Lecture 4: Markov chains and SDEs

2025–05–14

1.1 Timescales and convergence to equilibrium

In this section, we assume that the process X_t in $\mathbb{R}^d$ satisfies the Brownian dynamics

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}dB_t\,, \quad t > 0 \tag{1.1}$$

whose generator is given by

$$\mathcal{L}g = -\nabla V \cdot \nabla g + \frac{1}{\beta}\Delta g\,. \tag{1.2}$$

Recall the Boltzmann density

$$\pi(x) = \frac{1}{Z}\mathrm{e}^{-\beta V(x)}, \quad x \in \mathbb{R}^d, \tag{1.3}$$

where Z is a normalizing constant.

Let us denote by $L^2_\pi(\mathbb{R}^d)$ the Hilbert space induced by the weighted scalar product

$$\langle g_1, g_2\rangle_\pi = \int_{\mathbb{R}^d} g_1 g_2 \pi dx = \frac{1}{Z}\int_{\mathbb{R}^d} g_1 g_2 \mathrm{e}^{-\beta V}\,dx\,, \tag{1.4}$$

for two test functions $g_1, g_2 : \mathbb{R}^d \to \mathbb{R}$.

Lemma 1. We have

$$\langle \mathcal{L}g_1, g_2\rangle_\pi = \langle g_1, \mathcal{L}g_2\rangle_\pi = -\frac{1}{\beta}\int_{\mathbb{R}^d} \nabla g_1 \cdot \nabla g_2 \pi\,dx\,. \tag{1.5}$$

Therefore, $\mathcal{L}$ is a self-adjoint operator in $L^2_\pi(\mathbb{R}^d)$.

Proof. Let us compute

$$\begin{aligned}
\langle \mathcal{L}g_1, g_2\rangle_\pi =& \frac{1}{Z}\int_{\mathbb{R}^d}\Big(-\nabla V\cdot\nabla g_1 + \frac{1}{\beta}\Delta g_1\Big) g_2\,\mathrm{e}^{-\beta V}\,dx\\
=& \frac{1}{Z}\int_{\mathbb{R}^d}\Big[\big(-\nabla V\cdot\nabla g_1\big) g_2\,\mathrm{e}^{-\beta V} - \frac{1}{\beta}\nabla g_1\cdot\nabla\big(g_2\,\mathrm{e}^{-\beta V}\big)\Big]dx\\
=& -\frac{1}{\beta}\frac{1}{Z}\int_{\mathbb{R}^d}\nabla g_1\cdot\nabla g_2\,\mathrm{e}^{-\beta V}\,dx\\
=& -\frac{1}{\beta}\int_{\mathbb{R}^d}\nabla g_1\cdot\nabla g_2 \pi\,dx\,.
\end{aligned}$$

Switching g_1 and g_2, we obtain the first equality in (1.5). □

In particular, choosing $g_1 = g_2 = g$, we have

$$-\langle \mathcal{L}g, g\rangle_\pi = \frac{1}{\beta}\int_{\mathbb{R}^d}|\nabla g|^2\pi\,dx \geq 0\,. \tag{1.6}$$

We study the spectrum of $-\mathcal{L}$. We assume that the spectrum of $-\mathcal{L}$ consists of discrete eigenvalues. Lemma 1 and (1.6) imply that all eigenvalues of $-\mathcal{L}$ are non-negative real numbers. Denote by $\mathbf{1}$ the constant function whose value equals to 1. Then, we have $-\mathcal{L}\mathbf{1} = 0$, which implies that $\lambda = 0$ is an eigenvalue and $\mathbf{1}$ is the associated eigenfunction. Assume that all the eigenvalues are ordered as ($\lambda = 0$ is simple)

$$0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \dots\,, \tag{1.7}$$

and the corresponding eigenfunctions are $\varphi_0 = \mathbf{1}, \varphi_1, \varphi_2, \dots$. Without loss of generosity, we assume that φ_i are normalized, i.e. $|\varphi_i|_\pi^2 = \int_{\mathbb{R}^d} \varphi_i^2 \pi dx = 1$, for $i = 0, 1, 2, \cdots$, such that $\varphi_0, \varphi_1, \varphi_2, \dots$ form an orthonormal basis of $L^2_\pi(\mathbb{R}^d)$, that is, $\langle \varphi_i, \varphi_j\rangle_\pi = \delta_{ij}$, for $i, j = 0, 1, \dots$.

