1(\omega))}\left[\Phi_t(s, x, J_0, Y_0, J_1, Y_1, \ldots) \in B\right]$$

On the right side, we see that

$$\Phi_t(s, x, J_0, Y_0, J_1, Y_1, \ldots) = \begin{cases} x & \text{if } t < J_1(\omega) \\ \Phi_t(J_0, Y_0, J_1, Y_1, \ldots) & \text{if } t \ge J_1(\omega) \end{cases}$$

and hence

$$P_{(s,x)}\left[X_t \in B | \widetilde{G_1}\right](\omega) = \delta_x(B) \cdot I_{\{t<J_1\}}(\omega) + P_{(J_1(\omega),Y_1(\omega))}[X_t \in B] \cdot I_{\{t \ge J_1\}}(\omega)$$

$P_{(s,x)}$-a.s. We conclude

$$\begin{aligned} p_{s,t}(x,B) &= P_{(s,x)}[X_t \in B] \\ &= \delta_x(B) P_{(s,x)}[J_1 > t] + E_{(s,x)}[p_{J_1,t}(Y_1,B); t \geq J_1] \\ &= \delta_x(B) \cdot e^{-\int_s^t \lambda_r(x)\,dr} + \int_s^t \lambda_r(x) e^{-\int_s^r \lambda_u(x)\,du} \int \underbrace{\pi_r(x,dy) p_{r,t}(y,B)}_{=(\pi_r p_{r,t})(x,B)} \,dr \\ &= \delta_x(B) \cdot e^{-\int_s^t \lambda_r(x)\,dr} + \int_s^t e^{-\int_s^r \lambda_u(x)\,du} (q_r p_{r,t})(x,B)\,dr \end{aligned}$$

(3). This is a direct consequence of (3.1.1).

Fix a bounded function $f\colon S \to \mathbb{R}$. Note that

$$0 \leq (q_r p_{r,t} f)(x) = \lambda_r(x)(\pi_r p_{r,t} f)(x) \leq \lambda_r(x) \sup|f|$$

for all $0 \leq r \leq t$ and $x \in S$. Hence if $r \mapsto \lambda_r(x)$ is continuous (and locally bounded) for all $x \in S$, then

$$(p_{r,t} f)(x) \longrightarrow f(x) \tag{3.1.4}$$

as $r, t \downarrow s$ for all $x \in S$.

Thus by dominated convergence,

$$\begin{aligned} &(q_r p_{r,t} f)(x) - (q_s f)(x) \\ &= \int q_r(x,dy)(p_{r,t} f(y) - f(y)) + (q_r f)(x) - (q_s f)(x) \longrightarrow 0 \end{aligned}$$

as $r, t \downarrow s$ provided $r \mapsto (q_r f)(x)$ is continuous. The assertion now follows from (3.1.3).

□

Exercise (A first non-explosion criterion). Show that if $\bar{\lambda} := \sup_{\substack{t \geq 0 \\ x \in S}} \lambda_t(x) < \infty$, then

$$\zeta = \infty \quad P_{(t_0,\mu)}\text{-a.s. } \forall\, t_0,\ \mu$$

Remark. In the time-homogeneous case,

$$J_n = \sum_{k=1}^n \frac{E_k}{\lambda(Y_{n-1})}$$

is a sum of conditionally independent exponentially distributed random variables given $\{Y_k \mid k \geq 0\}$. From this one can conclude that the events

$$\{\zeta < \infty\} = \left\{ \sum_{k=1}^\infty \frac{E_k}{\lambda(Y_{k-1})} < \infty \right\} \text{ and } \left\{ \sum_{k=0}^\infty \frac{1}{\lambda(Y_k)} < \infty \right\}$$

coincide almost surely (apply Kolmogorov's 3-series Theorem).

3.1.4 Forward equation and martingale problem

Theorem 3.3 (Kolmogorov's forward equation). *Suppose that*

$$\bar{\lambda}_t = \sup_{0 \le s \le t} \sup_{x \in S} \lambda_s(x) < \infty$$

*for all $t > 0$. Then the **forward equation***

$$\frac{d}{dt}(p_{s,t}f)(x) = (p_{s,t}\mathcal{L}_t f)(x), \qquad (p_{s,s}f)(x) = f(x) \tag{3.1.5}$$

holds for all $0 \le s \le t,\ x \in S$ and all bounded functions $f\colon S \to \mathbb{R}$ such that $t \mapsto (q_t f)(x)$ and $t \mapsto \lambda_t(x)$ are continuous for all x.

Proof. (1). ***Strong continuity:*** Fix $t_0 > 0$. Note that $\|q_r f\|_{\sup} \le \bar{\lambda}_r \|f\|_{\sup}$ for all $0 \le r \le t_0$. Hence by the assumption and the integrated backward equation (3.1.3),

$$\begin{aligned} \|p_{s,t}f - p_{s,r}f\|_{\sup} &= \|p_{s,r}(p_{r,t}f - f)\|_{\sup} \\ &\le \|p_{r,t}f - f\|_{\sup} \le \varepsilon(t-r) \cdot \|f\|_{\sup} \end{aligned}$$

for all $0 \le s \le r \le t \le t_0$ and some function $\varepsilon\colon \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{h \downarrow 0} \varepsilon(h) = 0$.

(2). ***Differentiability:*** By 1.) and the assumption,

$$(r, u, x) \mapsto (q_r p_{r,u} f)(x)$$

is uniformly bounded for $0 \le r \le u \le t_0$ and $x \in S$, and

$$q_r p_{r,u} f = \underbrace{q_r(p_{r,u}f - f)}_{\longrightarrow 0 \text{ uniformly}} + q_r f \longrightarrow q_t f$$

pointwise as $r, u \longrightarrow t$. Hence by the integrated backward equation (3.1.3) and the continuity of $t \mapsto \lambda_t(x)$,

$$\frac{p_{t,t+h}f(x) - f(x)}{h} \xrightarrow{h \downarrow 0} -\lambda_t(x) f(x) + q_t f(x) = \mathcal{L}_t f(x)$$

for all $x \in S$, and the difference quotients are uniformly bounded.

Dominated convergence now implies

$$\frac{p_{s,t+h}f - p_{s,t}f}{h} = p_{s,t}\frac{p_{t,t+h}f - f}{h} \longrightarrow p_{s,t}\mathcal{L}_t f$$

pointwise as $h \downarrow 0$. A similar argument shows that also

$$\frac{p_{s,t}f - p_{s,t-h}f}{h} = p_{s,t-h}\frac{p_{t-h,t}f - f}{h} \longrightarrow p_{s,t}\mathcal{L}_t f$$

pointwise.

$\square$

Remark. (1). The assumption implies that the operators $\mathcal{L}_s, 0 \le s \le t_0$, are uniformly bounded with respect to the supremum norm:

$$\|\mathcal{L}_s f\|_{\sup} \le \lambda_t \cdot \|f\|_{\sup} \quad \forall\, 0 \le s \le t.$$

(2). Integrating (**??**) yields

$$p_{s,t}f = f + \int_s^t p_{s,r}\mathcal{L}_r f\, dr \tag{3.1.6}$$

In particular, the difference quotients $\frac{p_{s,t+h}f - p_{s,t}f}{h}$ converge uniformly for f as in the assertion.

Notation:

$$< \mu, f >:= \mu(f) = \int f\, d\mu$$

$$\mu \in M_1(S),\ s \ge 0,\ \mu_t := \mu p_{s,t} = P_{(s,\mu)} \circ X_t^{-1} \quad \textit{mass distribution at time } t$$

Corollary (Fokker-Planck equation). *Under the assumptions in the theorem,*

$$\frac{d}{dt} < \mu_t, f >=< \mu_t, \mathcal{L}_t f >$$

for all $t \ge s$ and bounded functions $f\colon S \to \mathbb{R}$ such that $t \mapsto q_t f$ and $t \mapsto \lambda_t$ are pointwise continuous. Abusing notation, one sometimes writes

$$\frac{d}{dt}\mu_t = \mathcal{L}_t^* \mu_t$$

Proof.

$$< \mu_t, f >=< \mu p_{s,t}, f >= \int \mu(dx) \int p_{s,t}(x, dy) f(y) =< \mu, p_{s,t} f >$$

hence we get

$$\frac{< \mu_{t+h}, f > - < \mu_t, f >}{h} =< \mu p_{s,t}, \frac{p_{t,t+h}f - f}{h} >\longrightarrow< \mu_t, \mathcal{L}_t f >$$

as $h \downarrow 0$ by dominated convergence. $\square$

Remark. **(Important!)**

$$\begin{aligned} & P_{(s,\mu)}[\zeta < \infty] > 0 \\ \Rightarrow & < \mu_t, 1 >= \mu_t(S) < 1 \quad \text{for large } t \end{aligned}$$

hence the Fokker-Planck equation does *not* hold for $f \equiv 1$:

$$\underbrace{< \mu_t, 1 >}_{<1} < \underbrace{< \mu, 1 >}_{=1} + \int_0^t < \mu_s, \mathcal{L}_s 1 > \, ds$$

where $\mathcal{L}_s 1 = 0$.

Example. Birth process on $S = \{0, 1, 2, \ldots\}$

$$\begin{aligned} q(i,j) &= \begin{cases} b(i) & \text{if } j = i+1 \\ 0 & \text{else} \end{cases} \\ \pi(i,j) &= \delta_{i+1,j}, \\ Y_n &= n, \\ S_n &= J_n - J_{n-1} \sim \operatorname{Exp}(b(n-1)) \quad \text{independent,} \\ \zeta &= \sup J_n = \sum_{n=1}^{\infty} S_n < \infty \iff \sum_{n=1}^{\infty} b(n)^{n-1} < \infty \end{aligned}$$

In this case, Fokker-Planck does not hold.

The question whether one can extend the forward equation to unbounded jump rates leads to the *martingale problem*.

Definition. *A Markov process $(X_t, P_{(s,x)} \mid 0 \leq s \leq t, x \in S)$ is called* ***non-explosive*** *(or* ***conservative****) if and only if $\zeta = \infty \quad P_{(s,x)}$-a.s. for all s, x.*

Now we consider again the minimal jump process $(X_t, P_{(t_0,\mu)})$ constructed above. A function

$$\begin{aligned} f \colon [0,\infty) \times S &\to \mathbb{R} \\ (t,x) &\mapsto f_t(x) \end{aligned}$$

is called ***locally bounded*** if and only if there exists an increasing sequence of open subsets $B_n \subseteq S$ such that $S = \bigcup B_n$, and

$$\sup_{\substack{x \in B_n \\ 0 \leq s \leq t}} |f_s(x)| < \infty$$

for all $t > 0,\ n \in \mathbb{N}$.

The following theorem gives a probabilistic form of Kolmogorov's forward equation:

Theorem 3.4 (Time-dependent martingale problem). *Suppose that $t \mapsto \lambda_t(x)$ is continuous for all x. Then:*

(1). The process

$$M_t^f := f_t(X_t) - \int_{t_0}^{t} \left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r(X_r)\, dr, \quad t \geq t_0$$

is a local $(\mathcal{F}_t^X)$-martingale up to ζ with respect to $P_{(t_0,\mu)}$ for any locally bounded function $f : \mathbb{R}^+ \times S \to \mathbb{R}$ such that $t \mapsto f_t(x)$ is C^1 for all x, $(t,x) \mapsto \frac{\partial}{\partial t} f_t(x)$ is locally bounded, and $r \mapsto (q_{r,t} f_t)(x)$ is continuous at $r = t$ for all t, x.

(2). If $\bar{\lambda}_t < \infty$ and f and $\frac{\partial}{\partial t} f$ are bounded functions, then M^f is a global martingale.

(3). More generally, if the process is non-explosive then M^f is a global martingale provided

$$\sup_{\substack{x \in S \\ t_0 \leq s \leq t}} \left(|f_s(x)| + \left| \frac{\partial}{\partial s} f_s(x) \right| + |(\mathcal{L}_s f_s)(x)| \right) < \infty \tag{3.1.7}$$

for all $t > t_0$.

Corollary. *If the process is conservative then the forward equation*

$$p_{s,t} f_t = f_s + \int_s^t p_{r,t} \left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r\, dr, \quad t_0 \leq s \leq t \tag{3.1.8}$$

holds for functions f satisfying (3.1.7).

Proof of corollary. M^f being a martingale, we have

$$\begin{aligned}(p_{s,t} f_r)(x) = E_{(s,x)}[f_t(X_t)] &= E_{(s,x)}\left[f_s(X_s) + \int_s^t \left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r(X_r)\, dr \right] \\ &= f_s(x) + \int_s^t p_{s,r} \left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r(x)\, dr\end{aligned}$$

for all $x \in S$. □

Remark. The theorem yields the Doob-Meyer decomposition

$$f_t(X_t) = \text{local martingale} + \text{bounded variation process}$$

Remark. (1). Time-homogeneous case:

If h is an harmonic function, i.e. $\mathcal{L}h = 0$, then $h(X_t)$ is a martingale

(2). In general:

If h_t is *space-time harmonic*, i.e. $\frac{\partial}{\partial t}h_t + \mathcal{L}_t h_t = 0$, then $h(X_t)$ is a martingale. In particular, $(p_{s,t}f)(X_t),\ (t \geq s)$ is a martingale for all bounded functions f.

(3). If h_t is superharmonic (or *excessive*), i.e. $\frac{\partial}{\partial t}h_t + \mathcal{L}_t h_t \leq 0$, then $h_t(X_t)$ is a supermartingale. In particular, $E[h_t(X_t)]$ is decreasing

$\rightsquigarrow$ stochastic Lyapunov function, stability criteria

e.g.

$$h_t(x) = e^{-tc}h(tc), \qquad \mathcal{L}_t h \leq ch$$

Proof of theorem. 2. Similarly to the derivation of the forward equation, one shows that the assumption implies

$$\frac{\partial}{\partial t}(p_{s,t}f_t)(x) = (p_{s,t}\mathcal{L}_t f_t)(x) + \left(p_{s,t}\frac{\partial}{\partial t}f_t\right)(x) \qquad \forall\, x \in S,$$

or, in a integrated form,

$$p_{s,t}f_t = f_s + \int_s^t p_{s,r}\left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r\, dr$$

for all $0 \leq s \leq t$. Hence by the Markov property, for $t_0 \leq s \leq t$,

$$\begin{aligned}
&E_{(t_0,\mu)}[f_t(X_t) - f_s(X_s) \mid \mathcal{F}_s^X] \\
=&E_{(s,X_s)}[f_t(X_t) - f_s(X_s)] = (p_{s,t}f_t)(X_s) - f_s(X_s) \\
=&\int_s^t \left(p_{s,r}\left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r\right)(X_s)\, dr \\
=&E_{(t_0,\mu)}\left[\int_s^t \left(\frac{\partial}{\partial r} + \mathcal{L}_r\right) f_r(X_r)\, dr \mid \mathcal{F}_r^X\right],
\end{aligned}$$

because all the integrands are uniformly bounded.

1. For $k \in \mathbb{N}$ let

$$q_t^{(k)}(x, B) := (\lambda_t(x) \wedge k) \cdot \pi_t(x, B)$$

denote the jump rates for the process $X_t^{(k)}$ with the same transition probabilities as X_t and jump rates cut off at k. By the construction above, the process $X_t^{(k)},\ k \in \mathbb{N}$, and X_t can be realized on the same probability space in such a way that

$$X_t^{(k)} = X_t \quad \text{a.s. on } \{t < T_k\}$$

where

$$T_k := \inf \{t \geq 0 \ : \ \lambda_t(X_t) \geq k,\ X_t \notin B_k\}$$

for an increasing sequence B_k of open subsets of S such that f and $\frac{\partial}{\partial t} f$ are bounded on $[0,t] \times B_k$ for all t, k and $S = \bigcup B_k$. Since $t \mapsto \lambda_t(X_t)$ is piecewise continuous and the jump rates do not accumulate before ζ, the function is locally bounded on $[0, \zeta)$. Hence

$$T_k \nearrow \zeta \quad \text{a.s. as } k \to \infty$$

By the theorem above,

$$M_t^{f,k} = f_t(X_t^{(k)}) - \int_{t_0}^{t} \left(\frac{\partial}{\partial r} + \mathcal{L}_r^{(k)} \right) f_r(X_r^{(k)})\, dr, \qquad t \geq t_0,$$

is a martingale with respect to $P_{(t_0,\mu)}$, which coincides a.s. with M_t^f for $t < T_k$. Hence M_t^f is a local martingale up to $\zeta = \sup T_k$.

3. If $\zeta = \sup T_k = \infty$ a.s. and f satisfies (3.1.7), then $(M_t^f)_{t \geq 0}$ is a bounded local martingale, and hence, by dominated convergence, a martingale.

□

3.1.5 Diffusion processes

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

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x_0, \tag{3.1.9}$$

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. [7]. 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 (3.1.9) 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. [9].

In Section 3.6 we sketch another construction of diffusion processes in $\mathbb{R}^n$ that avoids stochastic analysis techniques to some extent.

3.2 Lyapunov functions and stability

In this section we explain briefly how Lyapunov function methods similar to those considered in Section 1.3 can be applied to Markov processes in continuous time. An excellent reference is the book by Khasminskii [15] that focuses on diffusion processes in $\mathbb{R}^n$. Most results in [15] easily carry over to more general Markov processes in continuous time.

We assume that we are given a right continuous process $((X_t), P)$ with polish state space S, initial value $X_0 = x_0 \in S$, and life time ζ. Let $\hat{\mathcal{A}} \subseteq C^{1,0}([0,\infty) \times S)$ be a linear subspace, and 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 $\mathcal{L}_t$ acts only on the x-variable. For $f \in \hat{\mathcal{A}}$ and $t < \zeta$ we define

$$M_t^f = f(t, X_t) - \int_0^t (\hat{\mathcal{L}} f)(s, X_s)\, ds$$

where it is implicitly assumed that the integral exists almost surely and defines a measurable function. We assume that (X_t) is adapted to a filtration $(\mathcal{F}_t)$ and it 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

(i) $S = \bigcup B_k$

(ii) The exit times $T_k := \inf\{t \geq 0 : X_t \notin B_k\}$ satisfy

$$T_k < \zeta \text{ on } \{\zeta < \infty\} \text{ for any } k \in \mathbb{N}, \text{ and } \zeta = \sup T_k.$$

(iii) The stopped processes $\left(M^f_{t\wedge T_k}\right)_{t\geq 0}$ are $(\mathcal{F}_t)$ martingales for any $k \in \mathbb{N}$ and $f \in \hat{\mathcal{A}}$.

Examples. 1) **Minimal jump process:** A minimal jump process as constructed in Section 3.1 satisfies the assumption if (B_k) is an increasing sequence 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

$$\hat{\mathcal{A}} = \left\{ f \in C^{1,0} : f, \frac{\partial f}{\partial t} \text{ bounded on } [0,t] \times B_k \text{ for any } t \geq 0 \text{ an } k \in \mathbb{N} \right\}.$$

2) **Minimal diffusion process:** A minimal Itô diffusion in $\mathbb{R}^n$ satisfies the assumption with $B_k = B(0, k)$ and $\hat{\mathcal{A}} = C^{1,2}([0,\infty) \times \mathbb{R}^n)$.

3.2.1 Non-explosion criteria

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

Theorem 3.5 (Khasminskii). *Suppose that Assumption (A) is satisfied and there exists a function $V \in \hat{\mathcal{A}}$ such that*

(i) $V(t, x) \geq 0$ *for any* $t \geq 0$ *and* $x \in S$,

(ii) $\inf_{\substack{x\in B_k^c \\ s\in[0,t]}} V(s, x) \to \infty$ *as* $k \to \infty$ *for any* $t \geq 0$,

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

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

Proof. Since $V(t, X_t) = M_t^V + \int_0^t \left(\frac{\partial V}{\partial s} + \mathcal{L}_s V\right)(s, X_s)\, ds$, optional stopping and Conditions (iii) and (i) imply

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

for any $t \geq 0$ and $k \in \mathbb{N}$. Therefore, for any $t \geq 0$,

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

by (ii), and hence $P[\zeta < \infty] = \lim_{t\to\infty} P[\zeta \leq t] = 0$. □

Remark (Time-independent Lyapunov functions). Suppose that U is a continuous function on S such that

$$\text{(i) } U \geq 0, \qquad \text{(ii) } \lim_{k\to\infty} \inf_{B_k^c} U = 0, \qquad \text{(iii) } \mathcal{L}_t U \leq \alpha U \text{ for some } \alpha > 0.$$

The Theorem 3.5 can be applied with $V(t, x) = e^{-\alpha t} U(x)$ provided this function is contained in $\hat{\mathcal{A}}$.

