

Chapter 6: Spectral properties of Markov Chains

In this chapter we focus on finite state Markov chains, so that the linear operator P can just be represented by a matrix, and we can use some known spectral properties of finite-dimensional linear operators. We have seen that $P\mathbf{1} = \mathbf{1}$, namely the eigenvalue 1 is in the spectrum of P . And indeed P also admits a left eigenvector, the invariant probability m such that $mP = m$.

Theorem 0.1 (Perron-Frobenius for Stochastic Matrices). *Let P be the transition matrix of a Markov chain on a finite state space S . Then:*

- The constant function $\mathbf{1}$ (regarded as a vector with all entries equal to 1) is an eigenvector with eigenvalue $\lambda_0 = 1$. Any invariant probability is a left eigenvector of P with positive entries, and associated to the same eigenvalue.*
- Any eigenvalue $\lambda \in \text{Spec}(P)$ satisfies $|\lambda| \leq 1$.*

Assume now that the chain is irreducible.

- The eigenvalue $\lambda_0 = 1$ is simple, and thus the corresponding left eigenvector is, up to a multiplicative constant, the unique invariant probability measure m . All components of m are strictly positive.*

Assume now that the chain is also aperiodic.

- All other eigenvalues λ of P satisfy $|\lambda| < 1$.*

Proof.

- follows directly from $\sum_y p_{x,y} = 1$ for all $x \in S$, which means $P\mathbf{1} = \mathbf{1}$. Similarly $mP = m$ for any invariant probability m .*
- Let $\lambda \in \text{Spec}(P)$ be any eigenvalue and v an associated eigenvector (with possibly complex entries), $Pv = \lambda v$. For $t \geq 0$, $P^t v = \lambda^t v$ and thus (we consider a generic t for later use of this inequality, otherwise we may just take $t = 1$ here)*

$$|\lambda^t v_x| = |(P^t v)_x| = \left| \sum_{y \in S} p_{x,y}^{(t)} v_y \right| \leq \sum_{y \in S} p_{x,y}^{(t)} \sup_z |v_z| = \sup_z |v_z| \quad (1)$$

Optimizing over x we get $|\lambda^t| \sup_x |v_x| \leq \sup_z |v_z|$, namely $|\lambda| \leq 1$ since $v \neq 0$. In other words, the spectral radius of P is 1.

- Fix an eigenvector v associated to the eigenvalue $\lambda_0 = 1$. Take x such that $|v_x| = \sup_z |v_z|$ and normalize the eigenvector so that $v_x = 1$. Then we have for each $x \in S$*

$$1 = v_x = (P^t v)_x = \sum_{y \in S} p_{x,y}^{(t)} v_y \leq \sup_z |v_z| = 1 \quad (2)$$

Thus all inequalities are equalities, so $v_y = v_x = 1$ whenever $p_{x,y}^{(t)} > 0$. But the chain is irreducible, so for each $y \in S$ there is some $t \geq 0$ for which this holds, and therefore $v = \mathbf{1}$.

- d. Fix now an eigenvector v associated to an eigenvalue $|\lambda| = 1$. As in c., we can normalize v so that $v_x = \sup_z |v_z| = 1$. We want to prove (assuming irreducibility and aperiodicity), that $v = \mathbf{1}$, in particular $\lambda = 1$. The final inequality in b. reads, for $|\lambda| = 1$, $1 \leq \sup_z |v_z| = 1$, thus all inequalities in Equation 1 are equalities. But then, whenever $p_{x,y}^{(t)}, p_{x,z}^{(t)} > 0$, we have that $v_y = v_x$. Since the chain is aperiodic and irreducible, for t large enough all the transition probabilities are strictly positive, and therefore $v_y = v_x = 1$ for all $y \in S$. \square

The Perron-Frobenius theorem provides a powerful alternative way to understand convergence to the stationary distribution. If P is diagonalizable, we can write its spectral decomposition. For any initial distribution π , the distribution at time t is πP^t . The theorem implies that as $t \rightarrow \infty$, the terms associated with eigenvalues $|\lambda| < 1$ decay to zero exponentially fast, leaving only the projection onto the stationary distribution.

Exponential convergence for Finite State Chains

We have seen that, for irreducible positive-recurrent aperiodic chains the law of X_t converges to the invariant measure as $t \rightarrow \infty$. In particular, this occurs for irreducible finite (thus positive-recurrent) aperiodic chains. For finite chains however, we can have a sharper convergence result. The following proposition shows that, if λ_1 is the eigenvalue (different from $\lambda_0 = 1$) of P with largest absolute value, then

$$\sum_y |\mathbb{P}_x(X_t = y) - m_y| \simeq |\lambda_1|^{t(1+o(1))}$$

that is, the smaller is $|\lambda_1|$, the faster the law of X_t converges to m .