Recall that the semigroup

$$(T_t g)(x) = \mathbb{E}[g(X_t)|X_0 = x], \quad x \in \mathbb{R}^d\,. \tag{1.8}$$

satisfies

$$\frac{d}{dt} T_t g = \mathcal{L} T_t g\,, \tag{1.9}$$

which implies $T_t g = \mathrm{e}^{t\mathcal{L}}$. Therefore, the eigenvalues of T_t are $\nu_i = \mathrm{e}^{-\lambda_i t}$, with

$$1 = \nu_0 > \nu_1 \ge \nu_2 \ge \dots\,. \tag{1.10}$$

For a function g, with $g = \sum_{i=0}^{+\infty} \langle g, \varphi_i\rangle_\pi \varphi_i$, we have

$$\begin{aligned}
&\mathbb{E}[g(X_t)|X_0 = x]\\
=&(T_t g)(x)\\
=&(\mathrm{e}^{t\mathcal{L}} g)(x)\\
=&\sum_{i=0}^{+\infty} \langle g, \varphi_i\rangle_\pi (\mathrm{e}^{t\mathcal{L}} \varphi_i)(x)\\
=&\sum_{i=0}^{+\infty} \langle g, \varphi_i\rangle_\pi \mathrm{e}^{-\lambda_i t} \varphi_i(x)\\
=&\mathbb{E}_\pi(g) + \sum_{i=1}^{+\infty} \langle g, \varphi_i\rangle_\pi \mathrm{e}^{-\lambda_i t} \varphi_i(x)
\end{aligned} \tag{1.11}$$

Hence, the derivation above shows that $\mathbb{E}[g(X_t)|X_0 = x]$ converges to $\mathbb{E}_\pi(g)$ and the speed of convergence is dominated by the eigenvalue λ_1. The larger λ_1 is, the faster the convergence is.

Let us denote by

$$\mathcal{H}^1 := \left\{ g \in L^2_\pi(\mathbb{R}^d) \,\Big|\, \int_{\mathbb{R}^d} g\, \pi dx = 0, \int_{\mathbb{R}^d} |\nabla g|^2 \pi dx < \infty \right\}. \tag{1.12}$$

Proposition 1 (Variational principle).

$$\lambda_1 = \min_{g \in \mathcal{H}^1} \frac{\frac{1}{\beta} \int_{\mathbb{R}^d} |\nabla g|^2 \pi\, dx}{\int_{\mathbb{R}^d} g^2 \pi\, dx}\,. \tag{1.13}$$

Proof. For any $g \in \mathcal{H}^1$, we have $\langle g, \mathbf{1}\rangle_\pi = \int_{\mathbb{R}^d} g\pi dx = 0$. Let us consider the expansion $g = \sum_{i=1}^{+\infty} w_i\varphi_i$. For the denominator, using the fact that $\langle \varphi_i, \varphi_j\rangle_\pi = \delta_{ij}$, we have

$$\int_{\mathbb{R}^d} g^2\pi\, dx = \langle g, g\rangle_\pi = \langle \sum_{i=1}^{+\infty} w_i\varphi_i, \sum_{i=1}^{+\infty} w_i\varphi_i\rangle_\pi = \sum_{i=1}^{+\infty} w_i^2\,. \tag{1.14}$$