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. choose $\psi(n) = n$ as Lyapunov function. Then

$$(\mathcal{L}_t \psi)(n) = n \cdot b_t \cdot (n + 1 - n) + n \cdot d_t \cdot (n - 1 - n) = n \cdot (b_t - d_t) \leq n \sup_{t\geq 0} b_t$$

Since the individual birth rates $b_t,\ t \geq 0$, are bounded, 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

$$G_t(s) = 1 - \left(\frac{e^{\varrho_t}}{1-s} + \int_0^t b_n e^{\varrho_u}\, du\right)^{-1}$$

where

$$\varrho_t := \int_0^t (d_u - b_u)\, du$$

is the accumulated death rate. In particular, we obtain an explicit formula for the extinction probability:

$$\begin{aligned}P[X_t = 0] = \lim_{s \downarrow 0} G_t(s) &= \left(e^{\varrho_t} + \int_0^t b_n e^{\varrho_u}\, du\right)^{-1} \\ &= 1 - \left(1 + \int_0^t d_u e^{\varrho_u}\, du\right)^{-1}\end{aligned}$$

since $b = d - \varrho'$. Thus we have shown:

Theorem 3.6.

$$P[X_t = 0 \text{ eventually}] = 1 \iff \int_0^\infty d_u e^{\varrho_u}\, du = \infty$$

Remark. Informally, the mean and the variance of X_t can be computed by differentiating G_t at $s = 1$:

$$\frac{d}{ds} E\left[s^{X_t}\right]\Big|_{s=1} = E\left[X_t s^{X_t - 1}\right]\Big|_{s=1} = E[X_t]$$
$$\frac{d^2}{ds^2} E\left[s^{X_t}\right]\Big|_{s=1} = E\left[X_t(X_t-1)s^{X_t-2}\right]\Big|_{s=1} = \mathrm{Var}(X_t)$$

3.2.2 Hitting times and recurrence

Next, we 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.

Theorem 3.7 (Lyapunov bound for hitting times). *Suppose that Assumption A holds, and the process (X_t, P) is non-explosive. Furthermore, assume that there exist $V \in \hat{\mathcal{A}}$ and a measurable function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

(i) $V(t,x) \geq 0$ *for any* $t \geq 0$ *and* $x \in S$,

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

(iii) $\beta(t) := \int_0^t \alpha(s)\, ds \to \infty$ *as* $t \to \infty$.

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

$$E[\beta(T_A)] \leq V(0, x_0). \tag{3.2.1}$$

Proof. By Condition (ii),

$$V(t, X_t) \leq M_t^V - \int_0^t \alpha(s)\, ds = M_t^V - \beta(t)$$

holds for $t < T_A$. For any $k \in \mathbb{N}$, $M^V_{t\wedge T_k}$ is a martingale. Hence by (i),

$$0 \le E\left[V\left(t\wedge T_A\wedge T_k, X_{t\wedge T_A\wedge T_k}\right)\right] \le V(0,x_0) - E[\beta(t\wedge T_A\wedge T_k)].$$

As $k \to \infty, T_k \to \infty$ almost surely, and we obtain

$$\beta(t)P[t \le T_A] \le E[\beta(t\wedge T_A)] \le V(0,x_0)$$

for any $t > 0$. The assertion follows as $t \to \infty$. □

Example (Moments of hitting times).
If $\alpha(s) = cs^{n-1}$ for some $c > 0$ and $n \in \mathbb{N}$ then $\beta(s) = \frac{c}{n}s^n$. In this case, (3.2.1) is the moment bound

$$E[T_A^n] \le \frac{n}{c}V(0,x_0).$$

3.2.3 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. The following exercise covers the case of diffusions in $\mathbb{R}^n$:

Exercise (Explosion, occupation times and stationary distributions for diffusions on $\mathbb{R}^n$).
Consider a diffusion process (X_t, P_x) on $\mathbb{R}^n$ solving the local martingale problem for the generator

$$\mathcal{L}_t f \;=\; \frac{1}{2}\sum_{i,j=1}^n a_{i,j}(t,x)\frac{\partial^2 f}{\partial x_i\,\partial x_j} \;+\; \sum_{i=1}^n b_i(t,x)\frac{\partial f}{\partial x_i}, \qquad f \in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^n).$$

We assume that the coefficients are continuous functions and $P_x[X_0 = x] = 1$.

a) Prove that the process is non-explosive if there exist finite constants c_1, c_2, r such that

$$\operatorname{tr} a(t,x) \;\le\; c_1\,|x|^2 \quad \text{and} \quad x\cdot b(t,x) \;\le\; c_2\,|x|^2 \qquad \text{for } |x| \ge r.$$

b) Now suppose that $\zeta = \infty$ almost surely, and that there exist $V \in C^{1,2}(\mathbb{R}_+\times\mathbb{R}^n)$ and $\varepsilon, c \in \mathbb{R}_+$ such that $V \ge 0$ and

$$\frac{\partial V}{\partial t} + \mathcal{L}_t V \;\le\; \varepsilon + c1_B \qquad \text{on } \mathbb{R}_+\times\mathbb{R}^n,$$

where B is a ball in $\mathbb{R}^n$. Prove that

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

b) Conclude that if (X_t, P_x) is a time-homogeneous Markov process and the conditions above hold then there exists a stationary distribution.

Again the results carry over with similar proofs to general Markov processes. 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 $[0,t]$.

Lemma 3.8 (Lyapunov bound for occupation times). *Suppose 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*

$$\frac{\partial V}{\partial t} + \mathcal{L}_t V \le -\varepsilon + c1_B \quad \text{on } \mathbb{R}_+ \times S.$$

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 3.9 (Existence of stationary distributions). *Suppose that the assumptions in Lemma 3.8 hold, and moreover, assume that S is σ-compact, $V(t,x) = U(x)$ for some continuous function $U : S \to [0,\infty)$, and there exist $\varepsilon, c \in \mathbb{R}_+$ and a compact set $K \subseteq S$ such that*

$$\mathcal{L}U \le -\varepsilon + c1_K.$$

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

Proof. The assumptions imply

$$\liminf_{t\to\infty} \overline{p}_t(x_0,K) > 0.$$

The assertion now follows similarly as in discrete time, cf. Theorem **??**. □

3.3 Semigroups, generators and resolvents

In the discrete time case, there is a one-to-one correspondence between generators $\mathcal{L} = p - I$, transition semigroups $p_t = p^t$, 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 reference: for this section are Yosida: Functional Analysis [39], Pazy: Semigroups of Linear Operators [24], Davies: One-parameter semigroups [5] and Ethier/Kurtz [11].

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 any $t \geq 0$, and

(ii) $p_0(x, \cdot) = \delta_x$ and $p_t p_s = p_{t+s}$ for any $t, s \geq 0$ 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. $P_{(s,x)}$ withs state space $\mathbb{R}_+ \times S$ and transition function

$$\hat{P}_t\left((s, x), dudy\right) = \delta_{t+s}(du)p_{s,t+s}(x, dy).$$

3.3.1 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. *A positive measure $\mu \in M_+(S)$ 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*

$$\int p_t f d\mu \leq \int f d\mu \quad \textit{for any } f \in \mathcal{F}_+(S) \textit{ and } t \geq 0.$$

For processes with finite life-time, non-trivial invariant measures often do not exist, but in many cases non-trivial sub-invariant measures do exist.

Lemma 3.10 (Sub-Markov semigroup and contraction properties). *1) Any transition function $(p_t)_{t\geq 0}$ induces a **sub-Markovian semigroup** on $\mathcal{F}_b(S)$ or $\mathcal{F}_+(S)$ respectively, i.e., 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) The semigroup is contractive w.r.t. the supremum norm:

$$\|p_t f\|_{\mathit{sup}} \leq \|f\|_{\mathit{sup}} \quad \textit{for any } t \geq 0 \textit{ and } f \in \mathcal{F}_b(S).$$

3) If $\mu \in \mathcal{M}_+(S)$ is a sub-invariant measure then (p_t) is also contractive w.r.t. the $L^p(\mu)$ norm for every $p \in [1, \infty]$:

$$\int |p_t f|^p d\mu \leq \int |f|^p d\mu \quad \textit{for any } f \in \mathcal{L}^p(S, \mu).$$

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 $p \in [1, \infty)$:
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

$$|p_t f|^p \leq (p_t |f|)^p \leq p_t |f|^p$$

by Jensen's inequality. 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,

- $C_b(S)$ if p_t is Feller for any $t \geq 0$,
- $\hat{C}(S) = \{f \in C(S) : \forall \varepsilon > 0\ \exists K \subset S$ compact: $|f| < \varepsilon$ on $S \backslash K\}$ if p_t maps $\hat{C}(S)$ to $\hat{C}(S)$ (classical Feller property),
- $L^p(S, \mu), p \in [1, \infty]$, if μ is sub-invariant.

We will see below that for obtaining a densely defined generator, an additional property called strong continuity is required for the semigroups. This will exclude some of the Banach spaces above. Before discussing strong continuity, we introduce another fundamental object that will enable us to establish the connection between a semigroup and its generator: the resolvent.

Definition (Resolvent kernels). *The* ***resolvent kernels*** *associated to the transition function* $(p_t)_{t \geq 0}$ *are defined by*

$$g_\alpha(x, dy) = \int_0^\infty e^{-\alpha t} p_t(x, dy) dt \quad \text{for } \alpha \in (0, \infty),$$

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 absorbed Markov process with start in x and constant **absorption rate** α.

Lemma 3.11 (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., 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 any* $\alpha > 0$,

$$\|\alpha g_\alpha f\|_{sup} \leq \|f\|_{sup} \quad \text{for any } f \in \mathcal{F}_b(S).$$

3) ***Contractivity w.r.t.*** $\mathrm{L^p}$ ***norms:*** *If* $\mu \in \mathcal{M}_+(S)$ *is sub-invariant w.r.t.* (p_t) *then*

$$\|\alpha g_\alpha f\|_{L^p(S,\mu)} \le \|f\|_{L^p(S,\mu)} \quad \textit{for any } \alpha > 0, p \in [1,\infty], \textit{ and } f \in \mathcal{L}^p(S,\mu).$$

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 $(p_t)_{t\ge0}$ w.r.t. $\|\cdot\|$ implies that also (αg_α) is contractive w.r.t. $\|\cdot\|$:

$$\|\alpha g_\alpha f\| \le \int_0^\infty \alpha e^{-\alpha t} \|p_t f\| dt \le \int_0^\infty \alpha e^{-\alpha t} dt \, \|f\| = \|f\| \quad \text{for any } \alpha > 0.$$

□

The lemma shows that $(g_\alpha)_{\alpha>0}$ induces contraction resolvents of linear operators $(G_\alpha)_{\alpha>0}$ 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:

$$\text{(R)} \qquad \text{Range}(G_\alpha) = \text{Range}(G_\beta) \qquad \text{for any } \alpha, \beta \in (0,\infty).$$

This property will be important below.

3.3.2 Strong continuity and Generator

We now assume that $(P_t)_{t\ge0}$ is a semigroup of linear contractions on a Banach space E. Our goal is to define the infinitesimal generator L of (P_t) by $Lf = \lim\limits_{t\downarrow0} \frac{1}{t}(P_t f - f)$ 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 $\lim\limits_{t\downarrow0} \|P_t f - f\| = 0$ for any $f \in \mathcal{D}$, and hence, by contractivity of the operators P_t, for any $f \in E$. A semigroup with this property is called strongly continuous:

Definition ($\mathbf{C^0}$ semigroup, Generator). *1) The semigroup $(P_t)_{t\geq 0}$ on the Banach space E is called **strongly continuous** $(\mathbf{C^0})$ iff $P_0 = I$ and*

$$\|P_t f - f\| \to 0 \quad \textit{as } t \downarrow 0 \textit{ for any } f \in E.$$

*2) The **generator** of $(P_t)_{t\geq 0}$ is the linear operator $(L, Dom(L))$ given by*

$$Lf = \lim_{t\downarrow 0} \frac{P_t f - f}{t}, \quad Dom(L) = \left\{ f \in E : \lim_{t\downarrow 0} \frac{P_t f - f}{t} \textit{ exists} \right\}.$$

Here the limits are taken w.r.t. the norm on the Banach space E.

Remark (Strong continuity). 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 3.15 below shows that the generator of a C^0 contraction semigroup is densely defined.

Theorem 3.12 (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 $f \in Dom(L)$ then $P_t f \in Dom(L)$ for any $t \geq 0$, 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 $t > 0$,

$$\|P_{t+h} f - P_t f\| = \|P_t(P_h f - f)\| \leq \|P_h f - f\| \to 0 \quad \text{as } h \downarrow 0,$$

and, similarly, for any $t > 0$,

$$\|P_{t-h} f - P_t f\| = \|P_{t-h}(f - P_h f)\| \leq \|f - P_h f\| \to 0 \quad \text{as } h \downarrow 0.$$

2) Similarly, the forward equation $\frac{d}{dt}P_tf = P_tLf$ follows from the semigroup property, contractivity, strong continuity and the definition of the generator: For any $f \in \mathrm{Dom}(L)$ and $t \geq 0$,
$$\frac{1}{h}(P_{t+h}f - P_tf) = P_t\frac{P_hf - f}{h} \to P_tLf \quad \text{as } h \downarrow 0,$$
and, for $t > 0$,
$$\frac{1}{-h}(P_{t-h}f - P_tf) = P_{t-h}\frac{P_hf - f}{h} \to P_tLf \quad \text{as } h \downarrow 0$$
by strong continuity.

3) Finally, the backward equation $\frac{d}{dt}P_tf = LP_tf$ is a consequence of the forward equation: For $f \in \mathrm{Dom}(L)$ and $t \geq 0$,
$$\frac{P_hP_tf - P_tf}{h} = \frac{1}{h}(P_{t+h}f - P_tf) \to P_tLf \quad \text{as } h \downarrow 0.$$
Hence P_tf is in the domain of the generator, and $LP_tf = P_tLf = \frac{d}{dt}P_tf$.

□

3.3.3 Strong continuity of transition semigroups of Markov processes

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
$$(p_tf)(x) = (f * \varphi_t)(x) = \int_{\mathbb{R}} f(y)\varphi_t(x-y)\,dy,$$
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_tf is a continuous function for any $f \in \mathcal{F}_b(\mathbb{R})$,
$$\|p_t1_{(0,1)} - 1_{(0,1)}\|_\infty \geq \frac{1}{2} \quad \text{for any } t > 0.$$

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 3.14 below shows that the semigroups induced by (p_t) on the Banach spaces $\hat{C}(\mathbb{R})$ and $L^p(\mathbb{R}, dx)$ with $p \in [1, \infty)$ are strongly continuous.

Lemma 3.13.
If $(p_t)_{t\geq 0}$ is the transition function of a right-continuous Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ then

$$(p_t f)(x) \to f(x) \text{ as } t \downarrow 0 \quad \text{for any } f \in C_b(S) \text{ and } x \in S. \tag{3.3.1}$$

Moreover, if the linear operators induced by p_t are contractions w.r.t. the supremum norm or an L^p norm then

$$\|p_t f - f\| \to 0 \text{ as } t \downarrow 0 \quad \text{for any } f = g_\alpha h, \tag{3.3.2}$$

where $\alpha \in (0, \infty)$ and h is a function in $\mathcal{F}_b(S)$ or in the corresponding $\mathcal{L}^p$-space respectively.

Proof. For $f \in C_b(S), t \mapsto f(X_t)$ is right continuous and bounded. Therefore, by dominated convergence,

$$(p_t f)(x) = E_x\left[f(X_t)\right] \to E_x\left[f(X_0)\right] = f(x) \text{ as } t \downarrow 0.$$

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 $\mathcal{F}_b(S)$ or in the $\mathcal{L}^p$ space where (p_t) is contractive. Then for $t \geq 0$,

$$\begin{aligned} p_t f &= \int_0^\infty e^{-\alpha s} p_{s+t} h\, ds = e^{\alpha t} \int_t^\infty e^{-\alpha u} p_u h\, du \\ &= e^{\alpha t} f - e^{\alpha t} \int_0^t e^{-\alpha u} p_u h\, du, \end{aligned}$$

and hence

$$\|p_t f - f\| \leq (e^{\alpha t} - 1)\|f\| + e^{\alpha t} \int_0^t \|p_u h\| du.$$

Since $\|p_u h\| \leq \|h\|$ by assumption, the right-hand side converges to 0 as $t \downarrow 0$. □

Theorem 3.14 (Strong continuity of transition functions). *Suppose that (p_t) is the transition function of a right-continuous time-homogeneous Markov process on $(S, \mathcal{B})$.*

1) *If $\mu \in \mathcal{M}_+(S)$ is a sub-invariant measure for (p_t) then (p_t) induces a strongly continuous contraction semigroup of linear operators on $L^p(S, \mu)$ for every $p \in [1, \infty)$.*

2) *If S is locally compact and $p_t\left(\hat{C}(S)\right) \subseteq \hat{C}(S)$ for any $t \geq 0$ then (p_t) induces a strongly continuous contraction semigroup of linear operators on $\hat{C}(S)$.*

Proof. 1) We have to show that for any $f \in \mathcal{L}^p(S,\mu)$,

$$\|p_t f - f\|_{L^p(S,\mu)} \to 0 \quad \text{as } t \downarrow 0. \tag{3.3.3}$$

(i) We first show that (3.3.3) holds for $f \in C_b(S) \cap \mathcal{L}^1(S,\mu)$. To this end we may assume w.l.o.g. that $f \geq 0$. Then $p_t f \geq 0$ for all t, and hence $(p_t f - f)^- \leq f$. By sub-invariance of μ:

$$\int |p_t f - f| d\mu = \int (p_t f - f) d\mu + 2\int (p_t f - f)^- d\mu \leq 2\int (p_t f - f)^- d\mu,$$

and hence by dominated convergence and (3.3.1),

$$\limsup_{t\downarrow 0} \int |p_t f - f| d\mu \leq 0.$$

This proves (3.3.3) for $p = 1$. For $p > 1$, we now obtain

$$\int |p_t f - f|^p d\mu \leq \int |p_t f - f| d\mu \cdot \|p_t f - f\|_{\text{sup}}^{p-1} \to 0 \quad \text{as } t \downarrow 0,$$

where we have used that p_t is a contraction w.r.t. the supremum norm. For an arbitrary function $f \in \mathcal{L}^p(S,\mu)$, (3.3.3) follows by an $\varepsilon/3$ argument: Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $C_b(S) \cap \mathcal{L}^1(\mu)$ such that $f_n \to f$ in $L^p(S,\mu)$. Then, given $\varepsilon > 0$,

$$\begin{aligned}
\|p_t f - f\|_{L^p} &\leq \|p_t f - p_t f_n\|_{L^p} + \|p_t f_n - f_n\|_{L^p} + \|f_n - f\|_{L^p} \\
&\leq 2\|f - f_n\|_{L^p} + \|p_t f_n - f_n\|_{L^p} < \varepsilon
\end{aligned}$$

if n is chosen sufficiently large and $t \geq t_0(n)$.

2) We have to show that for any $f \in \hat{C}(S)$,

$$\|p_t f - f\|_{\text{sup}} \to 0 \quad \text{as } t \downarrow 0. \tag{3.3.4}$$

By Lemma 3.13, (3.3.4) holds if $f = g_\alpha h$ for some $\alpha > 0$ and $h \in \hat{C}(S)$. 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 $\alpha > 0$ - then 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

$$\mu(g_\alpha h) = 0 \quad \text{for any } h \in \hat{C}(S),$$

cf. [?]. 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, (3.3.1) implies that $\beta g_\beta h \to h$ pointwise as $\beta \to \infty$. 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.

