CONTINUOUS DYNAMICS AND TRANSFER OPERATOR DESCRIPTION

Up to this point, we have considered Markov chains on a finite state space and their analysis, and we have formulated methods to estimate a finite Markov model from data that comes from continuous dynamics. Now, we aim at formulating a framework to assess the approximation quality of such a model. The first step we need to take is to extend the concept of a Markov chain to a continuous state space. For the rest of this lecture, let Ω be a continuous space equipped with a sigma-algebra Σ, s.t. (Ω, Σ) is a measurable space.

1. Markov chains on continuous state space

In the finite setting, a Markov chain could be completely described by a matrix of conditional transition probabilities. Moreover, the Markov property could be expressed by the fact that path probabilities could be evaluated using only the transition matrix and the initial distribution:

$$\mathbb{P}_{p_0}(X_0 = s_0, \ldots, X_n = s_n) \quad = \quad p_0(s_0)\mathbf{T}(s_0, s_1)\ldots\mathbf{T}(s_{n-1}, s_n).$$

In the continuous setting, the conditional transition probabilities cannot be described by a matrix, but by a **transition kernel:**

Definition 1. A map $p : \Omega \times \Sigma \to [0, 1]$ is called **transition kernel** if

(1) for every $x \in \Omega$ the map $p(x, \cdot)$ is a probability measure on Σ, in particular $p(x, \Omega) = 1$.
(2) for every $A \in \Sigma$ the map $p(\cdot, A)$ is a measurable function on Ω.

Just as in the discrete case, a transition kernel and an initial distribution determine a Markov chain on Ω. In this setting, a Markov chain can be understood as a random variable on the product space $\Omega^{\mathbb{N}} = \prod_{n\in\mathbb{N}} \Omega$, which forms a measurable space if joined with the product sigma-algebra $\Sigma^{\mathbb{N}}$. Thus, a Markov chain is a random variable X that outputs infinite tuples of events in Ω:

$$X \quad = \quad (x_n)_{n\in\mathbb{N}} \in \Omega^{\mathbb{N}},$$

and the sequence of individual outputs are the projections onto the n-th entry,

$$X_n(X) \quad = \quad x_n \in \Omega.$$

The fact that a Markov chain is determined by the transition kernel and an initial distribution can be formulated as follows [4, Proposition 2.10]:

Lemma 2. *For any transition kernel p and any probability measure μ, there is a corresponding probability measure $\mathbb{P}_\mu$ on $\left(\Omega^{\mathbb{N}}, \Sigma^{\mathbb{N}}\right)$, s.t.*

$$\mathbb{P}_\mu\left(X_0 \in A_0, \ldots, X_n \in A_n\right) \quad = \quad \int_{A_0} \ldots \int_{A_{n-1}} p(y_{n-1}, A_n)p(y_{n-2}, dy_{n-1})\ldots p(y_0, dy_1)\mu(dy_0).$$

In particular, it follows that

$$\mathbb{P}_\mu(X_0 \in A,\, X_1 \in B) \;=\; \int_A p(x,B)\mu(dx)$$

and that we can inductively define a family of n-step transition kernels by

$$\begin{aligned} p_{n+1}(x,A) &= \int_\Omega p_n(y,A)p(x,dy), \\ p_1(x,A) &= p(x,A). \end{aligned}$$

2. Transfer Operator

For the finite state space, it turned out very helpful to assume an ensemble perspective, i.e. we have investigated the evolution of probability vectors in time. This evolution was given by the equation

$$\mathbf{p}_1^T \;=\; \mathbf{p}_0^T\mathbf{T}.$$

In this equation, the initial probability vector $\mathbf{p}_0$ is propagated in time by the action of a linear map $\mathbf{T}^T$. Let us imitate this idea to the continuous case. Instead of a vector, we need to consider a probability density $p_0 \in L^1(\mu)$, i.e. a function p_0 which satisfies

$$\begin{aligned} p_0 &\geq 0, \\ \int_\Omega p_0(x)\mu(dx) &= 1. \end{aligned}$$