Theorem 0.2. *Let P be the transition matrix of an irreducible aperiodic Markov chain on a finite state space S (with at least two elements) with invariant measure m , and let $\mu_t^x := (\delta_x P^t)$ be the law of X_t when $X_0 = x$. Let $\varrho := -\sup_{\lambda \in \text{Spec}(P) \setminus \{1\}} \log |\lambda| > 0$. Then*

$$\sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log \|\mu_t^x - m\|_{TV} := \sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log \left(\sum_y |p_{x,y}^{(t)} - m_y| \right) = -\varrho$$

Proof. Since $\lambda_0 = 1$ is simple, we can write $P = P_0 + Q$, where P_0 is a rank-one projection operator satisfying $\mu P = m$ for all $\mu \in \mathcal{P}(S)$, and $P_0 Q = Q P_0 = 0$. In other words we can write P in Jordan form as

$$P = V \begin{pmatrix} 1 & 0 \\ 0 & J \end{pmatrix} V^{-1}, \quad P_0 = V \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} V^{-1}, \quad Q = V \begin{pmatrix} 0 & 0 \\ 0 & J \end{pmatrix} V^{-1}.$$

and clearly P_0 and Q satisfy the aforementioned properties. In particular $P^t = P_0 + Q^t$ for $t \geq 1$ and $|Q_{x,y}^t| \leq C t^{|S|} e^{-t\varrho}$, since $e^{-\varrho}$ is the modulus of the largest eigenvalue of Q . We obtain $\delta_x P^t = \delta_x P_0 + \delta_x Q^t = m + \delta_x Q^t$. Thus

$$\sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log \|\mu_t^x - m\|_{TV} \leq \sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log (|S| \sup_y |Q_{x,y}^t|) \leq \sup_{x,y} \lim_{t \rightarrow \infty} \frac{1}{t} \log (|S| C t^{|S|} e^{-t\varrho}) = -\varrho \quad (3)$$

On the other hand, the set $(\delta_x)_{x \in S}$ represent $|S|$ linearly independent vectors. Thus (since we assumed $|S| \geq 2$) there exists at least one $x \in S$ such that δ_x has non-zero component along an eigenvector associated to an eigenvalue λ_1 of maximal modulus. It is then easy to check that for such an x , equality is achieved in Equation 3, since $|\delta_x Q^t|$ behaves as $c|\lambda_1|^t$ in this case. \square

Additional Exercises

Exercise 0.1. Consider a positive-recurrent aperiodic chain. If $f: S \rightarrow \mathbb{R}$ is such that $m(f) = 0$, then $m(Pf) = 0$. In other words, the restriction \hat{P} of the operator P on functions with vanishing m -mean is well-defined. Repeat the previous proof using directly the operator \hat{P} .

Exercise 0.2. Find an example of a Markov chain where there exists an $x \in S$ for which $\lim_{t \rightarrow \infty} \frac{1}{t} \log \|\mu_t^x - m\|_{TV} < \rho$. *Hint: it is enough to consider a space with 3 points.*

Exercise 0.3. Let \mathbf{X} be an aperiodic Markov chain on a state space S . Let $\xi \in S$ be an *absorbing* point, and let $S' = S \setminus \{\xi\}$. Assume that $x \leftrightarrow y$ for all $x, y \in S'$, $x \rightarrow \xi$ for all $x \in S$, but $p_{\xi, \xi} = 1$, so $\xi \nleftrightarrow x$ for $x \in S'$. For instance ξ represents that state at which a random algorithm terminates.

Let $\tau \equiv \tau_\xi$ be the hitting time of ξ , and let $Q := (p_{x,y})_{x,y \in S'}$. Prove that there exists a probability measure $m \in \mathcal{P}(S')$ such that for all $x, y \in S'$

$$\lim_t \mathbb{P}_x(X_t = y | \tau > t) = m_y$$

Exercise 0.4. Let P be the Markov operator of an irreducible Markov Chain on a finite state space S . Let m be its invariant measure, and P^\dagger the adjoint operator of P in $L^2(m)$.

- Check that if P, Q are Markov operators, then PQ and $\alpha P + (1 - \alpha)Q$ are Markov operators. In particular PP^\dagger and P^2 are Markov operators.
- Find the invariant measure of PP^\dagger and P^2 .
- Give an example where P is irreducible, but P^2 and PP^\dagger are not irreducible.
- Prove that, if P is irreducible and aperiodic, then P^2 is irreducible and aperiodic.
- Prove that PP^\dagger is aperiodic and each communicating class is closed.
- Assume that P is irreducible and aperiodic. Let X_t and Y_t be Markov Chains with Markov operators P^2 and PP^\dagger respectively, prove that X_t converges to its “equilibrium” faster than Y_t in the following sense. Let μ_t^x (respectively ν_t^x) be the law of X_t when $X_0 = x$ (respectively Y_t when $Y_0 = x$). Prove that

$$\sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log \|\mu_t^x - m\|_{TV} \leq \sup_x \lim_{t \rightarrow \infty} \frac{1}{t} \log \|\nu_t^x - m\|_{TV}$$