□

3.3.4 One-to-one correspondence

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 $(L, \mathrm{Dom}(L))$. Since $t \mapsto P_t f$ is a continuous function by Theorem 3.12, a corresponding resolvent can be defined as an E-valued Riemann integral:

$$G_\alpha f = \int_0^\infty e^{-\alpha t} P_t f \, dt \quad \text{for any } \alpha > 0 \text{ and } f \in E. \tag{3.3.5}$$

Exercise (Strongly continuous contraction resolvent).
Prove that the linear operators $G_\alpha, \alpha \in (0,\infty)$, defined by (3.3.5) form a **strongly continuous contraction resolvent**, i.e.,

(i) $G_\alpha f - G_\beta f = (\beta - \alpha) G_\alpha G_\beta f \quad$ for any $f \in E$ and $\alpha, \beta > 0$,

(ii) $\|\alpha G_\alpha f\| \leq \|f\| \quad$ for any $f \in E$ and $\alpha > 0$,

(iii) $\|\alpha G_\alpha f - f\| \to 0 \quad$ as $\alpha \to \infty$ for any $f \in E$.

Theorem 3.15 (Connection between resolvent and generator). *For any $\alpha > 0$, $G_\alpha = (\alpha I - L)^{-1}$. 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 that $G_\alpha f$ is contained in the domain of L. Indeed, as $t \downarrow 0$,

$$\begin{aligned}
\frac{P_t G_\alpha f - G_\alpha f}{t} &= \frac{1}{t}\left(\int_0^\infty e^{-\alpha s} P_{t+s} f \, ds - \int_0^\infty e^{-\alpha s} P_s f \, ds\right) \\
&= \frac{e^{\alpha t} - 1}{t} \int_0^\infty e^{-\alpha s} P_s f \, ds - e^{\alpha t} \frac{1}{t} \int_0^t e^{-\alpha s} P_0 f \, ds \\
&\to \alpha G_\alpha f - f
\end{aligned}$$

by strong continuity of $(P_t)_{t\geq 0}$. Hence $G_\alpha f \in \mathrm{Dom}(L)$ and

$$LG_\alpha f = \alpha G_\alpha f - f,$$

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,

$$\mathrm{Dom}(L) = \mathrm{Dom}(\alpha I - L) = \mathrm{Range}(G_\alpha) \quad \text{for any } \alpha > 0.$$

By strong continuity of the resolvent,

$$\alpha G_\alpha f \to f \quad \text{as } \alpha \to \infty \quad \text{for any } f \in E,$$

so the domain of L is dense in E. $\square$

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 theorem:

Theorem 3.16 (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 Dom$(L) \subset$ Dom$(\widetilde{L})$. Then*

$$P_t f - \widetilde{P}_t f = \int_0^t \widetilde{P}_s(\widetilde{L} - L)P_{t-s}f\,ds \quad \text{for any } t \geq 0 \text{ and } f \in \mathrm{Dom}(L). \tag{3.3.6}$$

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

$$P_{t-s}f \in \mathrm{Dom}(L) \subset \mathrm{Dom}(\widetilde{L})$$

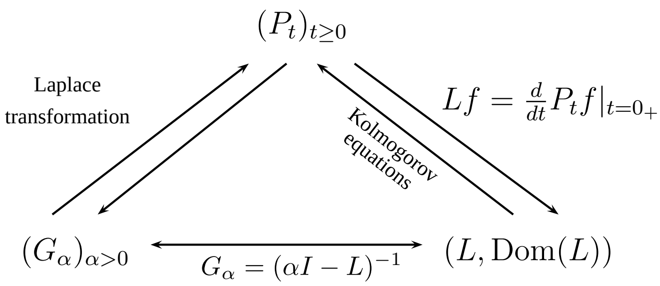
by Theorem 3.12. By combining the forward and backward equation in Theorem 3.12 we can then show that

$$\frac{d}{ds}\widetilde{P}_s P_{t-s} f = \widetilde{P}_s \widetilde{L} P_{t-s} f - \widetilde{P}_s L P_{t-s} f = \widetilde{P}_s (\widetilde{L} - L) P_{t-s} f$$

where the derivative is as usual taken in the Banach space E. The identity (3.3.6) now follows by the fundamental theorem of calculus for Banach-space valued functions, cf. e.g. Lang: Analysis 1 [17].

In particular, if the generator of $\widetilde{P}_t$ is an extension of L then (3.3.6) 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 3.15, this implies that the semigroups (P_t) and $(\widetilde{P}_t)$ coincide. $\square$

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 picture:



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 and tells us how to recover the semigroup from the resolvent.

3.3.5 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 [11] for a detailed proof.

Theorem 3.17 (Hille-Yosida). *A linear operator $(L, Dom(L))$ on the Banach space E is the generator of a strongly continuous contraction semigroup if and only if the following conditions hold*

(i) $Dom(L)$ is dense in E,

(ii) $Range(\alpha I - L) = E$ for some $\alpha > 0$ (or, equivalently, for any $\alpha > 0$),

*(iii) L is **dissipative**, i.e.,*

$$\|\alpha f - Lf\| \geq \alpha \|f\| \quad \text{for any } \alpha > 0, f \in Dom(L).$$

Proof. "$\Rightarrow$": If L generates a C^0 contraction semigroup then by Theorem 3.15, $(\alpha I - L)^{-1} = G_\alpha$ where (G_α) 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

$$\alpha\|f\| = \|\alpha G_\alpha g\| \leq \|g\| = \|\alpha f - Lf\|$$

by contractivity of αG_α.
"$\Leftarrow$": 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 $f \in \mathrm{Dom}(L)$,

$$Lf = \lim_{\alpha\to\infty} \alpha G_\alpha Lf = \lim_{\alpha\to\infty} L^{(\alpha)} f$$

where $L^{(\alpha)}$ is the **bounded** linear operator defined by

$$L^{(\alpha)} = \alpha L G_\alpha = \alpha^2 G_\alpha - \alpha I \quad \text{for } \alpha \in (0, \infty).$$

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 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 $L^{(\alpha)}$ for every $\alpha > 0$. Moreover, since $\left(L^{(\alpha)} f\right)_{\alpha \in \mathbb{N}}$ is a Cauchy sequence for any $f \in \mathrm{Dom}(L)$, Duhamel's formula (3.3.6) shows that also $\left(P_t^{(\alpha)} f\right)_{\alpha \in \mathbb{N}}$ is a Cauchy sequence for any $t \geq 0$ and $f \in \mathrm{Dom}(L)$. We can hence define

$$P_t f = \lim_{\alpha \to \infty} P_t^{(\alpha)} f \quad \text{for any } t \geq 0 \text{ and } f \in \mathrm{Dom}(L). \tag{3.3.7}$$

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 (3.3.7) 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. $\square$

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.,

$$(f, Lf) \leq 0 \quad \text{for any } f \in \mathrm{Dom}(L).$$

In this case, the C^0 semigroup generated by L is given by

$$P_t = e^{tL} \quad \text{for any } t \geq 0,$$

where the exponential is defined by spectral theory, cf. e.g. Reed& Simon:Methods of modern mathematical physics I [28], II [26], III [29], IV [27].

3.4 Martingale problems for Markov processes

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. We assume at first 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* $(p_t)_{t\geq 0}$. Suppose moreover that E is either 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 $t\geq 0,$ 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 $\mu\in\mathcal{M}_+(S)$.

3.4.1 From Martingale problem to Generator

In many situations it is known that for any $x\in S$, 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 3.18 (From the martingale problem to C_0 semigroups and generators). *Suppose that for any $x\in S$ 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 any $t\geq 0$, 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 $(L,Dom(L))$ of $(P_t)_{t\geq 0}$ is an extension of $(\mathcal{L},\mathcal{A})$.

Remark. In the case of Markov processes with finite life-time the statement 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

$$\begin{aligned}(p_t f)(x) - f(x) = E_x[f(X_t) - f(X_0)] &= E_x\left[\int_0^t (\mathcal{L}f)(X_s)ds\right] \\ &= \int_0^t E_x[(\mathcal{L}f)(X_s)]ds = \int_0^t (p_s\mathcal{L}f)(x)ds \end{aligned} \tag{3.4.1}$$

for any $t \geq 0, x \in S$ and $f \in \mathcal{A}$. By the assumptions and Lemma 3.10, p_t is contractive w.r.t. the norm on E for any $t \geq 0$. Therefore, by (3.4.1),

$$\|p_t f - f\|_E \leq \int_0^t \|p_s\mathcal{L}f\|_E ds \leq t\|\mathcal{L}f\|_E \to 0 \quad \text{as } t \downarrow 0$$

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, (3.4.1) implies that

$$\left\|\frac{p_t f - f}{t} - \mathcal{L}f\right\|_E \leq \frac{1}{t}\int_0^t \|p_s\mathcal{L}f - \mathcal{L}f\|_E ds \to 0 \quad \text{as } t \downarrow 0 \tag{3.4.2}$$

for any $f \in \mathcal{A}$. Here we have used that $\lim_{s\downarrow 0} p_s\mathcal{L}f = \mathcal{L}f$ by the strong continuity. By (3.4.2), the functions in $\mathcal{A}$ are contained in the domain of the generator L of (P_t), and $Lf = \mathcal{L}f$ for any $f \in \mathcal{A}$. □

3.4.2 Identification of the generator

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 martingale problem for any 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!

Example (Brownian motion with reflection and Brownian motion with absorption).
Let $S = [0, \infty)$ and $E = L^2(S, dx)$. We consider the linear operator $\mathcal{L} = \frac{1}{2}\frac{d^2}{dx^2}$ with dense domain $\mathcal{A} = C_0^\infty(0, \infty) \subset L^2(S, dx)$. Suppose that $((B_t)_{t\geq 0}, (P_x)_{x\in\mathbb{R}})$ is 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 by

$$X_t = |B_t| \qquad \text{for any } t \geq 0.$$

- **Brownian motion on $\mathbb{R}_+$ with absorption at 0** is defined by

$$\widetilde{X}_t = \begin{cases} B_t & \text{for } t < T_0^B, \\ \Delta & \text{for } t \geq T_0^B, \end{cases}$$

where $T_0^B = \inf\{t \geq 0 : B_t = 0\}$ is the first hitting time of 0 for (B_t).

Exercise. Prove that both (X_t, P_x) and $(\widetilde{X}_t, P_x)$ 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 example above 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 (Closability and closure of linear operators, operator cores). *1) A linear operator $(\mathcal{L}, \mathcal{A})$ is called **closable** iff it has a closed extension.*

*2) In this case, the smallest closed extension $(\overline{\mathcal{L}}, Dom(\overline{\mathcal{L}}))$ is called the **closure** of $(\mathcal{L}, \mathcal{A})$. It is given explicitly by*

$$\begin{aligned} Dom(\overline{\mathcal{L}}) &= \text{ completion of } \mathcal{A} \text{ w.r.t. the graph norm } \|\cdot\|_{\mathcal{L}}, \\ \overline{\mathcal{L}}f &= \lim_{n\to\infty} \mathcal{L}f_n \quad \text{for any sequence } (f_n)_{n\in\mathbb{N}} \text{ in } \mathcal{A} \text{ such that } f_n \to f \text{ in } E \\ &\qquad \text{and } (\mathcal{L}f_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence.} \end{aligned} \tag{3.4.3}$$

*3) Suppose that L is a linear operator on E with $\mathcal{A} \subseteq Dom(L)$. Then $\mathcal{A}$ is called a **core** for L iff $\mathcal{A}$ is dense in Dom(L) w.r.t. the graph norm $\|\cdot\|_L$.*

It is easy to verify that if an operator is closable then the extension defined by (3.4.3) is indeed the smallest closed extension. 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 $E \times E$. 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 $(L, \text{Dom}(L))$. The subspace $\mathcal{A} \subseteq \text{Dom}(L)$ is a core for L if and only if $(L, \text{Dom}(L))$ is the closure of $(\mathcal{L}, \mathcal{A})$.

Theorem 3.19 (Strong uniqueness). *Suppose that $\mathcal{A}$ is a dense subspace of the domain of the generator L. Then the following statements are equivalent:*

(i) $\mathcal{A}$ is a core for L.

(ii) $P_t f$ is contained in the completion of $\mathcal{A}$ w.r.t. the graph norm $\|\cdot\|_L$ for any $f \in Dom(L)$ and $t \in (0, \infty)$.

If (i) or (ii) hold then

(iii) $(P_t)_{t\geq 0}$ 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 3.12, $P_t f$ is contained in the domain of L for any $t > 0$ and $f \in \mathrm{Dom}(L)$.
(ii) $\Rightarrow$ (i): Let $f \in \mathrm{Dom}(L)$. We have to prove that f can be approximated by functions in the closure $\overline{\mathcal{A}}^L$ of $\mathcal{A}$ w.r.t. the graph norm of L. If (ii) holds this can be done by regularizing f via the semigroup: For any $t > 0$, $P_t f$ is contained in the closure of $\mathcal{A}$ w.r.t. the graph norm by (ii). Moreover, $P_t f$ converges to f as $t \downarrow 0$ by strong continuity, and

$$LP_t f = P_t L f \to Lf \quad \text{as } t \downarrow 0$$

by strong continuity and Theorem 3.12. So

$$\|P_t f - f\|_L \to 0 \quad \text{as } t \downarrow 0,$$

and thus f is also contained in the closure of $\mathcal{A}$ w.r.t. the graph norm.
(i) $\Rightarrow$ (iii): If (i) holds and $(\widetilde{P})_{t\geq 0}$ is a 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 operator by Theorem 3.15. Hence the semigroups $(\widetilde{P}_t)$ and (P_t) agree by Theorem 3.16. □

We now apply Theorem 3.19 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 3.20 (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*

$$L = \frac{1}{2}\Delta, \qquad \textit{Dom}(L) = \overline{C_0^\infty(\mathbb{R}^n)}^{\Delta},$$

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.*

Example (Generator of Brownian motion on $\mathbb{R}$**).** In the one-dimensional case, the generators are given explicitly by

$$Lf = \frac{1}{2}f'', \ \mathrm{Dom}(L) = \left\{ f \in \hat{C}(\mathbb{R}) \cap C^2(\mathbb{R}) : f'' \in \hat{C}(\mathbb{R}) \right\}, \tag{3.4.4}$$

$$Lf = \frac{1}{2}f'', \ \mathrm{Dom}(L) = \left\{ f \in L^p(\mathbb{R}, dx) \cap C^1(\mathbb{R}) : f' \text{ absolutely continuous, } f'' \in L^p(\mathbb{R}, dx) \right\}, \tag{3.4.5}$$

respectively.

Remark (Domain 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 $L^p(\mathbb{R}^n, dx)$, cf. e.g. $[XXX]$.

Proof. By Itô's formula, 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, 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 3.18, $(p_t)_{t\geq 0}$ 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 $\left(\frac{1}{2}\Delta, C_0^\infty(\mathbb{R}^n)\right)$. 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 3.19, the generators on $\hat{C}(S)$ and $L^p(\mathbb{R}^n, dx)$ coincide with the closures of $\left(\frac{1}{2}\Delta, C_0^\infty(\mathbb{R}^n)\right)$. □

Exercise (Generators of Brownian motions with absorption and reflection). 1) Show that Brownian motion with absorption at 0 induces a strongly continuous contraction semigroup $(P_t)_{t\geq 0}$ on the Banach space $E = \{f \in C(0,\infty) : \lim_{x\downarrow 0} f(x) = 0 = \lim_{x\uparrow\infty} f(x)\}$. Prove that

$$\mathcal{A} = \{f|_{(0,\infty)} : f \in C_0^\infty(\mathbb{R}) \text{ with } f(0) = 0\}$$

is a core for the generator L which is given by $Lf = \frac{1}{2}f''$ for $f \in \mathcal{A}$. Moreover, show that $C_0^\infty(0,\infty)$ is **not a core** for L.

2) Show that Brownian motion with reflection at 0 induces a strongly continuous contraction semigroup on the Banach space $E = \hat{C}([0,\infty))$, and prove that a core for the generator is given by

$$\mathcal{A} = \{f|_{[0,\infty)} : f \in C_0^\infty(\mathbb{R}) \text{ with } f'(0) = 0\}.$$

3.4.3 Uniqueness of martingale problems

From now on 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. 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 3.21 (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*

$$p_t f = P_t f \quad \text{for any } t \geq 0 \text{ and } f \in E. \tag{3.4.6}$$

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/Kurtz, Chapter 4, Theorem 4.1 [11].

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. [9].

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 $g \in E$,

$$E\left[\int_0^\infty e^{-\alpha t} g(X_{s+t}) \middle| \mathcal{F}_s^X\right] = (G_\alpha g)(X_s) \quad P\text{-a.s.} \tag{3.4.7}$$

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 (3.4.7) follows as $t \to \infty$.

Step 4 (Markov property in semigroup form). One can now conclude that

$$E[g(X_{s+t}) | \mathcal{F}_s^X] = (P_s g)(X_s) \quad P\text{-a.s.} \tag{3.4.8}$$

holds for any $s, t \geq 0$ and $g \in E$. The proof is based on the approximation

$$P_s g = \lim_{n \to \infty} \left(\frac{n}{s} G_{\frac{n}{s}}\right)^n g$$

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 (3.4.6). 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

$$\left(\frac{n}{t}G_{\frac{n}{t}}\right)^n g \to P_t g \quad \text{uniformly as } n \to \infty. \tag{3.4.9}$$

How could you derive (3.4.9) 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 3.21. Further uniqueness results for martingale problems with assumptions that may be easier to verify in applications can be found in Stroock/Varadhan [35] and Ethier/Kurtz [11].

3.4.4 Strong Markov property

In Theorem 3.21 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 [2]. 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 3.22 (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).

Sketch of proof of Theorem 3.22. 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 $(\mathcal{F}_{T+t}^X)_{t\geq 0}$.

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 $A \in \mathcal{F}_{T+s}^X$,

$$\begin{aligned}
&E_{Q_\omega}\left[f(X_{T+t}) - f(X_{T+s}) - \int_{T+s}^{T+t} (\mathcal{L}f)(X_r)dr; A\right] \\
&= E_x\left[\left(M_{T+t}^{[f]} - M_{T+s}^{[f]}\right) 1_A \middle| \mathcal{F}_T^X\right](\omega) \\
&= E_x\left[E_x\left[M_{T+t}^{[f]} - M_{T+s}^{[f]} \middle| \mathcal{F}_{T+s}^X\right] 1_A \middle| \mathcal{F}_T^X\right](\omega) = 0
\end{aligned}$$

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/Varadhan [35], Roger/Williams [31]+[32] or Ethier/Kurz [11]. $\square$

3.5 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. We start with a definition:

Definition (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 $t \geq 0$.

Lemma 3.23 (Feller processes, generators and martingales). *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 $p_t\big(\hat{C}(S)\big) \subseteq \hat{C}(S)$ 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. Strong continuity holds by 3.11. Filling in the other missing details is left as an exercise. □

3.5.1 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 3.24 (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*

$$E_x[f(X_t)|\mathcal{F}_s^X] = (P_{t-s}f)(X_s) \quad P_x\text{-almost surely} \tag{3.5.1}$$

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 $x \in S$.

Remark (Strong Markov property). In a similar way as for Brownian motion it can be shown that $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ is a strong Markov process, cf. e.g. Liggett [19].

Sketch of proof. We only mention the main steps in the proof, details can be found for instance in Rogers& Williams, [32]:

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

$$(P_t f)(x) = \int p_t(x, dy) f(y) \qquad \text{for any } f \in \hat{C}(S) \text{ and } x \in S.$$

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

$$P_x \circ (X_{t_1}, X_{t_2}, \dots, X_{t_n})^{-1} = p_{t_1}(x, dy_1) p_{t_2 - t_1}(y_1, dy_2) \dots p_{t_n - t_{n-1}}(y_{n-1}, dy_n)$$

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 (3.5.1) and the initial condition.