From Lemma 2, we know can compute the probability measure after one step:

$$\mathbb{P}_{p_0}(X_1 \in A) \;=\; \int_\Omega p(x,A)p_0(x)\mu(dx).$$

Now, we are looking for a linear map $\mathcal{T}$, s.t. the application of $\mathcal{T}$ to the density p_0 provides a density for the probability measure after one step:

$$\begin{aligned} \int_A \left[\mathcal{T}p_0\right](x)\mu(dx) &= \mathcal{P}_{p_0}(X_1 \in A) \\ &= \int_\Omega p(x,A)p_0(x)\mu(dx). \end{aligned} \tag{2.1}$$

This will serve as the defining equation for the **transfer operator**, the central object of this lecture:

Definition 3. Given a probability measure μ, an operator $\mathcal{T}$ which satisfies Eq. (2.1) for all probability densities $p_0 \in L^1(\mu)$ is called **transfer operator.**

We are now concerned with the existence of a transfer operator for the measure μ. The first step is to observe that Eq. (2.1) is a special case of the following equation, for $g = \chi_A$, the indicator function of set A:

$$\int_\Omega g(x)\left[\mathcal{T}p_0\right](x)\mu(dx) = \int_\Omega \left[\mathcal{U}g\right](x)p_0(x)\mu(dx), \tag{2.2}$$

$$\left[\mathcal{U}g\right](x) = \int_\Omega g(y)p(x,dy). \tag{2.3}$$

But this is an adjoint equation:

$$\begin{aligned}\langle g,\, [\mathcal{T}p_0]\rangle_\mu &= \langle [\mathcal{U}g]\,,\, p_0\rangle_\mu \\ \langle u,\, v\rangle_\mu &= \int_\Omega u(x)w(x)\mu(dx).\end{aligned}$$

From functional analysis, it is known that for a bounded linear operator on $L^1(\mu)$, there is an adjoint operator, which is a bounded linear operator on the space $L^\infty(\mu)$. The reverse statement is, in general, not true! In this special case, however, it is possible to show that the operator $\mathcal{U}$ possesses an adjoint operator on $L^1(\mu)$. The procedure to show the existence of the transfer operator is therefore as follows, see Refs. [2, 3] for details.

- Show that $\mathcal{U}$ as defined by Eq. (2.3) is well-defined and bounded on the space $L^\infty(\mu)$. It turns out that this is possible under a mild technical assumption: We need to assume that the transition kernel p is **μ-compatible**, which means that for every zero set A, i.e. $\mu(A) = 0$, it follows that $p(x, A) = 0$ μ-almost everywhere.
- Show that an adjoint operator $\mathcal{T}$ exists on the space $L^1(\mu)$, satisfying (2.2).
- We have already argued that this adjoint operator is the transfer operator.

In fact, it can be shown that μ-compatibility is also necessary for the existence of a transfer operator, i.e. the transfer operator exists if and only if p is μ-compatible. Finally, we observe that in the finite setting, the adjoint operator $\mathcal{U}$ reduces to right multiplication of the transition matrix. More importantly, it also possesses a physical meaning:

$$\begin{aligned}\left[\mathcal{U}g\right](x) &= \int_\Omega g(y)p(x,dy) \\ &= \mathbb{E}_x\left[g(X_1)\right],\end{aligned}$$

i.e. $\mathcal{U}$ applied to g provides the expectation value of the function g after starting the process from point x and performing one step of the chain. The operator $\mathcal{U}$ is frequently called the **Koopman operator**.

3. Stationary Measure

We can also define the concept of a stationary measure in the framework we have now discussed. A probability measure μ is called **stationary** if

$$\int_A \mu(dx) = \int_\Omega p(x,A)\mu(dx)$$

is true for all sets $A \in \Sigma$. We conclude at once that

$$\begin{aligned}\mathbb{P}_\mu(X_1 \in A) &= \int_\Omega p(x,A)\mu(dx) \\ &= \mu(A),\end{aligned}$$

i.e. the probability distribution of the process is unchanged after one step if the chain is started from μ. By induction, it can be shown that the same holds for all $n \in \mathbb{N}$, justifying the name stationary measure. Also, if μ is a stationary measure for the transition kernel p, then p is μ-compatible, as $\mu(A) = 0$ implies

$$\begin{aligned}0 &= \mu(A) \\ &= \int_\Omega p(x,A)\mu(dx).\end{aligned}$$

As $p(x, A)$ is a non-negative function, it must be zero μ-almost everywhere. We have thus shown that for a stationary measure, the transfer operator always exists, and we will from now on shift our attention to the study of transfer operatos w.r.t. to a stationary measure μ.