Solution

- Let $R = PQ$, thus $r_{x,z} = \sum_y p_{x,y} q_{y,z}$. And exchanging summations $r_{x,z} \geq 0$ and $\sum_z r_{x,z} = \sum_y p_{x,y} \sum_z q_{y,z} = \sum_y p_{x,y} = 1$. As for the convex combination, the proof is immediate by linearity.
- The invariant measure of all these operators is m . In general, if P, Q have the same invariant measure, then $m((PQ)f) = m(P(Qf)) = m(Qf) = m(f)$. So m is invariant for PQ .
- We can take a unique cycle $S = \mathbb{Z}_{2k}$ with an even number of points, P corresponds to moving clockwise with probability 1, $p_{x,x+1} = 1$ (sums are understood $(\text{mod } 2)k$). Then P is irreducible since there is a path joining any two points. PP^\dagger corresponds to making one step clockwise and one counter-clockwise, that is $PP^\dagger = I$, the walk just does not move. P^2 corresponds to making two steps clockwise $p_{x,x+2}^{(2)} = 1$. So there is not path joining points with different parity.
- For an irreducible aperiodic chain on a finite state space, there exists T large enough so that $p_{x,y}^{(t)} > 0$ for all $t > T$ and $x, y \in S$. In particular the same holds true, changing T to $T' = \lceil T/2 \rceil$, for the entries of P^2 .
- Let $(q_{x,y})$ denote the entries of $Q = PP^\dagger$. $q_{x,z} > 0$ iff there exists $y \in S$ such that $p_{x,y} p_{z,y} > 0$. In particular $q_{x,z} > 0$ iff $q_{z,x} > 0$ and $q_{x,x} > 0$ for all $x \in S$ (since each point x has at least one *successor* y_x such that $p_{x,y_x} > 0$). Therefore each Q -communicating class is closed.
- By the [exponential convergence theorem](#), we need to check that $|\lambda_1(P^2)| \leq |\lambda_1(PP^\dagger)|$, where $\lambda_1(Q)$ is

the largest (by modulus) eigenvalue of an irreducible Markov operator Q , except $\lambda_0 = 1$; or equivalently, the largest eigenvalue of Q restricted to functions f with $m(f) = 0$, where m is the invariant measure of Q .

For P^2 we have $\lambda_1(P^2) = \lambda_1(P)^2 = \overline{\lambda_1(P^\dagger)}^2$. Take f_1 an associated (complex) eigenfunction (with $m(f_1) = 0$) of P^\dagger , so that $P^\dagger f_1 = \lambda_1(P^\dagger) f_1$. Then, since PP^\dagger is symmetric, $\lambda_1(PP^\dagger) > 0$ and thus denoting $\langle \cdot, \cdot \rangle$ the Hermitian product in $L^2(m)$

$$|\lambda_1(PP^\dagger)| = \sup_{f : m(f)=0} \frac{\langle f, PP^\dagger f \rangle}{m(|f|^2)} = \sup_{f : m(f)=0} \frac{\langle P^\dagger f, P^\dagger f \rangle}{m(|f|^2)} \geq \frac{\langle P^\dagger f_1, P^\dagger f_1 \rangle}{m(|f_1|^2)} = \overline{\lambda_1(P^\dagger)} \lambda_1(P^\dagger) = |\lambda_1(P^2)|$$

i Abstraction

If S is countable (or more generally a Polish space), the previous theorem may fail. This is due to the fact that eigenvalues can accumulate around one point, or so to speak, one eigenvalue may have infinite multiplicity. An important class of operators for which this cannot happen are **compact operators**.

Consider an irreducible positive-recurrent aperiodic chain on a countable space S . Then it admits a unique invariant probability m and P is a **contraction** on $P : L^2(m) \rightarrow L^2(m)$. In particular it maps the closed ball $B_1 := \{f : \|f\|_{L^2(m)} \leq 1\}$ to itself. P is called **compact** (linear operator) if it maps the closed ball B_1 to a **compact** subset of $L^2(m)$ (this is always the case if S is finite since B_1 itself is compact in this case). If P is compact, we also have $\varrho := -\sup_{\lambda \in \text{Spec}(P) \setminus \{1\}} \log |\lambda| > 0$ and Theorem 0.2 carries over to the case of a countable state space with compact operator P (since the spectrum of compact operators consists of a countable set of eigenvalues without accumulation points except at 0, each eigenvalue having finite multiplicity, so the proof is basically the same).

Extensions of this result are possible, in particular if P is **quasi-compact**, the statement also holds true. This is particularly useful in dynamical systems. Let us describe quickly this case, assuming S countable to avoid discussing measurability issues.

If we have a map $\phi : S \rightarrow S$, we can consider $(X_0, X_1 = \phi(X_0), X_2 = \phi(X_1), \dots)$ as a Markov chain, although a rather specific one: all the randomness comes from X_0 , since ϕ is just deterministic. Here the transition probability is just

$$p_{x,y} = \delta_{\phi(x)}(y) = \begin{cases} 1 & \text{if } y = \phi(x) \\ 0 & \text{otherwise} \end{cases}$$

If m is an invariant probability, we know that the adjoint operator P^\dagger is also the transition operator of a Markov chain. A classical tool to study mixing of dynamical systems (e.g. exponential convergence to the invariant probability m) is indeed the quasi-compactness of P^\dagger .