□

We remark that alternatively it is possible to construct a Feller process as a limit of jump processes, cf. Chapter 4, Theorem 5.4. in Ethier&Kurtz [11]. Indeed, the Yosida approximation

$$\begin{aligned} Lf &= \lim_{\alpha\to\infty} \alpha G_\alpha L f = \lim_{\alpha\to\infty} L^{(\alpha)} f, \quad L^{(\alpha)} f := \alpha(\alpha G_\alpha f - f),\\ P_t f &= \lim_{\alpha\to\infty} e^{tL^{(\alpha)}} f, \end{aligned}$$

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 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 by P_x the limit, $((X_t), P_x)$ is a Markov process that solves the martingale problem for the generator $(L, \mathrm{Dom}(L))$. We return to this approximation approach for constructing solutions of martingale problems in Section 3.6.

3.5.2 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 3.25 (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* $f(x_0) = \sup f$ *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 3.14. 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 $\|f\|_{\mathrm{sup}} = |f(x_0)|$. Assuming w.l.o.g. $f(x_0) \geq 0$, we obtain

$$\alpha \|f\|_{\mathrm{sup}} \leq \alpha f(x_0) - (Lf)(x_0) \leq \|\alpha f - Lf\|_{\mathrm{sup}} \qquad \text{for any } \alpha > 0$$

by (iii). The Hille-Yosida Theorem 3.14 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

$$e^{tL^{(\alpha)}} f = e^{-\alpha t} \sum_{n=0}^{\infty} \frac{(\alpha t)^n}{n!} (\alpha G_\alpha)^n f \ \in [0,1) \qquad \text{for any } t \geq 0.$$

Hence also $P_t f = \lim_{\alpha\to\infty} e^{tL^{(\alpha)}} f \in [0,1]$ for any $t \geq 0$.

□

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:

Theorem 3.26 (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 Markov process $((X_t)_{t\geq 0}, (P_x)_{x\in\mathbb{R}^d})$ with continuous paths then there exist functions $a_{ij}, b_i, c \in C(\mathbb{R}^d)$ $(i,j = 1,\dots,d)$ such that for any x, $a_{ij}(x)$ is non-negative definite, $c(x) \leq 0$, and*

$$(Lf)(x) = \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x) + c(x) f(x) \quad \forall f \in C_0^\infty(\mathbb{R}^d). \tag{3.5.2}$$

Furthermore, if the process $((X_t)_{t\geq 0}, (P_x)_{x\in S})$ is non-explosive then $c \equiv 0$.

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

$$(Lf)(x) = \lim_{\varepsilon \downarrow 0} \frac{E_x\left[\int_0^{T_\varepsilon} (Lf)(X_s)ds\right]}{E_x[T_\varepsilon]} = \lim_{\varepsilon \downarrow 0} \frac{E_x[f(X_{T_\varepsilon})] - f(x)}{E_x[T_\varepsilon]}. \tag{3.5.3}$$

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 (3.5.3) 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 (3.5.3).

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 then $(Lf)(x) \le 0$. Indeed, in this case 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

$$(Lf)(x) = (L\widetilde{f})(x) \le 0.$$

3) **Taylor expansion:** For proving that L is a differential operator of the form (3.5.2) 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

$$f(y) = f(x)\varphi(y) + \sum_{i=1}^d \frac{\partial f}{\partial x_i}(x)\psi_i(y) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\psi_i(y)\psi_j(y) + R(y) \tag{3.5.4}$$

in a neighbourhood of x. Since L is a local linear operator, we obtain

$$(Lf)(x) = c(x)f(x) + \sum_{i=1}^d b_i(x)\frac{\partial f}{\partial x_i}(x) + \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) + (LR)(x) \tag{3.5.5}$$

with $c(x) := (L\varphi)(x), b_i(x) := (L\psi_i)(x)$, and $a_{ij}(x) := L(\psi_i\psi_j)(x)$. Since φ has a local maximum at x, $c(x) \le 0$. 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

$$\sum_{i,j=1}^d \xi_i\xi_j a_{ij}(x) = L\left(\sum_{i,j}\xi_i\xi_j\psi_i\psi_j\right) \ge 0,$$

i.e., the matrix $(a_{ij}(x))$ is non-negative definite. By (3.5.5), 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

$$0 \ge (LR_\varepsilon)(x) = (LR)(x) - \varepsilon \sum_{i=1}^{d} a_{ii}(x) \quad \forall \varepsilon > 0.$$

Letting ε tend to 0, we obtain $(LR)(x) \le 0$. On the other hand, R_ε has a local minimum at x for $\varepsilon < 0$, and in this case the local maximum principle implies

$$0 \le (LR_\varepsilon)(x) = (LR)(x) - \varepsilon \sum_{i=1}^{d} a_{ii}(x) \quad \forall \varepsilon < 0,$$

and hence $(LR)(x) \ge 0$. Thus $(LR)(x) = 0$.

4) **Vanishing of c:** If the process is non-explosive then $p_t 1 = 1$ for any $t \ge 0$. 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 3.26 has an extension to generators of general Feller semigroups including those corresponding to processes with discontinuous paths. We state the result without proof:

Theorem (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

$$\begin{aligned}(Lf)(x) = {} & \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x) + c(x) f(x) \\ & + \int_{\mathbb{R}^d \setminus \{x\}} \big(f(y) - f(x) - 1_{\{|y-x|<1\}}(y-x) \cdot \nabla f(x)\big)\, \nu(x, dy)\end{aligned}$$

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 that is derived in the Stochastic Analysis course [9].

3.6 Limits of martingale problems

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).

3.6.1 Weak convergence of stochastic processes

An excellent reference on this subject is the book by Billingsley [2]. 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 $\mathcal{C} = C([0,\infty), S)$ 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 $\mathcal{C}$ is again a polish space w.r.t. the topology of uniform convergence on compact time intervals:

$$x_n \xrightarrow{\mathcal{C}} x \ :\Leftrightarrow\ \forall T \in \mathbb{R}_+ : x_n \to x \ \text{ uniformly on } [0,T].$$

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 (**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*

$$x_n(\lambda_n(t)) \to x(t) \quad \textit{and} \quad \lambda_n(t) \to t \quad \textit{uniformly on } [0,T].$$

It can be shown that the Skorokhod space $\mathcal{D}$ is again a polish space, cf. [2]. Furthermore, the Borel σ-algebras on both $\mathcal{C}$ 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 $\mathcal{C}, \mathcal{D}$ 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 $K \subseteq \mathcal{C}$, $K \subseteq \mathcal{D}$ 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 $\mathcal{C}$ and $\mathcal{D}$. In the case of $\mathcal{C}$ 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 $x \in \mathcal{C}$ 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 a modification of $\omega_{\delta,T}$ by

$$\omega'_{\delta,T}(x) = \inf_{\substack{0=t_0<t_1<\cdots<t_{n-1}<T\le t_n \\ |t_i-t_{i-1}|>\delta}} \max_i \sup_{s,t\in[t_{i-1},t_i)} d(x(s), x(t)).$$

As $\delta \downarrow 0$, $\omega_{\delta,T}(x) \to 0$ for any $x \in \mathcal{C}$ and $T > 0$. 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 3.27 (Arzelà-Ascoli, Skorokhod). *1) A subset $K \subseteq \mathcal{C}$ 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 $K \subseteq \mathcal{D}$ 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 [2] or Ethier/Kurtz [11]. By combining Theorem 3.27 with Prokhorov's Theorem, one obtains:

Corollary 3.28 (Tightness of probability measures on function spaces).

1) A subset $\{P_n : n \in \mathbb{N}\}$ of $\mathcal{P}(\mathcal{C})$ 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

$$\sup_{n \in \mathbb{N}} P_n[X_0 \notin K] \le \varepsilon, \quad \textit{and}$$

(ii) For any $T \in \mathbb{R}_+$,

$$\sup_{n \in \mathbb{N}} P_n[\omega_{\delta,T} > \varepsilon] \to 0 \quad \textit{as } \delta \downarrow 0.$$

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

$$\sup_{n \in \mathbb{N}} P_n[X_t \notin K] \le \varepsilon, \quad \textit{and}$$

(ii) For any $T \in \mathbb{R}_+$,

$$\sup_{n\in\mathbb{N}} P_n[\omega'_{\delta,T} > \varepsilon] \to 0 \quad \textit{as } \delta \downarrow 0.$$

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 [11].

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

$$P_n[\omega_{\delta,T} > \varepsilon] \le \sum_{k=0}^{\lfloor T/\delta \rfloor} P_n\left[\sup_{t\le\delta} d(X_{k\delta+t}, X_{k\delta}) > \varepsilon/3\right]. \tag{3.6.1}$$

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 $X^i,\ i = 1, \ldots, d$. The suprema can then be controlled by applying a semimartingale decomposition and the maximal inequality to the component processes.

3.6.2 From Random Walks to Brownian motion

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

$$X^{(n)}_t := \frac{1}{\sqrt{n}} S_{nt} \quad \text{for } t \in \mathbb{R}_+ \text{ such that } nt \in \mathbb{Z}.$$

In between the partition points $t = k/n,\ k \in \mathbb{Z}_+$, the process $(X^{(n)}_t)$ is defined by linear interpolation so that $X^{(n)}$ has continuous paths.

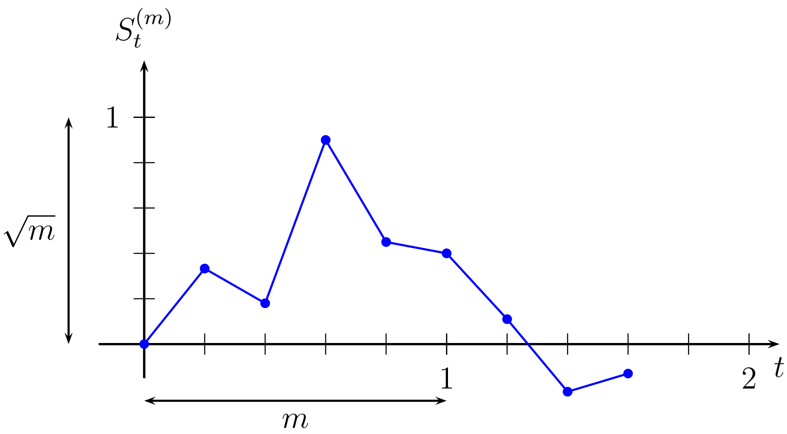


Figure 3.1: Rescaling of a Random Walk.

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 $0 \le t_0 < t_1 < t_2 < \cdots < t_n$,

$$(X_{t_1}^{(n)} - X_{t_0}^{(n)}, X_{t_2}^{(n)} - X_{t_1}^{(n)}, \ldots, X_{t_n}^{(n)} - X_{t_{n-1}}^{(n)}) \xrightarrow{\mathcal{D}} \bigotimes_{i=1}^{k} N(0, t_i - t_{i-1}). \tag{3.6.2}$$

This shows 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 $\mathcal{C}$, we can prove that not only marginals but the whole processes converge in distribution to a Brownian motion:

Theorem 3.29 (Invariance principle, functional central limit theorem). *Let P_n denote the law of the rescaled random walk $X^{(n)}$ on $\mathcal{C} = C([0,\infty), \mathbb{R})$. 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 (3.6.2), 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 $\mathcal{C}$. 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 (3.6.1) and time-homogeneity,

$$\begin{aligned}
P_n[\omega_{\delta,T} > \varepsilon] &\le \left(\left\lfloor \frac{T}{\delta} \right\rfloor + 1 \right) \cdot P\left[\sup_{t \le \delta} \left| X_t^{(n)} - X_t^{(0)} \right| \ge \frac{\varepsilon}{3} \right] \\
&\le \left(\left\lfloor \frac{T}{\delta} \right\rfloor + 1 \right) \cdot P\left[\max_{k \le \lceil n\delta \rceil} |S_k| \ge \frac{\varepsilon}{3}\sqrt{n} \right]
\end{aligned}$$

for any $\varepsilon, \delta > 0, T \in \mathbb{R}_+$ and $n \in \mathbb{N}$. By Corollary 3.28, tightness holds if the probability on the right hand side is of order $\mathrm{o}(S)$ uniformly in n, i.e., if

$$\limsup_{n\to\infty} P\left[\max_{k\leq m} |S_k| \geq \frac{\varepsilon}{3}\frac{\sqrt{m}}{\sqrt{\delta}}\right] = \mathrm{o}(\delta). \tag{3.6.3}$$

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. [8]. For general random walks one can show with some additional arguments that (3.6.3) also holds, see e.g. Billingsley [2]. □

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 $\left(X^{(n)}_{k/n}\right)_{k\in\mathbb{Z}_+}$ 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.

3.6.3 Regularity and tightness for solutions of martingale problems

We will now extend the martingale argument for proving Donsker's Theorem that has been sketched above to limits of general martingale problems on the space $\mathcal{C} = C([0, \infty), S)$ 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 3.33 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

$$(t, x) \mapsto (\mathcal{L}_t f)(x) \quad \text{is a function in } \mathcal{L}^2([0, T] \times \mathcal{C}, \lambda \otimes P)$$

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** $\mathrm{MP}(\mathcal{L}_\mathbf{t}, \mathcal{A})$ w.r.t. a probability measure P on $\mathcal{C}$ 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 $0 \leq s \leq t$,

$$f(X_t) - f(X_s) = M_t^f - M_s^f + \int_s^t \mathcal{L}_r f \, dr. \tag{3.6.4}$$

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 3.30 (Angle-bracket process for solutions of martingale problems). *Let $f, g \in \mathcal{A}$ such that $f \cdot g \in \mathcal{A}$. Then*

$$M_t^f \cdot M_t^g = N_t^{f,g} + \int_0^t \Gamma_r(f,g)\,dr \quad \text{for any } t \geq 0,$$

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 $(g_{ij}) = (a_{ij})^{-1}$:

$$\Gamma(f,f) = \|\operatorname{grad}_g f\|_g^2.$$

Proof of Theorem 3.30. 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 $0 \leq s \leq t$ and $f \in \mathcal{A}$,

$$(M_t^f)^2 - (M_s^f)^2 \sim \int_s^t \Gamma_r(f,f)\,dr.$$

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

$$\begin{aligned}I &:= f(X_t)^2 - f(X_s)^2 \sim_s \int_s^t \mathcal{L}_r f^2\,dr, \\ II &:= -2f(X_s)\left(f(X_t) - f(X_s) - \int_s^t \mathcal{L}_r f\,dr\right) \sim_s 0, \\ III &:= -2f(X_t)\int_s^t \mathcal{L}_r f\,dr = -2\int_s^t f(X_t)\mathcal{L}_r f\,dr, \quad \text{and} \\ IV &:= \left(\int_s^t \mathcal{L}_r f\,dr\right)^2 = 2\int_s^t \int_r^t \mathcal{L}_r f\,\mathcal{L}_u f\,dudr.\end{aligned}$$

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

$$III \sim_s -2\int_s^t f(X_r)\mathcal{L}_r f\,dr - 2\int_s^t \int_r^t \mathcal{L}_r f\,\mathcal{L}_u f\,dudr.$$

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

$$\omega_{\delta,T}^f := \omega_{\delta,T}(f \circ X), \qquad V_{s,t}^f := \sup_{r \in [s,t]} |f(X_r) - f(X_s)|.$$

Lemma 3.31 (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 $0 \le s \le t$,

$$\|V_{s,t}^f\|_{L^p(P)} \le C_p(t-s)^{1/2} \sup_{r \in [s,t]} \|\Gamma_r(f,f)\|_{L^{p/2}(P)}^{1/2} + (t-s) \sup_{r \in [s,t]} \|\mathcal{L}_r f\|_{L^p(P)}.$$

2) *For any* $\delta, \varepsilon, T \in (0, \infty)$,

$$P\left[\omega^f_{\delta,T} \geq \varepsilon\right] \leq \widetilde{C}_p\, \varepsilon^{-p}\left(1 + \left\lfloor \frac{T}{\delta} \right\rfloor\right) \cdot \left(\delta^{p/2} \sup_{r \leq T} \|\Gamma_r(f,f)\|^{p/2}_{L^{p/2}(P)} + \delta^p \sup_{r \leq t} \|\mathcal{L}_r f\|_{L^p(P)}\right).$$

Proof. 1) By (3.6.4),

$$V^f_{s,t} \leq \sup_{r \in [s,t]} |M^f_r - M^f_s| + \int_s^t |\mathcal{L}_u f|\, du.$$

Since $f(X_t)$ is continuous, M^f is a continuous martingale. Therefore, by **Burkholder's inequality**,

$$\begin{aligned}
&\Big\| \sup_{r \in [s,t]} \big| M^f_r - M^f_s \big| \Big\|_{L^p(P)} \leq C_p \big\| \langle M^f \rangle_t - \langle M^f \rangle_s \big\|^{1/2}_{L^{p/2}(P)} \\
&= C_p \left\| \int_s^t \Gamma_r(f,f)\, dr \right\|^{1/2}_{L^{p/2}(P)} \\
&\leq C_p (t-s)^{1/2} \sup_{r \in [s,t]} \|\Gamma_r(f,f)\|^{1/2}_{L^{p/2}(P)}.
\end{aligned}$$

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. [9].

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

$$\begin{aligned}
P\left[\omega^f_{\delta,T} \geq \varepsilon\right] &\leq \sum_{k=0}^{\lfloor T/\delta \rfloor} P\left[V^f_{k\delta,(k+1)\delta} \geq \varepsilon/3\right] \\
&\leq \sum_{k=0}^{\lfloor T/\delta \rfloor} \left(\frac{3}{\varepsilon}\right)^p \left\| V^f_{k\delta,(k+1)\delta} \right\|^p_{L^p(P)}.
\end{aligned}$$

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

$$\|\mathcal{L}_t f\|_{L^p(P)} = \|\mathcal{L}f\|_{L^p(\mu)}, \quad \|\Gamma_t(f,f)\|_{L^{p/2}(P)} = \|\Gamma(f,f)\|_{L^{p/2}(\mu)} \quad \text{for any } t \geq 0.$$

3.6.4 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 3.32 (Existence of diffusions in $\mathbb{R}^{\mathbf{d}}$). *For $1 \leq i,j \leq d$ 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 $\lfloor t \rfloor_n := \max\left\{s \in \frac{1}{n}\mathbb{Z} : s \leq t\right\}$, 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 $i = 1, \ldots, d$ let $f_i(x) := x_i$. Since $|x - y| \leq \sum_{i=1}^d |f_i(x) - f_i(y)|$ for any $x, y \in \mathbb{R}^d$, we have

$$\omega_{\delta,T} \leq \sum_{i=1}^{d} \omega_{\delta,T}^{f_i} \qquad \text{for any } \delta, T \in (0, \infty).$$

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 $\varepsilon, T \in (0, \infty)$,

$$P^{(n)}\left[\omega_{\delta,T} \geq \varepsilon\right] \leq \sum_{i=1}^{d} P^{(n)}\left[\omega_{\delta,T}^{f_i} \geq \varepsilon/d\right] \to 0$$

uniformly in n as $\delta \downarrow 0$ by Lemma 3.31.
Hence by Theorem 3.29, 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,

$$E^{(n)}\left[\left(f(X_t) - f(X_s) - \int_s^t \mathcal{L}_r^{(n)} f \, dr\right) g(X_{s_1}, \ldots, X_{s_k})\right] = 0$$

for any $0 \leq s_1 < s_2 < \cdots < s_k \leq s \leq t$ 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 3.32 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 [31] for a sketch of the proof.

3.6.5 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 3.33. *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*

$$P^{(n)}\left[\exists\, t \in [0, T] : X_t \notin K\right] \leq \gamma \quad \text{for any } n \in \mathbb{N}.$$

(ii) ***Uniform* L^p *bound:*** *There exists $p > 2$ such that for any $T \in \mathbb{R}_+$,*

$$\sup_{n \in \mathbb{N}} \sup_{t \leq T} \left(\left\| \Gamma_t^{(n)}(f, f) \right\|_{L^{p/2}(P^{(n)})} + \left\| \mathcal{L}_t^{(n)} f \right\|_{L^p(P^{(n)})} \right) < \infty.$$

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:***

$$\mathcal{L}_t^{(n)} f \to \mathcal{L}_t f \qquad \textit{uniformly for any } f \in \mathcal{A},$$

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 [11].