So far, we have defined the transfer operator on the space $L^1(\mu)$, which is suitable for the study of probability densities. From the perspective of spectral theory, it would be desirable to study $\mathcal{T}$ on the Hilbert space $L^2(\mu)$ of square-integrable functions. For a stationary measure μ, we will see that this is possible indeed. First, let us observe that

Lemma 4. *For a stationary measure μ, we have $L^\infty(\mu) \subset L^2(\mu) \subset L^1(\mu)$:*

Proof. The first inclusion follows by

$$\begin{aligned}\int_\Omega |f(x)|^2\mu(dx) &\leq \|f\|_\infty^2 \int_\Omega \mu(dx) \\ &= \|f\|_\infty^2\end{aligned}$$

for $f \in L^\infty(\mu)$, while the second follows from Hölder's inequality:

$$\begin{aligned}\int_\Omega |f(x)|\mu(dx) &= \int_\Omega |f(x)\cdot 1|\mu(dx) \\ &\leq \|f\|_2\|1\|_2 \\ &= \|f\|_2\end{aligned}$$

for $f \in L^2(\mu)$. □

It follows that the transfer operator can be applied to square-integrable functions, but we do not know yet if the result will still be contained in $L^2(\mu)$. The next step is to show that the Koopman operator $\mathcal{U}$ can be extended from $L^\infty(\mu)$ to $L^2(\mu)$ [1]:

Lemma 5. *For a stationary measure, the Koopman operator $\mathcal{U}$ is also well-defined for $f \in L^2(\mu)$:*

Proof. The proof uses Jensen's inequality, which states that for a real-valued and integrable function h on a probabilty space (Ω, Σ, μ) and a convex function φ of the real line, we have

$$\varphi\left[\int_\Omega h(x)\mu(dx)\right] \leq \int_\Omega \varphi(h(x))\mu(dx).$$

Now, consider the application of the Koopman operator to $f \in L^2(\mu)$:

$$\begin{aligned}
\|\mathcal{U}f\|_2^2 &= \int_\Omega \left[\int_\Omega f(y)p(x,dy)\right]^2 \mu(dx) \\
&= \int_\Omega \left[\mathbb{E}_{p(x,\cdot)}\left(f(X_1)\right)\right]^2 \mu(dx) \\
&\leq \int_\Omega \int_\Omega f(y)^2 p(x,dy)\mu(dx) \\
&= \int_\Omega f(y)^2 \mu(dy) \\
&= \|f\|_2^2.
\end{aligned}$$

Thus, the Koopman operator maps functions $f \in L^2(\mu)$ to functions $\mathcal{U}f \in L^2(\mu)$ and has operator norm ≤ 1. □

Lemma 6. *For a stationary measure, the transfer operator is a well-defined operator* $\mathcal{T} : L^2(\mu) \to L^2(\mu)$.

Proof. As the Koopman operator is well-defined on $L^2(\mu)$, it possesses a unique adjoint operator $\mathcal{T}^* : L^2(\mu) \to L^2(\mu)$, such that

$$\langle \mathcal{U}f, g\rangle_\mu = \langle f, \mathcal{T}^* g\rangle_\mu$$

holds for all $f, g \in L^2(\mu)$. For $f \in L^\infty(\mu)$, this implies

$$\begin{aligned}
\langle f, \mathcal{T}^* g\rangle_\mu &= \langle \mathcal{U}f, g\rangle_\mu \\
&= \langle f, \mathcal{T}g\rangle_\mu,
\end{aligned}$$

as $\mathcal{T}$ is the adjoint of the Koopman operator on $L^\infty(\mu)$. In particular, choosing $f = \chi_A \in L^\infty(\mu)$ for any set $A \in \Sigma$ implies

$$\int_A \mathcal{T}g(x)\mu(dx) = \int_A \mathcal{T}^* g(x)\mu(dx)$$

for all $A \in \Sigma$, thus $\mathcal{T}g = \mathcal{T}^* g$ μ-almost everywhere. Therefore, we have that $\mathcal{T}g \in L^2(\mu)$. □

4. Reversibility

The concept of reversibility leads to a self-adjoint operator on the Hilbert space $L^2(\mu)$.