For the numerator, we have

$$\begin{aligned}
\frac{1}{\beta}\int_{\mathbb{R}^d} |\nabla g|^2\pi\, dx =& \langle -\mathcal{L}g, g\rangle_\pi \\
=& \Big\langle -\mathcal{L}\sum_{i=1}^{+\infty} w_i\varphi_i, \sum_{i=1}^{+\infty} w_i\varphi_i\Big\rangle_\pi \\
=& \Big\langle \sum_{i=1}^{+\infty} w_i(-\mathcal{L})\varphi_i, \sum_{i=1}^{+\infty} w_i\varphi_i\Big\rangle_\pi \\
=& \Big\langle \sum_{i=1}^{+\infty} w_i\lambda_i\varphi_i, \sum_{i=1}^{+\infty} w_i\varphi_i\Big\rangle_\pi \\
=& \sum_{i=1}^{+\infty} w_i^2\lambda_i \\
\geq& \lambda_1\sum_{i=1}^{+\infty} w_i^2 \\
=& \lambda_1\int_{\mathbb{R}^d} g^2\pi\, dx
\end{aligned} \tag{1.15}$$

Therefore, we have

$$\frac{1}{\beta}\int_{\mathbb{R}^d} |\nabla g|^2\pi\, dx \geq \lambda_1\sum_{i=1}^{+\infty} w_i^2 = \lambda_1\int_{\mathbb{R}^d} g^2\pi\, dx\,,$$

or

$$\lambda_1 \leq \frac{\frac{1}{\beta}\int_{\mathbb{R}^d} |\nabla g|^2\pi\, dx}{\int_{\mathbb{R}^d} g^2\pi\, dx}\,.$$

On the other hand, taking $g = \varphi_1$ and using (1.6), we have

$$\frac{1}{\beta}\int_{\mathbb{R}^d} |\nabla g|^2\pi\, dx = \langle -\mathcal{L}g, g\rangle_\pi = \lambda_1\int_{\mathbb{R}^d} g^2\pi\, dx\,,$$

which shows that the equality in (1.13) can be attained. □

1.2 Markov chains

Consider the set $\mathcal{D} = \{1, 2, \cdots, M\}$, where $M > 1$. A Markov chain is a discrete-in-time Markovian process. Denote by P_n the $M \times M$ probability matrix at nth step for $n \geq 0$, whose entries are $\mathbb{P}(x_{n+1} = j | x_n = i)$, for $i, j \in \mathcal{D}$. The fact that the row sum of P_n equals one can be expressed as

$$P_n \mathbb{1} = \mathbb{1}\,, \tag{1.16}$$

where $\mathbb{1} \in \mathbb{R}^M$ denotes the vector with all elements equal to one.

Let $\pi^{(n)}$ be the distribution of the Markov chain at nth step, that is, $\pi_i^{(n)} = \mathbb{P}(x_n = i)$, for $i \in \mathcal{D}$. In particular, $\pi^{(0)}$ is the initial distribution of the state x_0 at step $n = 0$.

Proposition 2.

$$\pi^{(n+1)} = P_n^\top \pi^{(n)}\,. \tag{1.17}$$

Proof. For $i \in \mathcal{D}$, we have

$$\begin{aligned}
\pi_i^{(n+1)} =& \mathbb{P}(x_{n+1} = i) \\
=& \sum_{j=1}^{M} \mathbb{P}(x_{n+1} = i, x_n = j) \\
=& \sum_{j=1}^{M} \mathbb{P}(x_{n+1} = i | x_n = j) \mathbb{P}(x_n = j) \\
=& \sum_{j=1}^{M} (P_n)_{ji} \pi_j^{(n)}\,.
\end{aligned}$$

□

The equation (1.17) is called master equation. It also implies that, for any $0 \leq n_1 < n$, we have

$$\pi^{(n)} = (P_{n_1} P_{n_1+1} \cdots P_{n-1})^\top \pi^{(n_1)}\,. \tag{1.18}$$