Chapter 4

Convergence to equilibrium

Additional references for this chapter: Royer [33], Malrieu [21], Bakry, Gentil, Ledoux [1]

Our goal in the following sections is to relate the long time asymptotics $(t \uparrow \infty)$ of a time-homogeneous Markov process (respectively its transition semigroup) to its infinitesimal characteristics which describe the short-time behavior $(t \downarrow 0)$:

Asymptotic properties	$\leftrightarrow$	Infinitesimal behavior, generator
$t \uparrow \infty$		$t \downarrow 0$

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. On the other hand, we would like to take into account processes that jump instantaneously (as e.g. interacting particle systems on $\mathbb{Z}^d$) or have continuous trajectories (diffusion-processes). In this case it is not straightforward to describe the process completely in terms of infinitesimal characteristics, as we did for jump processes. A convenient general setup that can be applied to all these types of Markov processes is the martingale problem of Stroock and Varadhan.

Let S be a Polish space endowed with its Borel σ-algebra $\mathcal{S}$. By $\mathcal{F}_b(S)$ we denote the linear space of all bounded measurable functions $f\colon S \to \mathbb{R}$. Suppose that $\mathcal{A}$ is a linear subspace of $\mathcal{F}_b(S)$ such that

(A0) If μ is a signed measure on S with finite variation and

$$\int f \, d\mu = 0 \quad \forall f \in \mathcal{A},$$

then $\mu = 0$

Let

$$\mathcal{L}\colon \mathcal{A} \subseteq \mathcal{F}_b(S) \to \mathcal{F}_b(S)$$

be a linear operator.

From now on 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 $P_x[X_0 = x] = 1$.

Let $\bar{\mathcal{A}}$ denote the closure of $\mathcal{A}$ with respect to the supremum norm. For most results derived below, we will impose two additional assumptions:

Assumptions:

(A1) If $f \in \mathcal{A}$, then $\mathcal{L}f \in \bar{\mathcal{A}}$.

(A2) There exists a linear subspace $\mathcal{A}_0 \subseteq \mathcal{A}$ such that if $f \in \mathcal{A}_0$, then $p_t f \in \mathcal{A}$ for all $t \geq 0$, and $\mathcal{A}_0$ is dense in $\mathcal{A}$ with respect to the supremum norm.

Example. (1). For a diffusion process in $\mathbb{R}^d$ with continuous non-degenerated coefficients satisfying an appropriate growth constraint at infinity, (A1) and (A2) hold with $\mathcal{A}_0 = C_0^\infty(\mathbb{R}^d)$, $\mathcal{A} = \mathcal{S}(\mathbb{R}^d)$ and $B = \bar{\mathcal{A}} = C_\infty(\mathbb{R}^d)$.

(2). In general, it can be difficult to determine explicitly a space $\mathcal{A}_0$ such that (A2) holds. In this case, a common procedure is to approximate the Markov process and its transition semigroup by more regular processes (e.g. non-degenerate diffusions in $\mathbb{R}^d$), and to derive asymptotic properties from corresponding properties of the approximands.

(3). For an interacting particle system on $T^{\mathbb{Z}^d}$ with bounded transition rates $c_i(x, \eta)$, the conditions (A1) and (A2) hold with

$$\mathcal{A}_0 = \mathcal{A} = \left\{ f\colon T^{\mathbb{Z}^d} \to \mathbb{R} \ : \ |||f||| < \infty \right\}$$

where

$$|||f||| = \sum_{x\in\mathbb{Z}^d} \Delta_f(x), \qquad \Delta_f(x) = \sup_{i\in T} \left| f(\eta^{x,i}) - f(\eta) \right|,$$

cf. Liggett [18].

Theorem (From the martingale problem to the Kolmogorov equations).
Suppose (A1) and (A2) hold. Then $(p_t)_{t\geq 0}$ induces a C_0 contraction semigroup $(P_t)_{t\geq 0}$ on the Banach space $B = \bar{\mathcal{A}} = \bar{\mathcal{A}}_0$, and the generator is an extension of $(\mathcal{L}, \mathcal{A})$. In particular, the forward and backward equations

$$\frac{d}{dt} p_t f = p_t \mathcal{L} f \quad \forall\, f \in \mathcal{A}$$

and

$$\frac{d}{dt} p_t f = \mathcal{L} p_t f \quad \forall\, f \in \mathcal{A}_0$$

hold.

Proof. Since M_t^f is a bounded martingale with respect to P_x, we obtain the integrated forward equation by Fubini:

$$\begin{aligned}(p_t f)(x) - f(x) = E_x[f(X_t) - f(X_0)] &= E_x\left[\int_0^t (\mathcal{L}f)(X_s)\, ds\right] \\ &= \int_0^t (p_s \mathcal{L} f)(x)\, ds\end{aligned} \tag{4.0.1}$$

for all $f \in \mathcal{A}$ and $x \in S$. In particular,

$$\|p_t f - f\|_{\sup} \leq \int_0^t \|p_s \mathcal{L} f\|_{\sup}\, ds \leq t \cdot \|\mathcal{L} f\|_{\sup} \to 0$$

as $t \downarrow 0$ for any $f \in \mathcal{A}$. This implies strong continuity on $B = \bar{\mathcal{A}}$ since each p_t is a contraction with respect to the sup-norm. Hence by (A1) and (4.0.1),

$$\frac{p_t f - f}{t} - \mathcal{L} f = \frac{1}{t} \int_0^t (p_s \mathcal{L} f - \mathcal{L} f)\; ds \to 0$$

uniformly for all $f \in \mathcal{A}$, i.e. $\mathcal{A}$ is contained in the domain of the generator L of the semigroup $(P_t)_{t\geq 0}$ induced on B, and $Lf = \mathcal{L} f$ for all $f \in \mathcal{A}$. Now the forward and the backward equations follow from the corresponding equations for $(P_t)_{t\geq 0}$ and Assumption (A2). $\square$

4.1 Stationary distributions and reversibility

4.1.1 Stationary distributions

Theorem 4.1 (Infinitesimal characterization of stationary distributions). *Suppose (A1) and (A2) hold. Then for $\mu \in M_1(S)$ the following assertions are equivalent:*

(i) The process (X_t, P_μ) is stationary, i.e.

$$(X_{s+t})_{t\geq 0} \sim (X_t)_{t\geq 0}$$

with respect to P_μ for all $s \geq 0$.

(ii) μ is a stationary distribution for $(p_t)_{t\geq 0}$

(iii) $\int \mathcal{L}f\, d\mu = 0 \quad \forall\, f \in \mathcal{A}$ (i.e. μ is infinitesimally invariant, $\mathcal{L}^*\mu = 0$).

Proof. (i)⇒(ii) If (i) holds then in particular

$$\mu p_s = P_\mu \circ X_s^{-1} = P_\mu \circ X_0^{-1} = \mu$$

for all $s \geq 0$, i.e. μ is a stationary initial distribution.

(ii)⇒(i) By the Markov property, for any measurable subset $B \subseteq \mathcal{D}(\mathbb{R}^+, S)$,

$$P_\mu[(X_{s+t})_{t\geq 0} \in B \mid \mathcal{F}_s] = P_{X_s}[(X_t)_{t\geq 0} \in B] \quad P_\mu\text{-a.s., and thus}$$

$$P_\mu[(X_{s+t})_{t\geq 0} \in B] = E_\mu[P_{X_s}((X_t)_{t\geq 0} \in B)] = P_{\mu p_s}[(X_t)_{t\geq 0} \in B] = P_\mu[X \in B]$$

(ii)⇒(iii) By the theorem above, for $f \in \mathcal{A}$,

$$\frac{p_t f - f}{t} \to \mathcal{L}f \text{ uniformly as } t \downarrow 0,$$

so

$$\int \mathcal{L}f\, d\mu = \lim_{t\downarrow 0} \frac{\int (p_t f - f)\, d\mu}{t} = \lim_{t\downarrow 0} \frac{\int f\, d(\mu p_t) - \int f\, d\mu}{t} = 0$$

provided μ is stationary with respect to $(p_t)_{t\geq 0}$.

(iii)⇒(ii) By the backward equation and (iii),

$$\frac{d}{dt}\int p_t f\, d\mu = \int \mathcal{L} p_t f\, d\mu = 0$$

since $p_t f \in \mathcal{A}$ for $f \in \mathcal{A}_0$ and hence

$$\int f\, d(\mu p_t) = \int p_t f\, d\mu = \int f\, d\mu \tag{4.1.1}$$

for all $f \in \mathcal{A}_0$ and $t \geq 0$. Since $\mathcal{A}_0$ is dense in $\mathcal{A}$ with respect to the supremum norm, (4.1.1) extends to all $f \in \mathcal{A}$. Hence $\mu p_t = \mu$ for all $t \geq 0$ by (A0).

□

Remark. Assumption (A2) is required only for the implication (iii)⇒(ii).

Applicaton to Itô diffusions:

Suppose that we are given non-explosive weak solutions $(X_t, P_x), x \in \mathbb{R}^d$, of the 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 in $\mathbb{R}^d$, 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. Then by Itô's formula (X_t, P_x) solves the martingale problem for the operator

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x)\cdot\nabla, \quad a = \sigma\sigma^T,$$

with domain $\mathcal{A} = C_0^\infty(\mathbb{R}^n)$. 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/Williams [32]. Therefore by the remark above, (X_t, P_x) is a Markov process.

Theorem 4.2. *Suppose μ is a stationary distribution of (X_t, P_x) that has a smooth density ϱ with respect to the Lebesgue measure. Then*

$$\mathcal{L}^*\varrho := \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}\varrho) - \operatorname{div}(b\varrho) = 0$$

Proof. Since μ is a stationary distribution,

$$0 = \int \mathcal{L}f\,d\mu = \int_{\mathbb{R}^n} \mathcal{L}f\,\varrho\,dx = \int_{\mathbb{R}^n} f\mathcal{L}^*\varrho\,dx \quad \forall\, f \in C_0^\infty(\mathbb{R}^n) \tag{4.1.2}$$

Here the last equation follows by integration by parts, because f has compact support. □

Remark. In general, μ is a distributional solution of $\mathcal{L}^*\mu = 0$.

Example (One-dimensional diffusions). In the one-dimensional case,

$$\mathcal{L}f = \frac{a}{2}f'' + bf',$$

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:

$$\begin{aligned}\mathcal{L}f = \frac{a}{2}f'' + bf' = 0 \quad &\Leftrightarrow \quad f' = C_1 \exp - \int_0^{\bullet} \frac{2b}{a}\,dx, \quad C_1 \in \mathbb{R}\\ &\Leftrightarrow \quad f = C_2 + C_1 \cdot s, \quad C_1, C_2 \in \mathbb{R}\end{aligned}$$

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:

(1). **Brownian motion:** $a \equiv 1, b \equiv 0, s(y) = y$. There is no stationary distribution. Lebesgue measure is an infinite stationary measure.

(2). **Ornstein-Uhlenbeck process:**

$$\begin{aligned}
dX_t &= dB_t - \gamma X_t\,dt, & \gamma &> 0, \\
\mathcal{L} &= \frac{1}{2}\frac{d^2}{dx^2} - \gamma x \frac{d}{dx}, & a &\equiv 1,
\end{aligned}$$

$$b(x) = -\gamma x, \qquad s(y) = \int_0^y e^{\int_0^y 2\gamma x\,dx}\,dy = \int_0^y e^{\gamma y^2}\,dy \text{ recurrent,}$$

$$m(dy) = e^{-\gamma y^2}\,dy, \qquad \mu = \frac{m}{m(\mathbb{R})} = N\left(0, \frac{2}{\gamma}\right) \text{ is the unique stationary distribution}$$

(3).

$$dX_t = dB_t + b(X_t)\,dt, \qquad b \in C^2, \qquad b(x) = \frac{1}{x} \text{ for } |x| \geq 1$$

transient, two independent non-negative solutions of $\mathcal{L}^*\varrho = 0$ with $\int \varrho\,dx = \infty$.

(**Exercise:** stationary distributions for $dX_t = dB_t - \frac{\gamma}{1+|X_t|}\,dt$)

Example (Deterministic diffusions).

$$\begin{aligned} dX_t &= b(X_t)\,dt, && b \in C^2(\mathcal{R}^n) \\ \mathcal{L}f &= b \cdot \nabla f \\ \mathcal{L}^*\varrho &= -\operatorname{div}(\varrho b) = -\varrho \operatorname{div} b - b \cdot \nabla \varrho, && \varrho \in C^1 \end{aligned}$$

Lemma 4.3.

$$\begin{aligned} \mathcal{L}^*\varrho = 0 \quad &\Leftrightarrow \quad \operatorname{div}(\varrho b) = 0 \\ &\Leftrightarrow \quad (\mathcal{L}, C_0^\infty(\mathbb{R}^n)) \text{ anti-symmetric on } L^2(\mu) \end{aligned}$$

Proof. First equivalence: cf. above

Second equivalence:

$$\begin{aligned} \int f\mathcal{L}g\,d\mu = \int fb \cdot \nabla g \varrho\,dx &= -\int \operatorname{div}(fb\varrho)g\,dx \\ &= -\int \mathcal{L}fg\,d\mu - \int \operatorname{div}(\varrho b)fg\,dx \qquad \forall\, f,g \in C_0^\infty \end{aligned}$$

Hence $\mathcal{L}$ is anti-symmetric if and only if $\operatorname{div}(\varrho b) = 0$

□

4.1.2 Reversibility

Theorem 4.4. *Suppose (A1) and (A2) hold. Then for $\mu \in M_1(S)$ the following assertions are equivalent:*

*(i) The process (X_t, P_μ) is **invariant with respect to time reversal**, i.e.,*

$$(X_s)_{0 \leq s \leq t} \sim (X_{t-s})_{0 \leq s \leq t} \quad \textit{with respect to } P_\mu\ \forall\, t \geq 0$$

(ii)

$$\mu(dx)p_t(x, dy) = \mu(dy)p_t(y, dx) \quad \forall\, t \geq 0$$

(iii) p_t *is* μ*-symmetric, i.e.,*

$$\int f p_t g \, d\mu = \int p_t f g \, d\mu \quad \forall\, f, g \in \mathcal{F}_b(S)$$

(iv) $(\mathcal{L}, \mathcal{A})$ *is* μ*-symmetric, i.e.,*

$$\int f \mathcal{L} g \, d\mu = \int \mathcal{L} f g \, d\mu \quad \forall\, f, g \in \mathcal{A}$$

Remark. (1). A reversible process (X_t, P_μ) is stationary, since for all $s, u \geq 0$,

$$(X_{s+t})_{0 \leq t \leq u} \sim (X_{u-t})_{0 \leq t \leq u} \sim (X_t)_{0 \leq t \leq u} \quad \text{with respect to } P_\mu$$

(2). Similarly (ii) implies that μ is a stationary distribution:

$$\int \mu(dx)p_t(x, dy) = \int p_t(y, dx)\mu(dy) = \mu(dy)$$

Proof of the Theorem. (i)⇒(ii):

$$\mu(dx)p_t(x, dy) = P_\mu \circ (X_0, X_t)^{-1} = P_\mu \circ (X_t, X_0)^{-1} = \mu(dy)p_t(y, dx)$$

(ii)⇒(i): By induction, (ii) implies

$$\begin{aligned} &\mu(dx_0)p_{t_1-t_0}(x_0, dx_1)p_{t_2-t_1}(x_1, dx_2) \cdots p_{t_n-t_{n-1}}(x_{n-1}, dx_n) \\ =&\mu(dx_n)p_{t_1-t_0}(x_n, dx_{n-1}) \cdots p_{t_n-t_{n-1}}(x_1, dx_0) \end{aligned}$$

for $n \in \mathbb{N}$ and $0 = t_0 \leq t_1 \leq \cdots \leq t_n = t$, and thus

$$E_\mu[f(X_0, X_{t_1}, X_{t_2}, \ldots, X_{t_{n-1}}, X_t)] = E_\mu[f(X_t, \ldots, X_{t_1}, X_0)]$$

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,

$$\int f p_t g \, d\mu = \iint f(x)g(y)\mu(dx)p_t(x, dy)$$

is symmetric for all $f, g \in \mathcal{F}_b(S)$ if and only if $\mu \otimes p_t$ is a symmetric measure on $S \times S$.

(iii)⇔(iv): Exercise.

□

4.1.3 Application to diffusions in $\mathbb{R}^n$

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

μ probability measure on $\mathbb{R}^n$ (more generally locally finite positive measure)

Question: For which process is μ stationary?

Theorem 4.5. *Suppose $\mu = \varrho\, dx$ with $\varrho_i a_{ij} \in C^1, b \in C, \varrho > 0$. Then*

(1). We have

$$\mathcal{L}g = \mathcal{L}_s g + \mathcal{L}_a g$$

for all $g \in C_0^\infty(\mathbb{R}^n)$ where

$$\mathcal{L}_s g = \frac{1}{2}\sum_{i,j=1}^{n}\frac{1}{\varrho}\frac{\partial}{\partial x_i}\left(\varrho a_{ij}\frac{\partial g}{\partial x_i}\right)$$

$$\mathcal{L}_a g = \beta\cdot\nabla g, \quad \beta_j = b_j - \sum_i \frac{1}{2\varrho}\frac{\partial}{\partial x_i}(\varrho a_{ij})$$

(2). The operator $(\mathcal{L}_s, C_0^\infty)$ is symmetric with respect to μ.

(3). The following assertions are equivalent:

(i) $\mathcal{L}^*\mu = 0$ *(i.e. $\int \mathcal{L}f\, d\mu = 0$ for all $f \in C_0^\infty$).*

(ii) $\mathcal{L}_a^*\mu = 0$

(iii) $\mathrm{div}(\varrho\beta) = 0$

(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 $\operatorname{div}(\varrho\beta) = 0$.

Remark. Probabilistic proof of reversibility for $b := -\frac{1}{2}\nabla H,\ H \in C^1$:

$$X_t = x + B_t + \int_0^t b(X_s)\,ds, \quad \text{non-explosive}, \quad b = -\frac{1}{2}\nabla h$$

Hence $P_\mu \circ X_{0:T}^{-1} \ll P_\lambda^{\mathrm{BM}}$ with density

$$\exp\left(-\frac{1}{2}H(B_0) - \frac{1}{2}H(B_T) - \int_0^T \left(\frac{1}{8}|\nabla H|^2 - \frac{1}{4}\Delta H\right)(B_s)\,ds\right)$$

which shows that (X_t, P_μ) is reversible.

4.2 Poincaré inequalities and convergence to equilibrium

Suppose now that μ is a stationary distribution 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 the stationarity of μ. As before, we assume that we are given a Markov process with transition semigroup $(p_t)_{t\geq 0}$ solving the martingale problem for the operator $(\mathcal{L}, \mathcal{A})$. The assumptions on $\mathcal{A}_0$ and $\mathcal{A}$ can be relaxed in the following way:

(A0) as above

(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 $g \in \mathcal{L}^q(S,\mu)$, $\frac{1}{q} + \frac{1}{q} = 1$, 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 $L^p(S,\mu)$.

4.2.1 Decay of variances and correlations

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. *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

$$\mathcal{E}(f, g) = -(f, L^{(2)} g)_\mu = -\frac{d}{dt}(f, p_t g)_\mu \Big|_{t=0}$$

Theorem 4.6. *For all* $f \in \mathcal{A}_0$ *and* $t \geq 0$

$$\frac{d}{dt} \mathrm{Var}_\mu(p_t f) = \frac{d}{dt} \int (p_t f)^2 \, d\mu = -2\mathcal{E}(p_t f, p_t f) = -2\mathcal{E}_s(p_t f, p_t f)$$

Remark. (1). In particular,

$$\mathcal{E}(f, f) = -\frac{1}{2} \int (p_t f)^2 \, d\mu = -\frac{1}{2}\frac{d}{dt} \mathrm{Var}_\mu(p_t f)\Big|_{t=0},$$

infinitesimal change of variance

(2). The assertion extends to all $f \in \mathrm{Dom}(L^{(2)})$ 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,

$$\frac{d}{dt} \operatorname{Var}_\mu(p_t f) = \frac{d}{dt} \int (p_t f)^2 \, d\mu$$

□

Remark. (1). 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.