Definition 7. A transition kernel p is called **reversible** with respect to a measure μ if

$$\begin{aligned} \int_A p(x,B)\mu(dx) &= \int_B p(x,A)\mu(dx) \\ \mathbb{P}_\mu(X_0 \in A,\, X_1 \in B) &= \mathbb{P}_\mu(X_0 \in B,\, X_1 \in A). \end{aligned}$$

If we choose $B = \Omega$, we conclude that μ is automatically stationary if p is reversible.

Lemma 8. *For a reversible transition kernel p with stationary measure μ, the transfer operator $\mathcal{T}$ is self-adjoint on $L^2(\mu)$.*

Proof. We verify that

$$\begin{aligned} \langle \mathcal{T}f,\, g\rangle_\mu &= \int_\Omega f(x) \int_\Omega g(y)p(x,dy)\mu(dx) \\ &= \int_\Omega f(x) \int_\Omega g(y)p(y,dx)\mu(dy) \\ &= \langle f,\, \mathcal{T}g\rangle_\mu. \end{aligned}$$

□

5. Spectral Decomposition

Finally, we can get back to the spectral analysis of a transfer operator. Here, we will make three assumptions that simplify the setting. First, let us assume that both the stationary measure μ as well as the transition kernel possess a density with respect to ordinary Lebesgue measure:

$$\begin{aligned} \mu(A) &= \int_A \mu(x)\,\mathrm{dx}, \\ p(x,A) &= \int_A p(x,y)\,\mathrm{dy}. \end{aligned}$$

Reversibility then implies that $\mu(x)p(x,y) = \mu(y)p(y,x)$ pointwise for $x, y \in \Omega$. If the density of μ is positive everywhere in Ω, we can verify that the transfer operator is given by the expression

$$\mathcal{T}f(y) = \frac{1}{\mu(y)} \int_\Omega p(x,y)\mu(x)f(x)\,\mathrm{dx}. \tag{5.1}$$

$$= \int_\Omega p(y,x)f(x)\,\mathrm{dx}. \tag{5.2}$$

This can be checked by inserting Eq. (5.1) into the duality pair

$$\langle \mathcal{T}f,\, g\rangle_\mu = \langle f,\, \mathcal{U}g\rangle_\mu$$

for $f \in L^1(\mu)$, $g \in L^\infty(\mu)$. It is easy to check that the constant function $\psi_1 = \chi_\Omega$ is an eigenfunction of $\mathcal{T}$ with eigenvalue one. We assume that the eigenvalue $\lambda_1 = 1$ is simple and there is no eigenvalue equal to -1. If we make the additional assumption that the transition kernel is square-integrable in both x and y, it follows that $\mathcal{T}$ is a Hilbert-Schmidt-operator. Therefore, it is not only self-adjoint, but also compact. All of this implies that $\mathcal{T}$ possesses a sequence of eigenvalues $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots$

and mutally orthonormal eigenfunctions ψ_m, such that the action of $\mathcal{T}$ can be decomposed completely by

$$\begin{aligned} \mathcal{T}f &= \sum_{m=1}^{\infty} \langle f, \psi_m \rangle_\mu \psi_m \\ &= \langle f \rangle_\mu + \sum_{m=2}^{\infty} \langle f, \psi_m \rangle_\mu \psi_m. \end{aligned}$$

References

[1] J.R. Baxter and Jeffrey S. Rosenthal. Rates of convergence for everywhere-positive Markov chains. *Statist. Probab. Lett.*, 22(4):333 – 338, 1995.

[2] Adam Nielsen. Von Femtosekunden zu Minuten. Master's thesis, Freie Universität Berlin, 2012.

[3] Adam Nielsen. *Computation Schemes for Transfer Operators*. PhD thesis, Freie Universität Berlin, 2015.

[4] Daniel Revuz. *Markov Chains*. North-Holland Mathematical Library, 1984.