In particular, choosing $n_1 = 0$, we have

$$\pi^{(n)} = \Big(\prod_{l=0}^{n-1} P_l\Big)^\top \pi^{(0)}\,. \tag{1.19}$$

In the following, we assume that the Markov chain is time-homogeneous, i.e. the probabilities $\mathbb{P}(x_{n+1} = j | x_n = i)$ are independent of n, and we denote its probability matrix by P. In this case, (1.16) implies that $\lambda = 1$ is an eigenvalue of P and $\mathbb{1}$ is the corresponding (right) eigenvector. (1.19) becomes

$$\pi^{(n)} = \big(P^n\big)^\top \pi^{(0)}\,, \quad n \geq 0\,. \tag{1.20}$$

Definition 1. Assume that the Markov chain is time-homogeneous. A distribution π is invariant, if

$$\pi_i = \sum_{j=1}^{M} \pi_j P_{ji}\,, \quad \text{or,}\ \ \pi = P^\top \pi\,. \tag{1.21}$$

Therefore, π is a invariant distribution, if and only if π is a (non-negative) left eigenvector of P corresponding to the eigenvalue $\lambda = 1$.

Definition 2. The Markov chain is ergodic, if there is a distribution π, such that

$$\lim_{n\to+\infty} (P^n)^\top \pi^{(0)} = \pi\,, \tag{1.22}$$

for all initial distribution $\pi^{(0)}$.

Proposition 3. Let π be the limiting distribution in Definition 2, then π is the unique invariant distribution of the Markov chain.

Proof. On the one hand, choosing $\pi^{(0)} = \pi$, from (1.22) we have

$$\pi = \lim_{n\to+\infty} (P^n)^\top \pi = P^\top \lim_{n\to+\infty} (P^{n-1})^\top \pi = P^\top \pi\,.$$

Hence, π is invariant. On the other hand, let π' be an invariant distribution, then we have $(P^n)^\top \pi' = (P^{n-1})^\top \pi' = \cdots = \pi'$. Therefore,

$$\pi = \lim_{n\to+\infty} (P^n)^\top \pi' = \pi'\,,$$

which shows that π is unique. □

We say P is a primitive matrix, if there exists $n \geq 1$ such that all entries of P^n are positive. This means that, for any two states $i, j \in \mathcal{D}$, the probability of reaching j in n steps starting from i is positive.

Theorem 1 (Perron-Frobenius Theorem)**.** Assume that P is primitive. Then,

1. $\lambda = 1$ is an eigenvalue of P and any other eigenvalues λ is strictly smaller than 1.
2. $\lambda = 1$ is simple, i.e. the eigenspace associated to $\lambda = 1$ is one-dimensional.
3. There exists an eigenvector $v \in \mathbb{R}^M$ with eigenvalue $\lambda = 1$ such that all components of v are positive and $P^\top v = v$.
4. $\lim\limits_{n\to\infty} P^n = \mathbb{1} v^\top$.

The eigenvector v is called Perron eigenvector or principal eigenvector. Notice that $\mathbb{1}^\top \pi^{(0)} = 1$ for any probability distribution $\pi^{(0)}$. From the last claim in Theorem 1, we can derive

$$\lim_{n\to+\infty} (P^n)^\top \pi^{(0)} = v \mathbb{1}^\top \pi^{(0)} = v\,. \tag{1.23}$$

Therefore, Theorem 1 implies that, when P is primitive, the Markov chain is ergodic and v is the unique invariant distribution.

Let $h, g : \mathcal{D} \to \mathbb{R}$ and $\mathcal{D}' \subset \mathcal{D}$ be a subset of $\mathcal{D}$. Consider the problem

$$\begin{aligned} [(P-I)w]_i =& -h_i, \quad i \in \mathcal{D}'\,, \\ w_i =& g_i \quad i \notin \mathcal{D}'\,. \end{aligned} \tag{1.24}$$

Proposition 4. We have

$$w(x) = \mathbb{E}\Big(g(\tau_{D'}) + \sum_{n=0}^{\tau_{D'}-1} h(x_n)\Big| x_0 = x\Big), \quad x \in \mathcal{D}\,, \tag{1.25}$$

where

$$\tau_{D'} = \inf\left\{n \geq 0, x_n \notin \mathcal{D}'\right\} \tag{1.26}$$

is the first time that the Markov chain leaves $\mathcal{D}'$.