(2). 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 4.7 (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****:*

$$\operatorname{Var}_\mu(p_t f) \leq e^{-2\lambda t} \operatorname{Var}_\mu(f) \quad \forall\, f \in L^2(S,\mu) \tag{4.2.1}$$

(iii) ***Spectral gap****:*

$$\operatorname{Re}\alpha \geq \lambda \quad \forall\, \alpha \in \operatorname{spec}\left(-L^{(2)}\Big|_{\operatorname{span}\{1\}^\perp}\right)$$

Remark. Optimizing over λ, the corollary says that (4.2.1) holds with

$$\lambda := \inf_{f\in\mathcal{A}} \frac{\mathcal{E}(f,f)}{\operatorname{Var}_\mu(f)} = \inf_{\substack{f\in\mathcal{A} \\ f\perp 1 \text{ in } L^2(\mu)}} \frac{(f, -\mathcal{L}f)_\mu}{(f,f)_\mu}$$

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 (4.2.1) 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

$$-(L^{(2)} f, f)_\mu \geq \lambda \int f^2 \, d\mu \quad \text{for } f \perp 1$$

which is equivalent to (iii).

$(iii) \Rightarrow (i)$ Follows by the equivalence above.

$\square$

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.

Example. (1). **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

$$Q(x,y) = \mu(x)\mathcal{L}_s(x,y) = \frac{1}{2}\left(\mu(x)\mathcal{L}(x,y) + \mu(y)\mathcal{L}(y,x)\right)$$

(2). **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 $\mathcal{A} = C_0^\infty,\ \mu = \varrho\, dx,\ \varrho, a_{ij} \in C^1,\ b \in C\ \varrho \geq 0$,

$$\mathcal{E}_s(f,g) = \frac{1}{2}\int \sum_{i,j=1}^{n} a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\, d\mu$$

$$\mathcal{E}(f,g) = \mathcal{E}_s(f,g) - (f, \beta\cdot\nabla g), \qquad \beta = b - \frac{1}{2\varrho}\operatorname{div}(\varrho a_{ij})$$

4.2.2 Divergences

Definition ("Distances" of probability measures). *$\mu,\ \nu$ probability measures on S, $\mu - \nu$ signed measure.*

(i) ***Total variation distance****:*

$$\|\nu - \mu\|_{TV} = \sup_{A\in\mathcal{S}} |\nu(A) - \mu(A)|$$

(ii) χ^2***-divergence:***

$$\chi^2(\mu|\nu) = \begin{cases} \int \left(\frac{d\mu}{d\nu} - 1\right)^2 d\mu = \int \left(\frac{d\nu}{d\mu}\right)^2 d\mu - 1 & \textit{if } \nu \ll \mu \\ +\infty & \textit{else} \end{cases}$$

(iii) ***Relative entropy (Kullback-Leibler divergence):***

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}\, d\mu = \int \log\frac{d\nu}{d\mu}\, d\nu & \textit{if } \nu \ll \mu \\ +\infty & \textit{else} \end{cases}$$

(where $0\log 0 := 0$).

Remark. By Jensen's inequality,

$$H(\nu|\mu) \geq \int \frac{d\nu}{d\mu}\, d\mu \log \int \frac{d\nu}{d\mu}\, d\mu = 0$$

Lemma 4.8 (Variational characterizations).

(i)

$$\|\nu - \mu\| = \frac{1}{2}\sup_{\substack{f\in\mathcal{F}_b(S)\\ |f|\leq 1}}\left(\int f\, d\nu - \int f\, d\mu\right)$$

(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)

$$H(\nu|\mu) = \sup_{\substack{f\in\mathcal{F}_b(S)\\ \int e^f\,d\mu\le 1}} \int f\,d\nu = \sup_{f\int\mathcal{F}_b(S)} \int f\,d\nu - \log\int e^f\,d\mu$$

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) " $\le$ "

$$\nu(A)-\mu(A) = \frac{1}{2}\left(\nu(A)-\mu(A)+\mu(A^c)-\nu(A^c)\right) = \frac{1}{2}\left(\int f\,d\nu - \int f\,d\mu\right)$$

and setting $f := I_A - I_{A^c}$ leads to

$$\|\nu-\mu\|_{\mathrm{TV}} = \sup_A\left(\nu(A)-\mu(A)\right) \le \frac{1}{2}\sup_{|f|\le 1}\left(\int f\,d\nu - \int f\,d\mu\right)$$

" $\ge$ " If $|f|\le 1$ then

$$\begin{aligned}\int f\,d(\nu-\mu) &= \int_{S_+} f\,d(\nu-\mu) + \int_{S_-} f\,d(\nu-\mu)\\ &\le (\nu-\mu)(S_+) - (\nu-\mu)(S_-)\\ &= 2(\nu-\mu)(S_+) \qquad \text{(since } (\nu-\mu)(S_+)+(\nu-\mu)(S_-) = (\nu-\mu)(S)=0)\\ &\le 2\|\nu-\mu\|_{\mathrm{TV}}\end{aligned}$$

where $S = S_+\dot{\bigcup} S_-$, $\nu-\mu\ge 0$ on S_+, $\nu-\mu\le 0$ 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,

$$\begin{aligned} \int f\, d\nu &= \int f \varrho\, d\mu \\ &\leq \int \varrho \log \varrho\, d\mu - \int \varrho\, d\mu + \int e^f\, d\mu \\ &= H(\nu|\mu) - 1 + \int e^f\, d\mu \qquad \forall\, f \in \mathcal{F}_b(S) \\ &\leq H(\nu|\mu) \qquad \text{if } \int e^f\, d\mu \leq 1 \end{aligned}$$

" $\leq$ " $\nu \ll \mu$ 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 [6].

$\square$

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 $\nu \not\ll \mu$.

4.2.3 Decay of χ^2 divergence

Corollary 4.9. *The assertions $(i) - (iii)$ in the corollary above are also equivalent to*

(iv) ***Exponential decay of χ^2 distance to equilibrium:***

$$\chi^2(\nu p_t | \mu) \leq e^{-2\lambda t} \chi^2(\nu | \mu) \quad \forall\, \nu \in M_1(S)$$

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

$$\|p_t f\|_{L^2(\mu)} \leq e^{-\lambda t} \|f\|_{L^2(\mu)}$$

$\square$

Example: $d = 1$!

Example (**Gradient type diffusions in $\mathbb{R}^n$**).

$$dX_t = dB_t + b(X_t)\,dt, \qquad\qquad b \in C(\mathbb{R}^n, \mathbb{R}^n)$$

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:

$$\mathrm{Var}_{\varrho\,dx}(f) \le \frac{1}{2\lambda} \cdot \int |\nabla f|^2 \varrho\,dx$$

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. $b(x) = -\alpha x, \varrho(x) =$ const. $e^{-\alpha x^2}, \mu =$ Gauss measure.

Bounds on the variation norm:

Lemma 4.10. *(i)*

$$\|\nu - \mu\|_{TV}^2 \le \frac{1}{4}\chi^2(\nu|\mu)$$

(ii) Pinsker's inequality:

$$\|\nu - \mu\|_{TV}^2 \le \frac{1}{2}H(\nu|\mu) \quad \forall\, \mu, \nu \in M_1(S)$$

Proof. If $\nu \not\ll \mu$, then $H(\nu|\mu) = \chi^2(\nu|\mu) = \infty$.
Now let $\nu \ll \mu$:

(i)

$$\|\nu - \mu\|_{\mathrm{TV}} = \frac{1}{2}\|\varrho - 1\|_{L^1(\mu)} \le \frac{1}{2}\|\varrho - 1\|_{L^2(\mu)} = \frac{1}{2}\chi^2(\nu|\mu)^{\frac{1}{2}}$$

(ii) We have the inequality

$$3(x-1)^2 \le (4+2x)(x\log x - x + 1) \quad \forall\, x \ge 0$$

and hence

$$\sqrt{3}|x-1| \leq (4+2x)^{\frac{1}{2}}(x\log x - x + 1)^{\frac{1}{2}}$$

and with the Cauchy Schwarz inequality

$$\begin{aligned}\sqrt{3}\int |\varrho - 1|\, d\mu &\leq \left(\int (4+2\varrho)\, d\mu\right)^{\frac{1}{2}} \left(\int (\varrho\log\varrho - \varrho + 1)\, d\mu\right)^{\frac{1}{2}} \\ &= \sqrt{6}\cdot H(\nu|\mu)^{\frac{1}{2}}\end{aligned}$$

□

Remark. If S is finite and $\mu(x) > 0$ for all $x \in S$ then conversely

$$\begin{aligned}\chi^2(\nu|\mu) = \sum_{x\in S}\left(\frac{\nu(x)}{\mu(x)} - 1\right)^2 \mu(x) &\leq \frac{\left(\sum_{x\in S}\left|\frac{\nu(x)}{\mu(x)} - 1\right|\mu(x)\right)^2}{\min_{x\in S}\mu(x)} \\ &= \frac{4\|\nu-\mu\|_{\mathrm{TV}}^2}{\min\mu}\end{aligned}$$

Corollary 4.11. *(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

$$\|\nu p_t - \mu\|_{TV} \leq \frac{1}{2}e^{-\lambda t}\chi^2(\nu|\mu)^{\frac{1}{2}} \tag{4.2.2}$$

(ii) In particular, if S is finite then

$$\|\nu p_t - \mu\|_{TV} \leq \frac{1}{\min_{x\in S}\mu(x)^{\frac{1}{2}}}e^{-\lambda t}\|\nu - \mu\|_{TV}$$

where $\|\nu - \mu\|_{TV} \leq 1$. This leads to a bound for the ***Dobrushin coefficient*** *(contraction coefficient with respect to $\|\cdot\|_{TV}$).*

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: $\varepsilon \in (0,1)$,

$$\begin{aligned} T_{\text{mix}}(\varepsilon) &= \inf \{t \geq 0 \; : \; \|\nu p_t - \mu\|_{\text{TV}} \leq \varepsilon \text{ for all } \nu \in M_1(S)\} \\ &\leq \frac{1}{\lambda} \log \frac{1}{\varepsilon} + \frac{1}{2\lambda} \log \frac{1}{\min \mu(x)} \end{aligned}$$

where the first summand is the $\mathbf{L^2}$ **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 (4.2.2) is often problematic since $\chi^2(\nu|\mu)$ can be very large (whereas $\|\nu - \mu\|_{\text{TV}} \leq 1$). 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?

$$\int p_t f \, d\nu - \int f \, d\mu = \int p_t f \, d(\nu - \mu) \leq \|p_t f\|_{\sup} \cdot \|\nu - \mu\|_{\text{TV}}$$

Analysis: From the Sobolev inequality follows

$$\|p_t f\|_{\sup} \leq c \cdot \|f\|_{L^p}$$

However, Sobolev constants are dimension dependent! This leads to a replacement by the log Sobolev inequality.

4.3 Central Limit theorem for Markov processes

When are stationary Markov processes ergodic?
Let $(L, \text{Dom}(L))$ denote the generator of $(p_t)_{t\geq 0}$ on $L^2(\mu)$.

Theorem 4.12. *The following assertions are equivalent:*

(i) P_μ *is ergodic*

(ii) $\ker L = \text{span}\{1\}$*, i.e.*

$$h \in \mathcal{L}^2(\mu) \textit{harmonic} \quad \Rightarrow \quad h = \textit{const. } \mu\textit{-a.s.}$$

(iii) p_t *is* μ*-irreducible, i.e.*

$$B \in \mathcal{S} \text{ such that } p_t I_B = I_B \quad \mu\text{-a.s.}\ \forall\, t \geq 0 \quad \Rightarrow \quad \mu(B) \in \{0, 1\}$$

If reversibility holds then (i)-(iii) are also equivalent to:

(iv) p_t *is* $L^2(\mu)$*-ergodic, i.e.*

$$\left\| p_t f - \int f \, d\mu \right\|_{L^2(\mu)} \to 0 \quad \forall\, f \in L^2(\mu)$$

Let $(M_t)_{t\geq 0}$ be a continuous square-integrable $(\mathcal{F}_t)$ martingale and $\mathcal{F}_t$ 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 [14]).

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

$$\Gamma(f,g) = L^{(1)}(f\cdot g) - fL^2 g - gL^{(2)}f \in L^1(\mu)$$

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

$$\Gamma(f,g)(x) = \sum_{i,j} a_{ij}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial g}{\partial x_j}(x) = \left|\sigma^T(x)\nabla f(x)\right|^2_{\mathcal{R}^n}$$

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:

$$\mathcal{E}(f,f) = -\int fL^{(2)}f\,d\mu + \underbrace{\left(\frac{1}{2}\int L^{(1)}f^2\,d\mu\right)}_{=0} = \frac{1}{2}\int \Gamma(f,f)\,d\mu$$

4.3.1 CLT for continuous-time martingales

Theorem 4.13 (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

$$\frac{M_t}{\sqrt{t}} \xrightarrow{\mathcal{D}} N(0,\sigma^2)$$

4.3.2 CLT for Markov processes

Corollary 4.14 (Central limit theorem for Markov processes (elementary version)). *Let (X_t, P_μ) be a stationary ergodic Markov process. Then for $f \in \mathrm{Range}(L),\ f = Lg$:*

$$\frac{1}{\sqrt{t}}\int_0^t f(X_s)\,ds \xrightarrow{\mathcal{D}} N(0,\sigma_f^2)$$

where

$$\sigma_f^2 = 2\int g(-L)g\,d\mu = 2\mathcal{E}(g,g)$$

Remark. (1). If μ is stationary then

$$\int f\,d\mu = \int Lg\,d\mu = 0$$

i.e. the random variables $f(X_s)$ are centered.

(2). $\ker(L) = \mathrm{span}\{1\}$ 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.,

$$\mathrm{spec}(-L) \subset \{0\} \cup [\lambda, \infty)$$

hence there is a $G = (-L\big|_{L_0^2(\mu)})^{-1}$, $\mathrm{spec}(G) \subseteq [0, \frac{1}{\lambda}]$ and hence

$$\sigma_f^2 \le \frac{2}{\lambda}\|f\|_{L^2(\mu)}^2$$

is a bound for asymptotic variance.

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: $\mathrm{Range}(L) \neq L^2$, replace $-L$ by $\alpha - L$ (bijective), then $\alpha \downarrow 0$. Cf. Landim [16].

4.4 Logarithmic Sobolev inequalities and entropy bouds

We consider the setup from section 4.3. In addition, we now assume that $(\mathcal{L}, \mathcal{A})$ is symmetric on $L^2(S, \mu)$.

4.4.1 Logarithmic Sobolev inequalities and hypercontractivity

Theorem 4.15. *With assumptions (A0)-(A3) and $\alpha > 0$, the following statements are equivalent:*

(i) ***Logarithmic Sobolev inequality (LSI)***

$$\int_S f^2 \log \frac{f^2}{\|f\|_{L^2(\mu)}^2}\, d\mu \leq 2\alpha\mathcal{E}(f,f) \quad \forall\, f \in \mathcal{A}$$

(ii) ***Hypercontractivity*** *For $1 \leq p < q < \infty$,*

$$\|p_t f\|_{L^q(\mu)} \leq \|f\|_{L^p(\mu)} \quad \forall\, f \in L^p(\mu),\ t \geq \frac{\alpha}{2} \log \frac{q-1}{p-1}$$

(iii) Assertion (ii) holds for $p = 2$.

Remark. Hypercontractivity and Spectral gap implies

$$\|p_t f\|_{L^q(\mu)} = \|p_{t_0} p_{t-t_0} f\|_{L^q(\mu)} \leq \|p_{t-t_0} f\|_{L^2(\mu)} \leq e^{-\lambda(t-t_0)} \|f\|_{L^2(\mu)}$$

for all $t \geq t_0(q) := \frac{\alpha}{4} \log(q-1)$.

Proof. (i)⇒(ii) **Idea:** WLOG $f \in \mathcal{A}_0,\ f \geq \delta > 0$ (which implies that $p_t f \geq \delta \,\forall\, t \geq 0$).

Compute

$$\frac{d}{dt}\|p_t f\|_{L^{q(t)}(\mu)}, \quad q\colon \mathcal{R}^+ \to (1,\infty) \text{ smooth:}$$

(1). 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

$$\int (p_t f)^{q(t)-1}\mathcal{L} p_t f\,d\mu = -\mathcal{E}\left((p_t f)^{q(t)-1}, p_t f\right)$$

(2). 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.

$$\begin{aligned}
\mathcal{E}(f^{q-1}, f) &= -\left(f^{q-1}, \mathcal{L} f\right)_\mu = \lim_{t\downarrow 0}\frac{1}{t}\left(f^{q-1}, f - p_t f\right)_\mu \\
&= \lim_{t\downarrow 0}\frac{1}{2t}\iint \left(f^{q-1}(y) - f^{q-1}(x)\right)\left(f(y) - f(x)\right) p_t(x,dy)\,\mu(dx) \\
&\geq \frac{4(q-1)}{q^2}\lim_{t\downarrow 0}\frac{1}{2t}\iint \left(f^{\frac{q}{2}}(y) - f\frac{q}{2}(x)\right)^2 p_t(x,dy)\,\mu(dx) \\
&= \frac{4(q-1)}{q^2}\mathcal{E}\left(f^{\frac{q}{2}}, f^{\frac{q}{2}}\right)
\end{aligned}$$

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.

(3). 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

$$\begin{aligned} & q(t) \cdot \|p_t f\|_{q(t)}^{q(t)-1} \frac{d}{dt} \|p_t f\|_{q(t)} \\ & \leq -\frac{4(q(t)-1)}{q(t)} \mathcal{E}\left((p_t f)^{\frac{q(t)}{2}}, (p_t f)^{\frac{q(t)}{2}}\right) + \frac{q'(t)}{q(t)} \cdot \int (p_t f)^{q(t)} \log \frac{(p_t f)^{q(t)}}{\int (p_t f)^{q(t)} \, d\mu} \, d\mu \end{aligned}$$

(4). 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 $\frac{d}{dt} \|p_t f\|_{L^{q(t)}(\mu)}$). □

Theorem 4.16 (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 $x \log x = x - 1 + \frac{1}{2}(x-1)^2 + \mathrm{O}(|x-1|^3)$. □

4.4.2 Decay of relative entropy

Theorem 4.17 (Exponential decay of relative entropy). *(1). $H(\nu p_t | \mu) \leq H(\nu | \mu)$ for all $t \geq 0$ and $\nu \in M_1(S)$.*

(2). 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 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

$$\int f^2 \log \frac{f^2}{\|f\|_{L^2(\mu)}^2}\, d\mu \leq \frac{\alpha}{2} \int |\nabla f|^2\, d\mu = \alpha \mathcal{E}(f,f)$$

(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

$$\frac{d}{dt} H(\nu p_t|\mu) = -2 \int |\nabla \sqrt{p_t g}|^2\ d\mu \tag{4.4.1}$$

(1). $-2 \int \left|\nabla \sqrt{p_t g}\right|^2\ d\mu \leq 0$

(2). The Logarithmic Sobolev Inequality yields that

$$-2 \int |\nabla \sqrt{p_t g}|^2\ d\mu \leq -\frac{4}{\alpha} \int p_t g \log \frac{p_t g}{\int p_t g\, d\mu}\, d\mu$$

where $\int p_t g\, d\mu = \int g\, d\mu = 1$ and hence

$$-2 \int |\nabla \sqrt{p_t g}|^2\ d\mu \leq -\frac{4}{\alpha} H(\nu p_t|\mu)$$

(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

$$g_{a,b} := (g \vee a) \wedge b, \quad 0 < a < b,$$
$$\nu_{a,b} := g_{a,b} \cdot \mu.$$

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. (1). The proof in the general case is analogous, just replace (4.4.1) by inequality

$$4\mathcal{E}(\sqrt{f}, \sqrt{f}) \le \mathcal{E}(f, \log f)$$

(2). 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 4.18 (Total variation bound). *For all $t \ge 0$ and $\nu \in M_1(S)$,*

$$\begin{aligned} \|\nu p_t - \mu\|_{\text{TV}} &\le \frac{1}{\sqrt{2}} e^{-\frac{t}{\alpha}} H(\nu|\mu)^{\frac{1}{2}} \\ &\left(\le \frac{1}{\sqrt{2}} \log \frac{1}{\min \mu(x)} e^{-\frac{t}{\alpha}} \quad \textit{if } S \textit{ is finite} \right) \end{aligned}$$

Proof.

$$\|\nu p_t - \mu\|_{\text{TV}} \le \frac{1}{\sqrt{2}} H(\nu p_t|\mu)^{\frac{1}{2}} \le \frac{1}{\sqrt{2}} e^{-\frac{t}{\alpha}} H(\nu|\mu)^{\frac{1}{2}}$$

where we use Pinsker's Theorem for the first inequality and Theorem 4.17 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)

$$\begin{aligned} T_{\text{mix}}(\varepsilon) &= \inf \{ t \ge 0 \ : \ \|\nu p_t - \mu\|_{\text{TV}} \le \varepsilon \text{ for all } \nu \in M_1(S) \} \\ &\le \alpha \cdot \log \frac{1}{\sqrt{2}\varepsilon} + \log \log \frac{1}{\min_{x \in S} \mu(x)} \end{aligned}$$

Hence we have $\log \log$ instead of $\log$!