Proof. Let w be the solution to (1.24). For $x \notin \mathcal{D}'$, we have $\tau_{\mathcal{D}'} = 0$ and therefore (1.25) reduces to the second equation in (1.24). Assume now that $x \in \mathcal{D}'$ and consider the Markov chain with $x_0 = x$. It is straightforward to verify that

$$w(x_{n+1}) = w(x_n) + \sum_{j=1}^{M} \big(w(j) - w(x_n)\big)\mathbb{P}(j|x_n) + \Big(w(x_{n+1}) - \sum_{j=1}^{M} w(j)\mathbb{P}(j|x_n)\Big)\,. \tag{1.27}$$

The first equation in (1.24) implies

$$\sum_{j=1}^{M} \big(w(j) - w(x_n)\big)\mathbb{P}(j|x_n) = -h(x_n), \quad x_n \in \mathcal{D}'. \tag{1.28}$$

Therefore, we can derive

$$\begin{aligned} w(x_{n+1}) =& w(x_n) + \sum_{j=1}^{M} \big(w(j) - w(x_n)\big)\mathbb{P}(j|x_n) + \Big(w(x_{n+1}) - \sum_{j=1}^{M} w(j)\mathbb{P}(j|x_n)\Big) \\ =& w(x_n) - h(x_n) + \Big(w(x_{n+1}) - \sum_{j=1}^{M} w(j)\mathbb{P}(j|x_n)\Big)\,. \end{aligned} \tag{1.29}$$

Summing up the above equality from $n = 0$ to $n = \tau_{D'} - 1$, we obtain

$$w(x_{\tau_{D'}}) = w(x) - \sum_{n=0}^{\tau_{D'}-1} h(x_n) + \sum_{n=0}^{\tau_{D'}-1} \Big(w(x_{n+1}) - \sum_{j=1}^{M} w(j)\mathbb{P}(j|x_n)\Big)\,. \tag{1.30}$$

Taking expectation on both sides of the equality above and noticing that the mean of the last term on the right hand side vanishes, we obtain (1.25). □

Choosing $h \equiv 1$ and $g \equiv 0$, we obtain that the solution to the problem

$$\begin{aligned} [(P-I)w]_i =& -1, \quad i \in \mathcal{D}', \\ w_i =& 0 \quad i \notin \mathcal{D}', \end{aligned} \tag{1.31}$$

is given by

$$w(x) = \mathbb{E}\big(\tau_{D'} \big| x_0 = x\big), \quad x \in \mathcal{D}. \tag{1.32}$$

Definition 3 (Detailed Balnace)**.** The Markov chain satisfies detailed balance condition with a probability distribution π on $\mathcal{D}$, if

$$\pi_i P_{ij} = \pi_j P_{ji}, \quad \forall i, j \in \mathcal{D}. \tag{1.33}$$

Lemma 2. If the detailed balance condition holds, then π is invariant.

Proof. For $i \in \mathcal{D}$, from (1.33) we have $\sum_{j=1}^M \pi_j P_{ji} = \sum_{j=1}^M \pi_i P_{ij} = \pi_i$, which implies $P^\top \pi = \pi$. Therefore, π is invariant. $\square$

Define the diagonal matrix $\Pi = \mathrm{diag}\{\pi_1, \pi_2, \dots, \pi_M\}$. Introduce the scalar product

$$\langle u, v\rangle_\pi = u^\top \Pi v = \sum_{i=1}^M u_i \pi_i v_i, \quad u, v \in \mathbb{R}^M. \tag{1.34}$$

Lemma 3. The following claims are equivalent.

1. The Markov chain satisfies detailed balance condition (1.33).
2. $\Pi P = P^\top \Pi$.
3. $\langle Pu, v\rangle_\pi = \langle u, Pv\rangle_\pi$, for all $u, v \in \mathbb{R}^M$.

The second claim can be written as

$$\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} = \Pi^{-\frac{1}{2}} P^\top \Pi^{\frac{1}{2}}. \tag{1.35}$$

Therefore, $\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}}$ is symmetric and its all eigenvalues are real. From this we know that all eigenvalues of P are real as well. Denote them by (assume P is primitive)

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_M.$$

The corresponding left eigenvectors are $v_1 = \pi, v_2, \dots, v_M$. We can verify that

$$\begin{aligned} & v_1^\top \mathbb{1} = 1 \\ & v_j^\top \mathbb{1} = v_j^\top P \mathbb{1} = \lambda_j v_j \mathbb{1} \Longrightarrow v_j^\top \mathbb{1} = 0, \quad j > 1. \end{aligned} \tag{1.36}$$

Therefore, for any distribution $\pi^{(0)} \in \mathbb{R}^M$, we can write it as $\pi^{(0)} = \pi + \sum_{i=2}^M \alpha_i v_i$, and we have

$$(P^n)^\top \pi^{(0)} = \pi + \sum_{i=2}^M \alpha_i \lambda_i^n v_i = \pi + \sum_{i=2}^M \alpha_i \lambda_i^n v_i. \tag{1.37}$$

Since $|\lambda_2|, \cdots, |\lambda_M|$ are all smaller than 1, we know that (1.22) holds. Moreover, the speed of convergence in (1.22) is determined by the magnitude of the eigenvalue $\lambda_2 < 1$. The smaller λ_2 is, the faster the convergence is.

The following result shows the connection between the left and right eigenvectors.

Lemma 4. v is a left eigenvector of P with eigenvalue λ if and only if $\Pi^{-1}v$ is a right eigenvector of P with the same eigenvalue λ.

Proof. Using the second claim in Lemma 3, we have

$$\begin{aligned} & P^\top v = \lambda v \\ \Longleftrightarrow & (P^\top \Pi)\Pi^{-1} v = \lambda \Pi(\Pi^{-1} v) \\ \Longleftrightarrow & (\Pi P)\Pi^{-1} v = \lambda \Pi(\Pi^{-1} v) \\ \Longleftrightarrow & P(\Pi^{-1} v) = \lambda(\Pi^{-1} v) \end{aligned} \tag{1.38}$$

□

Example. $\mathcal{D} = \{1, 2\}$.

- Let $p_{11} = 0, p_{12} = 1, p_{21} = 1, p_{22} = 0$. Accordingly,

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.39}$$

  The invariant distribution is $\pi = (\frac{1}{2}, \frac{1}{2})$.

- Let $p_{11} = 0, p_{12} = 1, p_{21} = p_{22} = \frac{1}{2}$. Accordingly,

$$P = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{1.40}$$

  The invariant distribution is $\pi = (\frac{1}{3}, \frac{2}{3})$.

- Let $p_{11} = 1, p_{12} = 0, p_{21} = 0, p_{22} = 1$. Accordingly,

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1.41}$$

  The Markov chain is not ergodic and any distribution is invariant.

Example (Random walk on graphs)**.** Consider a graph with nodes $\mathcal{D} = \{1, 2, \ldots, M\}$ and set of edges E. We write $i \sim j$ if there is an edge between nodes i and j. The transition probability density is

$$P_{ij} = \begin{cases} d_i^{-1}, & i \sim j \\ 0, & \text{otherwise} \end{cases} \tag{1.42}$$

where d_i is the number of nodes connected to i (i.e. degree of node i). The Markov chain satisfies the detailed balance condition with the distribution π, which is defined as $\pi_i = \frac{d_i}{2|E|}$, where $|E|$ is the number of edges of the graph.