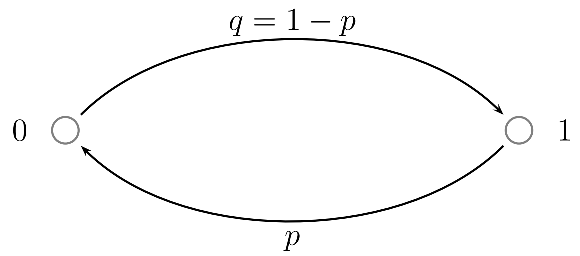
4.4.3 LSI on product spaces

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:

$$\begin{aligned} \mathcal{E}(f, f) &= \frac{1}{2} \sum_{x,y} (f(y) - f(x))^2 \, \mu(x) \mathcal{L}(x, y) \\ &= pq \cdot |f(1) - f(0)|^2 = \mathrm{Var}_\mu(f) \end{aligned}$$

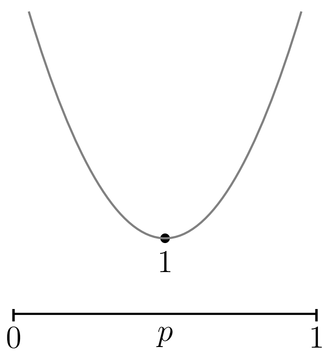
Spectral gap:

$$\lambda(p) = \inf_{f \text{not const.}} \frac{\mathcal{E}(f, f)}{\mathrm{Var}_\mu(f)} = 1 \quad \text{independent of } p\,!$$

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 $p \uparrow \infty$!



Spectral gap and Logarithmic Sobolev Inequality for product measures:

$$\mathrm{Ent}_\mu(f) := \int f \log f \, d\mu, \;\; f > 0$$

Theorem 4.19 (Factorization property). *$(S_i, \mathcal{S}_i, \mu_i)$ probability spaces, $\mu = \otimes_{i=1}^n \mu_i$. Then*

(1).

$$\mathrm{Var}_\mu(f) \leq \sum_{i=1}^n E_\mu \left[\mathrm{Var}_{\mu_i}^{(i)}(f) \right]$$

where on the right hand side the variance is taken with respect to the i-th variable.

(2).

$$\mathrm{Ent}_\mu(f) \leq \sum_{i=1}^n E_\mu \left[\mathrm{Ent}_{\mu_i}^{(i)}(f) \right]$$

Proof. (1). Exercise.

(2).

$$\mathrm{Ent}_\mu(f) = \sup_{g \,:\, E_\mu[e^g]=1} E_\mu[fg], \quad \text{cf. above}$$

Fix $g\colon S^n \to \mathcal{R}$ 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 4.20. *(1). 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

$$\lambda = \min_{1\le i\le n} \lambda_i$$

(2). 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. $S = \{0,1\}^n,\ \mu^n$ product of $\mathrm{Bernoulli}(p)$,

$$\begin{aligned}
&\mathrm{Ent}_{\mu^n}(f^2)\\
&\leq 2\alpha(p)\cdot p\cdot q\cdot\sum_{i=1}^n \int \left|f(x_1,\dots,x_{i-1},1,x_{i+1},\dots,x_n) - f(x_1,\dots,x_{i-1},0,x_{i+1},\dots,x_n)\right|^2\,\mu^n(dx)
\end{aligned}$$

independent of n.

Example. Standard normal distribution $\gamma = N(0,1)$,

$$\varphi_n\colon \{0,1\}^n \to \mathcal{R}, \qquad \varphi_n(x) = \frac{\sum_{i=1}^n \left(x_i - \frac{1}{2}\right)}{\sqrt{\frac{n}{4}}}$$

The Central Limit Theorem yields that $\mu = \mathrm{Bernoulli}(\frac{1}{2})$ and hence

$$\mu^n \circ \varphi_n^{-1} \overset{w}{\to} \gamma$$

Hence for all $f \in C_0^\infty(\mathbb{R})$,

$$\begin{aligned}
\mathrm{Ent}_\gamma(f^2) &= \lim_{n\to\infty} \mathrm{Ent}_{\mu^n}(f^2\circ\varphi_n)\\
&\leq \liminf \frac{1}{2}\sum_{i=1}^n \int \left|\Delta_i f\circ\varphi_n\right|^2\,d\mu^n\\
&\leq \dots \leq 2\cdot\int |f'|^2\,d\gamma
\end{aligned}$$

4.4.4 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

$$x \cdot \nabla H(x) \geq \kappa \cdot |x|^2 - c, \tag{4.4.2}$$

$$H(x) \geq \frac{\kappa}{2}|x|^2 - \widetilde{c} \tag{4.4.3}$$

with constants $c, \widetilde{c} \in \mathbb{R}$. By (4.4.2) 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 $\mathcal{A}_0 = \operatorname{span}\left(\mathbb{C}_0^\infty(\mathbb{R}^n), 1\right)$, $\mathcal{A} = \operatorname{span}\left(\mathcal{S}(\mathbb{R}^n), 1\right)$. By (4.4.3), the measure μ is finite, hence by our results above, the normalized measure is a stationary distribution for p_t.

Lemma 4.21. *If* $\operatorname{Hess} H \geq \kappa I$ *then*

$$|\nabla p_t f| \leq e^{-\kappa t} p_t |\nabla f| \qquad f \in C_b^1(\mathbb{R}^n)$$

Remark. (1). Actually, both statements are equivalent.

(2). 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).

Informal analytic proof:

$$\begin{aligned}
\nabla \mathcal{L} f &= \nabla \left(\Delta - \nabla H \cdot \nabla\right) f \\
&= \left(\Delta - \nabla H \cdot \nabla - \partial^2 H\right) \nabla f \\
&=: \vec{\mathcal{L}} \ \text{ operator on one-forms (vector fields)}
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial}{\partial t} |\nabla p_t f| &= \frac{\partial}{\partial t} \left(\nabla p_t f \cdot \nabla p_t f\right)^{\frac{1}{2}} = \frac{\left(\frac{\partial}{\partial t} \nabla p_t f\right) \cdot \nabla p_t f}{|\nabla p_t f|} \\
&= \frac{\left(\vec{\mathcal{L}}\, \nabla p_t f\right) \cdot \nabla p_t f}{|\nabla p_t f|} \leq \frac{\mathcal{L} \nabla p_t f \cdot \nabla p_t f}{|\nabla p_t f|} - \kappa \cdot \frac{|\nabla p_t f|^2}{|\nabla p_t f|} \\
&\leq \cdots \leq \mathcal{L} |\nabla p_t f| - \kappa |\nabla p_t f|
\end{aligned}$$

We get that $v(t) := e^{\kappa t} p_{s-t} |\nabla p_t f|$ with $0 \leq t \leq s$ satisfies

$$v'(t) \leq \kappa v(t) - p_{s-t}\mathcal{L} |\nabla p_t f| + p_{s-t}\mathcal{L} |\nabla p_t f| - \kappa p_{s-t} |\nabla p_t f| = 0$$

and hence

$$e^{\kappa s} |\nabla p_s f| = v(s) \leq v(0) = p_s |\nabla f|$$

□

- 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[6].
- The assertion extends to general diffusion operators.

Probabilistic proof: $p_t f(x) = E[f(X_t^x)]$ where X_t^x is the solution flow of the stochastic differential equation

$$dX_t = \sqrt{2} dB_t - (\nabla H)(X_t)\, dt, \quad \text{i.e.,}$$
$$X_t^x = x + \sqrt{2} B_t - \int_0^t (\nabla H)(X_s^x)\, ds$$

By the assumption on H one can show that $x \to X_t^x$ is smooth and the derivative flow $Y_t^x = \nabla_x X_t$ satisfies the differentiated stochastic differential equation

$$dY_t^x = -(\partial^2 H)(X_t^x) Y_t^x\, dt,$$
$$Y_0^x = I$$

which is an ordinary differential equation. Hence if $\partial^2 H \geq \kappa I$ then for $v \in \mathbb{R}^n$,

$$\frac{d}{dt} |Y_t \cdot v|^2 = -2 \left(Y_t \cdot v,\ (\partial^2 H)(X_t) Y_t \cdot v \right)_{\mathcal{R}^n} \leq 2\kappa \cdot |Y_t \cdot v|^2$$

where $Y_t \cdot v$ is the derivative of the flow in direction v. Hence

$$|Y_t \cdot v|^2 \leq e^{-2\kappa t} |v|$$
$$\Rightarrow \quad |Y_t \cdot v| \leq e^{-\kappa t} |v|$$

This implies that for $f \in C_b^1(\mathbb{R}^n)$, $p_t f$ is differentiable and

$$v \cdot \nabla p_t f(x) = E\left[(\nabla f(X_t^x) \cdot Y_t^x \cdot v) \right]$$
$$\leq E\left[|\nabla f(X_t^x)| \right] \cdot e^{-\kappa t} \cdot |v| \quad \forall\, v \in \mathbb{R}^n$$

i.e.

$$|\nabla p_t f(x)| \leq e^{-\kappa t} p_t |\nabla f|(x)$$

□

Theorem 4.22 (Bakry-Emery). *Suppose that*

$$\partial^2 H \geq \kappa \cdot I \quad \textit{with } \kappa > 0$$

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:

$$\int g \log g\, d\mu \leq \frac{1}{\kappa} \int |\nabla \sqrt{g}|^2\ d\mu + \int g\, d\mu \log \int g\, d\mu$$

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:

$$|\nabla p_t g| \leq e^{-\kappa t} p_t |\nabla g| \leq e^{-\kappa t} |\nabla g|$$

and hence

$$|p_t g(x) - p_t g(y)| \leq e^{-\kappa t} \sup |\nabla g| \cdot |x - y|$$

which leads to

$$\begin{aligned} \left| p_t g(x) - \int g \, d\mu \right| &= \left| \int \left(p_t g(x) - p_t g(y) \right) \, \mu(dy) \right| \\ &\leq e^{-\kappa t} \sup |\nabla g| \cdot \int |x - y| \, \mu(dy) \to 0 \end{aligned}$$

Since $p_t g \geq \delta \geq 0$, 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,

$$\begin{aligned} -u'(t) = \int \nabla p_t g \cdot \nabla \log p_t g \, d\mu &= \int \frac{|\nabla p_t g|^2}{p_t g} \, d\mu \\ &\leq e^{-2\kappa t} \int \frac{(p_t |\nabla g|)^2}{p_t g} \, d\mu \leq e^{-2\kappa t} \int p_t \frac{|\nabla g|^2}{g} \, d\mu \\ &= e^{-2\kappa t} \int \frac{|\nabla g|^2}{g} \, d\mu = 4 e^{-2\kappa t} \int |\nabla \sqrt{g}|^2 \, d\mu \end{aligned}$$

□

Example. An Ising model with real spin: (Reference: Royer [33])
$S = \mathbb{R}^\Lambda = \{(x_i)_{i \in \Lambda} \mid x_i \in \mathbb{R}\}$, $\Lambda \subset \mathbb{Z}^d$ 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 $V: \mathbb{R} \to \mathbb{R}$ is a non-constant polynomial, bounded from below, and $\vartheta: \mathbb{Z} \to \mathbb{R}$ is a function such that $\vartheta(0) = 0$, $\vartheta(i) = \vartheta(-i)\ \forall\, i$, (symmetric interactions), $\vartheta(i) = 0\ \forall\, |i| \geq R$ (finite range), $z \in \mathcal{R}^{\mathbb{Z}^d \setminus \Lambda}$ fixed boundary condition.

Glauber-Langevin dynamics:

$$dX_t^i = -\frac{\partial H}{\partial x_i}(X_t)\, dt + dB_t^i, \quad i \in \Lambda \tag{4.4.4}$$

Dirichletform:

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{i \in \Lambda} \int \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_i}\, d\mu$$

Corollary 4.23. *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.

$$\begin{aligned} \frac{\partial^2 H}{\partial x_i \partial x_j}(x) &= V''(x_i) \cdot \delta_{ij} - \vartheta(i-j) \\ \Rightarrow \quad \partial^2 H &\geq \left(\inf V'' - \sum_i |\vartheta(i)| \right) \cdot I \end{aligned}$$

in the sense of ???. □

Consequence: There is a unique Gibbs measure on $\mathbb{Z}^d$ corresponding to H, cf. Royer [33]. What can be said if V is not convex?

4.4.5 Stability under bounded perturbations

Theorem 4.24 (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

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\mu)}}\, d\mu \leq 2\alpha \cdot \int |\nabla f|^2\, d\mu \quad \forall\, f \in \mathbb{C}_0^\infty$$

then

$$\int f^2 \log \frac{f^2}{\|f\|^2_{L^2(\nu)}}\, d\nu \leq 2\alpha \cdot e^{\operatorname{osc}(U)} \cdot \int |\nabla f|^2\, d\nu \quad \forall\, f \in C_0^\infty$$

where

$$\operatorname{osc}(U) := \sup U - \inf U$$

Proof.

$$\int f^2 \log \frac{|f|^2}{\|f\|^2_{L^2(\nu)}}\, d\nu \leq \int \left(f^2 \log f^2 - f^2 \log \|f\|^2_{L^2(\mu)} - f^2 + \|f\|^2_{L^2(\mu)} \right) d\nu \tag{4.4.5}$$

since

$$\int f^2 \log \frac{|f|^2}{\|f\|^2_{L^2(\nu)}}\, d\nu \leq \int f^2 \log f^2 - f^2 \log t^2 - f^2 + t^2 \, d\nu \quad \forall\, t > 0$$

Note that in (4.4.5) 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

(1). **No interactions:**

$$H(x) = \sum_{i \in \Lambda} \left(\frac{x_i^2}{2} + V(x_i) \right), \quad V : \mathcal{R} \to \mathcal{R} \text{ bounded}$$

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!

(2). **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 4.25. *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 $\mathrm{Cov}_\mu(x_i, x_j)$ for Gibbs measure, cf. ???, Course ???.

(3). **Discrete Ising model:** One can show that for $\beta < \beta_c$ a logarithmic Sobolev inequality holds on$\{-N, \dots, N\}^d$ 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. [???].

4.5 Concentration of measure

$(\Omega, \mathcal{A}, P)$ probability space, $X_i \colon \Omega \to \mathbb{R}^d$ independent identically distributed, $\sim \mu$.
Law of large numbers:

$$\frac{1}{N} \sum_{i=1}^N U(X_i) \to \int U \, d\mu \quad U \in \mathcal{L}^1(\mu)$$

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 4.26 (Herbst). *If μ satisfies a logarithmic Sobolev inequality with constant α then for any function $U \in C_b^1(\mathbb{R}^d)$ with $\|U\|_{Lip} \leq 1$:*

(i)

$$\frac{1}{t}\log \int e^{tU}\,d\mu \leq \frac{\alpha}{2}t + \int U\,d\mu \quad \forall\, t > 0 \tag{4.5.1}$$

*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 4.27 (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 $\|U\|_{Lip} \leq 1$, $N \in \mathbb{N}$ 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,

$$\left|\nabla\widetilde{U}(x)\right| = \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N} |\nabla U(x_i)|^2\right)^{\frac{1}{2}} \leq 1$$

□

Appendix

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 expectation

For more details and proofs of the following statements see [Eberle:Stochastic processes] [10].

Definition (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.1. *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:*

(1). Linearity: $E[\lambda X + \mu Y|\mathcal{F}] = \lambda E[X|\mathcal{F}] + \mu E[Y|\mathcal{F}]$ *P-almost surely for all* $\lambda, \mu \in \mathbb{R}$.

(2). Monotonicity: $X \geq 0$ *P-almost surely implies that* $E[X|\mathcal{F}] \geq 0$ *P-almost surely.*

(3). $X = Y$ *P-almost surely implies that* $E[X|\mathcal{F}] = E[Y|\mathcal{F}]$ *P-almost surely.*

(4). Monotone convergence: If (X_n) *is growing monotone with* $X_1 \geq 0$*, then*

$$E[\sup X_n|\mathcal{F}] = \sup E[X_n|\mathcal{F}] \text{ } P\text{-almost surely.}$$

(5). Projectivity / Tower property: If $\mathcal{G} \subset \mathcal{F}$*, then*

$$E[E[X|\mathcal{F}]|\mathcal{G}] = E[X|\mathcal{G}] \text{ } P\text{-almost surely.}$$

In particular:

$$E[E[X|Y,Z]|Y] = E[X|Y] \text{ } P\text{-almost surely.}$$

(6). Let Y *be* $\mathcal{F}$*-measurable with* $Y \cdot X \in \mathcal{L}^1$ *or* ≥ 0*. This implies that*

$$E[Y \cdot X|\mathcal{F}] = Y \cdot E[X|\mathcal{F}] \text{ } P\text{-almost surely.}$$

(7). Independence: If X *is independent of* $\mathcal{F}$*, then* $E[X|\mathcal{F}] = E[X]$ *P-almost surely.*

(8). 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 (Conditional probability). *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space,* $\mathcal{F}$ *a* σ*-algebra. The* ***conditional probability*** *is defined as*

$$P[A|\mathcal{F}](\omega) := E[\mathbb{1}_A|\mathcal{F}](\omega) \; \forall A \in \mathcal{F}, \omega \in \Omega.$$

A.2 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.

A.2.1 Filtrations

We fix a probability space $(\Omega, \mathcal{A}, P)$. Moreover, we assume that we are given an increasing sequence $\mathcal{F}_n$ $(n = 0, 1, 2, \ldots)$ 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 (Filtration, adapted process). *(1). A filtration on $(\Omega, \mathcal{A})$ is an increasing sequence*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$$

of σ-algebras $\mathcal{F}_n \subseteq \mathcal{A}$.

(2). 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. (1). The *canonical filtration* $(\mathcal{F}_n^X)$ generated by a stochastic process (X_n) is given by

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \ldots, X_n).$$

If the filtration is not specified explicitly, we will usually consider the canonical filtration.

(2). Alternatively, filtrations containing additional information are of interest, for example the filtration

$$\mathcal{F}_n = \sigma(Z, X_0, X_1, \ldots, X_n)$$

generated by the process (X_n) and an additional random variable Z, or the filtration

$$\mathcal{F}_n = \sigma(X_0, Y_0, X_1, Y_1, \ldots, X_n, Y_n)$$

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 $n \geq 0$.

A.2.2 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 (Martingale, supermartingale, submartingale).

(1). A sequence of real-valued random variables $M_n : \Omega \to \mathbb{R} \quad (n = 0, 1, \ldots)$ *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)* $E[M_n \mid \mathcal{F}_{n-1}] \ = \ M_{n-1}$ *for any* $n \in \mathbb{N}$*.*

(2). 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 "*$\leq$*", "*$\geq$*" 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 "$\leq$" or "$\geq$" for super- or submartingales.

Intuitively, a martingale is a ”fair gameÅ$\frac{1}{2}$Å$\frac{1}{2}$, 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.2. *If* (M_n) *is a martingale (respectively a supermartingale) w.r.t. a filtration* $(\mathcal{F}_n)$ *then*

$$E[M_{n+k} \mid \mathcal{F}_n] \ \overset{(\leq)}{=} \ M_n \qquad P\text{-almost surely for any } n, k \geq 0.$$

A.2.3 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 (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.3 (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)*

$$Y_n \quad = \quad M_n + A_n \tag{A.2.1}$$

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

$$A_n \quad = \quad \sum_{k=1}^{n} E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}], \qquad \textit{and} \quad M_n \;=\; Y_n - A_n. \tag{A.2.2}$$

Remark. (1). The increments $E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}]$ of the process (A_n) are the predicted increments of (Y_n) given the previous information.

(2). The process (Y_n) is a supermartingale (resp. a submartingale) if and only if the predictable part (A_n) is decreasing (resp. increasing).

Proof of Theorem A.3. Uniqueness: For any decomposition as in (A.2.1) 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

$$E[Y_k - Y_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad E[A_k - A_{k-1} \,|\, \mathcal{F}_{k-1}] \quad = \quad A_k - A_{k-1} \qquad P\text{-a.s.}$$

This implies that (A.2.2) holds almost surely if $A_0 = 0$.

Existence: Conversely, if (A_n) and (M_n) are defined by (A.2.2) then (A_n) is predictable with $A_0 = 0$ and (M_n) is a martingale, since

$$E[M_k - M_{k-1} \,|\, \mathcal{F}_{k-1}] \;=\; 0 \qquad P\text{-a.s. for any } k \in \mathbb{N}.$$

□

A.3 Gambling strategies and stopping times

Throughout this section, we fix a filtration $(\mathcal{F}_n)_{n\geq 0}$ on a probability space $(\Omega, \mathcal{A}, P)$.

A.3.1 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 (Martingale transform). *The stochastic process $C_\bullet M$ defined by*

$$(C_\bullet M)_n \;:=\; \sum_{k=1}^{n} C_k \cdot (M_k - M_{k-1}) \qquad \textit{for any } n \geq 0,$$

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

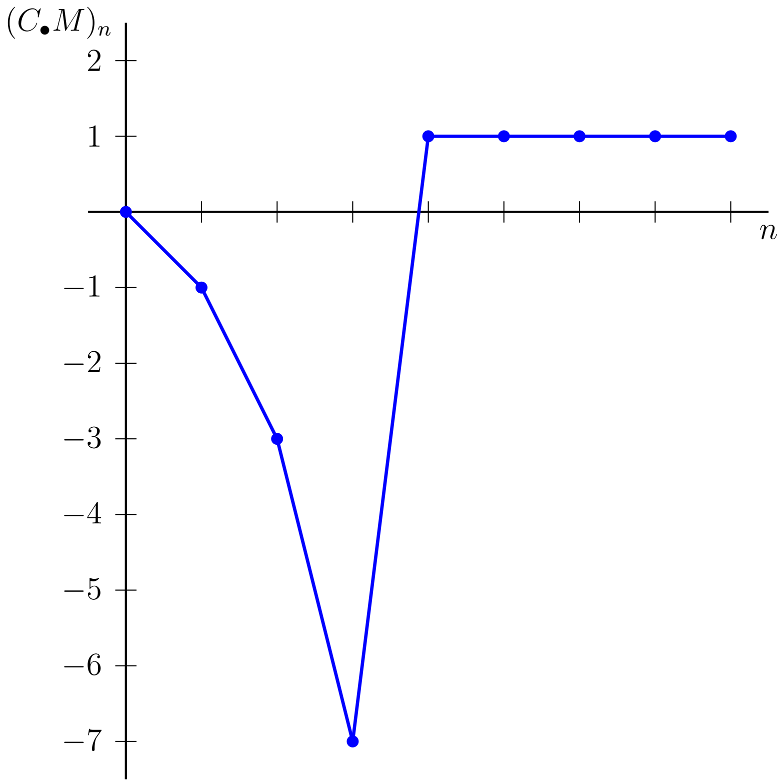
$$M_n \;=\; \eta_1 + \ldots + \eta_n, \qquad \eta_i \text{ i.i.d. with } P[\eta_i = 1] = P[\eta_i = -1] = 1/2,$$

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.4 (You can't beat the system!). *(1). 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.*

(2). 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.4 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.

A.3.2 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.5 (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*

$$E[M_{T\wedge n}] \overset{(\leq)}{=} E[M_0] \qquad \textit{for any } n \geq 0.$$

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,

$$M_{T\wedge n} - M_0 = (C_\bullet M)_n \qquad \text{for any } n \geq 0. \tag{A.3.1}$$

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.3.1) 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). (1). In general, it is NOT TRUE under the assumptions in Theorem A.5 that

$$E[M_T] = E[M_0], \quad E[M_T] \leq E[M_0] \quad \text{respectively}. \tag{A.3.2}$$

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 $M_0 = 0$.

(2). 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.

(1). *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

$$x \;=\; E[X_0] \;=\; E[X_{T\wedge n}] \;\stackrel{n\to\infty}{\to}\; E[X_T] \;=\; a \cdot r(x) + b \cdot (1 - r(x))$$

by the Optional Stopping Theorem and the Dominated Convergence Theorem. Hence

$$r(x) \quad = \quad \frac{b - x}{a - x}. \tag{A.3.3}$$

(2). *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

$$\begin{aligned} x^2 \quad &= \quad E[M_0] \;=\; E[M_{T\wedge n}] \;=\; E[X_{T\wedge n}^2] - E[T \wedge n] \\ &\stackrel{n\to\infty}{\longrightarrow} \quad E[X_T^2] - E[T]. \end{aligned}$$

Therefore, by (A.3.3),

$$\begin{aligned} E[T] \quad &= \quad E[X_T^2] - x^2 \;=\; a^2 \cdot r(x) + b^2 \cdot (1 - r(x)) - x^2 \\ &= \quad (b - x) \cdot (x - a). \end{aligned} \tag{A.3.4}$$

(3). *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.3.4), 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.

(4). *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,

$$1 \ = \ E[M_0^\lambda] \ = \ E[M_{T\wedge n}^\lambda] \ = \ E[e^{\lambda X_{T\wedge n}}/(\cosh \lambda)^{T\wedge n}] \tag{A.3.5}$$

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.3.5) converges to $E[e^\lambda/(\cosh \lambda)^T \ ; \ T < \infty]$, and we obtain the identity

$$E[(\cosh \lambda)^{-T} \ ; \ T < \infty] \ = \ e^{-\lambda} \qquad \text{for any } \lambda > 0. \tag{A.3.6}$$

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.3.6), we can now compute the generating function of T explicitly:

$$E[s^T] \ = \ e^{-\lambda} \ = \ (1 - \sqrt{1 - s^2})/s \qquad \text{for any } s \in (0, 1). \tag{A.3.7}$$

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 $m \geq 1$.

A.3.3 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.3.2) holds nevertheless. A first general criterion is obtained by applying the Dominated Convergence Theorem:

Theorem A.6 (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*

$$|M_{T\wedge n}| \le Y \qquad P\text{-almost surely for any } n \in \mathbb{N}.$$

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.5, $E[M_0]=E[M_{T\wedge n}]$, 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.,

$$\sup_{n\in\mathbb{N}} E\big[|M_{T\wedge n}|\,;\,|M_{T\wedge n}|\ge c\big] \quad\to\quad 0 \qquad \text{as } c\to\infty.$$

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.7 (Optional Stopping Theorem, Version 3). *If (M_n) is a non-negative supermartingale w.r.t. $(\mathcal{F}_n)$, then*

$$E[M_0] \;\ge\; E[M_T\,;\,T<\infty]$$

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.5 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.4 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.8 (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.*

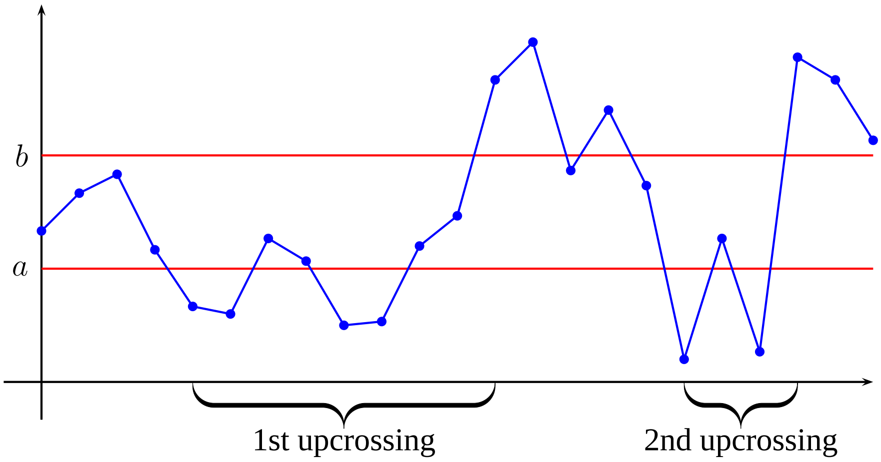
Remark (L^1 boundedness and L^1 convergence).

(1). Although the limit is integrable, L^1 convergence does *not* hold in general.

(2). 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.

A.4.1 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

$$U_n^{(a,b)} \;=\; \max\big\{k \geq 0 : \exists\, 0 \leq s_1 < t_1 < s_2 < t_2 \ldots < s_k < t_k \leq n : Z_{s_i} \leq a, Z_{t_i} \geq b\big\}.$$

Lemma A.9 (Doob). *If (Z_n) is a supermartingale then*

$$(b-a) \cdot E[U_n^{(a,b)}] \;\leq\; E[(Z_n - a)^-] \qquad \text{for any } a < b \text{ and } n \geq 0.$$

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:

repeat

- 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

$$C_k = \begin{cases} 1 & \text{if } (C_{k-1} = 1 \text{ and } Z_{k-1} \le b) \text{ or } (C_{k-1} = 0 \text{ and } Z_{k-1} \le a), \\ 0 & \text{otherwise.} \end{cases}$$

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.4 and the remark below, the process

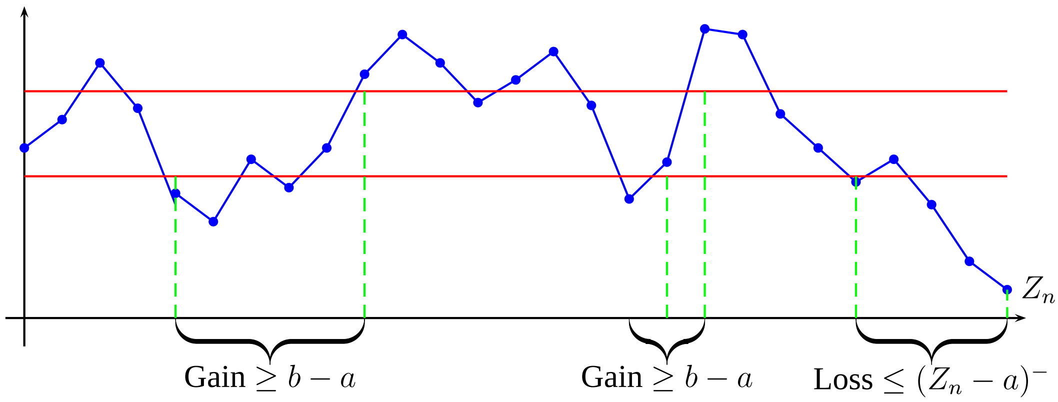
$$(C_\bullet Z)_k = \sum_{i=1}^{k} C_i \cdot (Z_i - Z_{i-1}), \qquad 0 \le k \le n,$$

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

$$(b-a)E[U_n^{(a,b)}] \le E[(C_\bullet Z)_n] + E[(Z_n - a)^-] \le E[(Z_n - a)^-].$$

□

A.4.2 Proof of Doob's Convergence Theorem

We can now complete the proof of Theorem A.8.

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,

$$E[U^{(a,b)}] \;=\; \lim_{n\to\infty} E[U_n^{(a,b)}] \;\leq\; \frac{1}{b-a}\cdot \sup_{n\in\mathbb{N}} E[(Z_n-a)^-]. \qquad \text{(A.4.1)}$$

Assuming $\sup E[Z_n^-] < \infty$, the right hand side of (A.4.1) is finite since $(Z_n-a)^- \leq |a| + Z_n^-$. 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.8, $\sup E[Z_n^-] < \infty$ 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. □

A.4.3 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:

$$\eta_i \;\text{ i.i.d. with } |\eta_i| \leq c \text{ and } E[\eta_i] = 0, \qquad c\in\mathbb{R}.$$

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.,

$$\limsup S_n \;=\; \infty \qquad \text{almost surely.}$$

Since (S_n) is also a submartingale, we obtain

$$\liminf S_n \;=\; -\infty \qquad \text{almost surely}$$

by an analogue argument.

Remark (Almost sure vs. Lᵖ 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:

(1). $\alpha < 1$ *(subcritical):* In this case,

$$Z_n \quad = \quad M_n \cdot \alpha^n$$

converges to 0 exponentially fast with probability one.

(2). $\alpha = 1$ *(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.

(3). $\alpha > 1$ *(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:

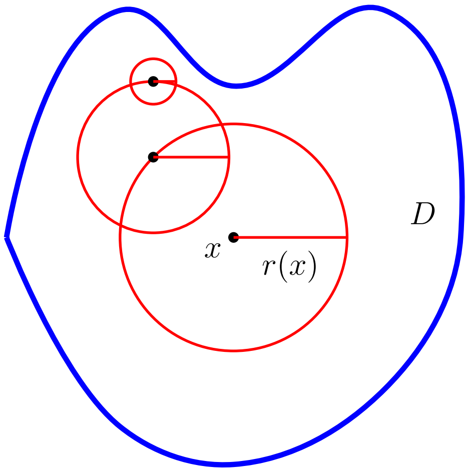
$$\Delta h(x) \;=\; 0 \quad \text{for any } x \in D, \quad \inf_{x \in D} h(x) > -\infty. \tag{A.4.2}$$

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

$$0 \;<\; r(x) \;<\; \operatorname{dist}(x, \partial D) \qquad \text{for any } x \in D, \tag{A.4.3}$$

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.4.3), 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.4.2), 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:

(1). *Boundary regularity:* If h is harmonic and bounded from below on D then the limit $\lim\limits_{n\to\infty} h(X_n)$ exists along almost every trajectory X_n to the boundary ∂D.

(2). *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!

A.5 Brownian Motion

Definition (Brownian motion)**.**

(1). 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) $B_0(\omega) = a$ *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:

$$B_t - B_s \sim N(0, (t-s) \cdot I_d) \quad \text{for any } 0 \leq s \leq t.$$

Bibliography

[1] D. Bakry, I. Gentil, and M. Ledoux. Analysis and Geometry of Markov Diffusion Operators. Grundlehren der mathematischen Wissenschaften. Springer, 2013.

[2] P. Billingsley. Convergence of probability measures. Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics. Wiley, 1968.

[3] E. Bolthausen, A.S. Sznitman, and Deutschen Mathematiker-Vereinigung. Ten Lectures on Random Media. DMV Seminar. Momenta, 2002.

[4] M. Chen. From Markov Chains to Non-equilibrium Particle Systems. World Scientific, 2004.

[5] E.B. Davies. One-parameter semigroups. L.M.S. monographs. Academic Press, 1980.

[6] J.D. Deuschel and D.W. Stroock. Large deviations. Pure and Applied Mathematics. Elsevier Science, 1989.

[7] Andreas Eberle. Lecture „Introduction to Stochastic Analysis“. Institute for Applied Mathematics, University of Bonn, 2011. `http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/St`

[8] Andreas Eberle. Lecture „Einführung in die Wahrscheinlichkeitstheorie“. Institute for Applied Mathematics, University of Bonn, 2013. `http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/Wt`

[9] Andreas Eberle. Lecture „Stochastic Analysis“. Institute for Applied Mathematics, University of Bonn, 2013. `http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/St`

[10] Andreas Eberle. Lecture „Stochastic processes“. Institute for Applied Mathematics, University of Bonn, 2014. `http://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/St`

[11] 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.

[12] Martin Hairer. Lecture „Convergence of Markov processes“. Mathematics Department, University of Warwick, 2010. `http://www.hairer.org/notes/Convergence.pdf`.

[13] Galin L. Jones. On the markov chain central limit theorem. Probability Surveys, 1:299–320, 2004.

[14] I. Karatzas and S.E. Shreve. Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics. Springer New York, 1991.

[15] R. Khasminskii. Stochastic Stability of Differential Equations. Stochastic Modelling and Applied Probability. Springer, 2011.

[16] Claudio Landim. Central limit theorem for markov processes. In Pierre Picco and Jaime San Martin, editors, From Classical to Modern Probability, volume 54 of Progress in Probability, pages 145–205. Birkhaeuser Basel, 2003.

[17] S. Lang. Analysis I. Addison-Wesley world student series. Addison-Wesley, 1978.

[18] T.M. Liggett. Interacting Particle Systems. Classics in Mathematics. Springer, 2004.

[19] T.M. Liggett. Continuous Time Markov Processes: An Introduction. Graduate studies in mathematics. American Mathematical Soc., 2010.

[20] Georg Lindgren. „Lecture notes on Stationary Stochastic Processes“. Lund Institute of Technology Centre for Mathematical Sciences Mathematical Statistics, 2002. `http://www.maths.lth.se/matstat/staff/georg/Publications/lecture2002.pc`

[21] Florent Malrieu. Lecture „Processus de Markov et inégalités fonctionelles“, 2005/2006. `http://perso.univ-rennes1.fr/florent.malrieu/MASTER2/markov180406.pdf`.

[22] Robert J. McCann. Exact solutions to the transportation problem on the line. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 455(1984):1341–1380, 1999.

[23] Sean Meyn and Richard L. Tweedie. Markov Chains and Stochastic Stability. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

[24] A. Pazy. Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences. Springer New York, 1992.

[25] S.T. Rachev and L. Rüschendorf. Mass Transportation Problems: Volume I: Theory. Mass Transportation Problems. Springer, 1998.

[26] M. Reed and B. Simon. Methods of modern mathematical physics. 2. Fourier analysis, self-adjointness. Fourier Nanlysis, Self-adjointness. Academic Press, 1975.

[27] M. Reed and B. Simon. Methods of modern mathematical physics: Analysis of operators. Methods of Modern Mathematical Physics. Academic Press, 1978.

[28] M. Reed and B. Simon. Methods of Modern Mathematical Physics: Functional analysis. Number Bd. 1 in Methods of Modern Mathematical Physics. Academic Press, 1980.

[29] M. Reed and B. Simon. Methods of Modern Mathematical Physics: Scattering theory. Number Bd. 3. Acad. Press, 1986.

[30] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. Probab. Surv., 1:20–71, 2004.

[31] L. C. G. Rogers and David Williams. Diffusions, Markov processes, and martingales. Vol. 2. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[32] L.C.G. Rogers and D. Williams. Diffusions, Markov Processes, and Martingales: Volume 1, Foundations. Cambridge Mathematical Library. Cambridge University Press, 2000.

[33] G. Royer. An initiation to logarithmic Sobolev inequalities. American Mathematical Society, 2007.

[34] Daniel W. Stroock. c. Cambridge University Press, New York, NY, USA, 2nd edition, 2010.

[35] D.W. Stroock and S.R.S. Varadhan. Multidimensional Diffussion Processes. Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Springer, 1979.

[36] S. R. S. Varadhan. Probability Theory. Courant Lecture Notes. Courant Institute of Mathematical Sciences, 2001.

[37] C. Villani. Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften. Springer, 2008.

[38] Cédric Villani. Topics in Optimal Transportation. Graduate studies in mathematics. American Mathematical Society, 2003.

[39] K. Yosida. Functional Analysis. Classics in Mathematics. Cambridge University Press, 1995.