1. **(9 pt.) Fundamental Theorem of Markov Chains: A Special Case**

Let $X_0, X_1, \ldots$ be a Markov chain over n states (labeled $1, 2, \ldots, n$) with transition matrix $P \in \mathbb{R}^{n\times n}$, i.e., for any $t \geq 0$, $\Pr[X_{t+1} = j | X_t = i] = P_{ij}$. In addition, we assume that $P_{ij} > 0$ for all $i, j \in [n]$, and define $p_{\min} \coloneqq \min_{i,j\in[n]} P_{ij} > 0$. In this problem, we will prove part of the fundamental theorem of Markov chains for this special case. In particular, we will show that there exists a unique stationary distribution π such that for all $i, j \in [n]$,

$$\lim_{t\to+\infty} \Pr[X_t = j | X_0 = i] = \pi_j.$$

(a) **(2 pt.)** As a warmup, show that the assumption $P_{ij} > 0$ for all $i, j \in [n]$ implies that the Markov chain is irreducible and aperiodic. Thus, the assumption that we made is not weaker than the one in the original theorem.

(b) **(2 pt.)** Let $a = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$ be a row vector that satisfies $\sum_{i=1}^n a_i = 0$. Prove that $\|aP\|_1 \leq (1 - np_{\min}/2)\|a\|_1$.

[**HINT:** *You can use the following fact: For vectors* $a, b \in \mathbb{R}^n$ *satisfying* $\sum_{i=1}^n a_i = 0$ *and* $\min_{i\in[n]} b_i \geq \epsilon > 0$, $|\sum_{i=1}^n a_i b_i| \leq \sum_{i=1}^n |a_i| b_i - \frac{\epsilon}{2} \sum_{i=1}^n |a_i|$.]

(c) **(3 pt.)** Prove that there exists an n-dimensional row vector $\pi = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \end{bmatrix}$ such that: (1) $\pi = \pi P$; (2) $\sum_{i=1}^n \pi_i = 1$.

[**HINT:** *First prove the existence of a non-zero vector* π *satisfying* $\pi = \pi P$*, and then show that the second condition can be satisfied by scaling* π*. For the first step, you may use the following fact without proof: if* λ *is an eigenvalue of a square matrix* A*,* λ *is also an eigenvalue of* A^T*. Part 1b might be helpful for the second step.*]

(d) **(2 pt.)** Let $v = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ be a row vector that satisfies $\sum_{i=1}^n v_i = 1$. Let π be a vector chosen as in Part 1c. Prove that $\lim_{t\to+\infty} vP^t = \pi$. Then, derive that for all $i, j \in [n]$,

$$\lim_{t\to+\infty} \Pr[X_t = j | X_0 = i] = \pi_j.$$

[**HINT:** *Apply Part 1b to* $(v - \pi), (v - \pi)P, (v - \pi)P^2, \ldots$*.*]

(e) **(0 pt.)** **[Optional: this won't be graded.]** Extend the proof to the general case, where the Markov chain is irreducible and aperiodic but $P_{ij} > 0$ might not hold.

SOLUTION:

(a) The assumption $P_{ij} > 0$ implies that there is a positive probability of reaching state j from state i for arbitrary $i, j \in [n]$, so the Markov chain is irreducible. Moreover, if we choose $i = j$, we have $P_{ii} > 0$, i.e., if we start from state i, there is a positive probability of staying at state i in the next step. This implies that the Markov chain is aperiodic.

(b) Applying the hint to every column of P, for all $j \in [n]$, we have $|\sum_{i=1}^n a_i P_{ij}| \leq \sum_{i=1}^n |a_i| P_{ij} - \frac{p_{\min}}{2} \|a\|_1$. Summing over $j \in [n]$, we have

$$\begin{aligned}
\|aP\|_1 &= \sum_{j=1}^n \left| \sum_{i=1}^n a_i P_{ij} \right| \\
&\leq \sum_{j=1}^n \left(\sum_{i=1}^n |a_i| P_{ij} - \frac{p_{\min}}{2} \|a\|_1 \right) \\
&= \left(\sum_{i=1}^n |a_i| \left(\sum_{j=1}^n P_{ij} \right) \right) - \frac{n p_{\min}}{2} \|a\|_1 \\
&= \left(\sum_{i=1}^n |a_i| \right) - \frac{n p_{\min}}{2} \|a\|_1 && \text{(because } \textstyle\sum_{j=1}^n P_{ij} = 1\text{)} \\
&= (1 - n p_{\min}/2) \|a\|_1.
\end{aligned}$$

(c) Let us use $\mathbf{1}$ to denote the n dimensional all-ones column vector $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$. For every $i \in [n]$, the entries in the i-th row of P represent the probability mass function of the conditional distribution of X_{t+1} given $X_t = i$, so they add up to one: $\sum_{j=1}^n P_{ij} = 1$. This implies that $P\mathbf{1} = \mathbf{1}$, so $\lambda = 1$ is an eigenvalue of P, and thus also an eigenvalue of P^T. Therefore, there exists a non-zero row vector $\tilde{\pi}$ such that $P^T \tilde{\pi}^T = \tilde{\pi}^T$, or $\tilde{\pi} P = \tilde{\pi}$. We will show later that $\sum_{i=1}^n \tilde{\pi}_i \neq 0$, so we can define $\pi = \tilde{\pi} / \sum_{i=1}^n \tilde{\pi}_i$ which satisfies both requirements of the problem.

We prove by contradiction that $\sum_{i=1}^n \tilde{\pi}_i \neq 0$. Assuming $\sum_{i=1}^n \tilde{\pi}_i = 0$, part (b) gives us a contradiction: $\|\tilde{\pi}\|_1 = \|\tilde{\pi} P\|_1 \leq (1 - n p_{\min}/2) \|\tilde{\pi}\|_1 < \|\tilde{\pi}\|_1$. The last inequality is strict because $\tilde{\pi}$ is not the zero vector.

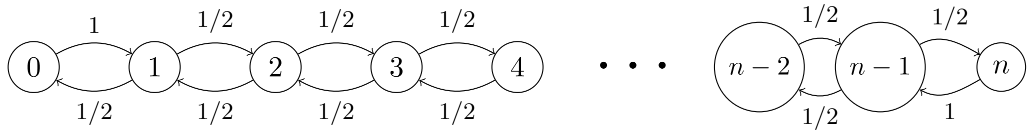
(d) The assumption $\sum_{i=1}^n v_i = 1$ implies that $v\mathbf{1} = 1$, where $\mathbf{1}$ was defined in the solution to part (c). Similarly, we have $\pi \mathbf{1} = 1$, so $(v - \pi)\mathbf{1} = 0$. We have showed $P\mathbf{1} = \mathbf{1}$, so for every $t = 0, 1, \ldots$, we have $(v - \pi) P^t \mathbf{1} = (v - \pi)\mathbf{1} = 0$. This means that all the coordinates of $(v - \pi)P^t$ add up to zero. We can thus apply part (b) to $(v - \pi)P^t$ and get

$$\|(v - \pi) P^{t+1}\|_1 = \|((v - \pi) P^t) P\|_1 \leq (1 - n p_{\min}/2) \|(v - \pi) P^t\|_1. \tag{1}$$

It is clear that $1 - n p_{\min}/2 < 1$, and we also have $1 - n p_{\min}/2 = 1 - \frac{1}{2} \sum_{j=1}^n p_{\min} \geq 1 - \frac{1}{2} \sum_{j=1}^n P_{1j} = 1/2 \geq 0$. Therefore, $1 - n p_{\min}/2 \in [0, 1)$. By induction on t using (1), we know $\|(v - \pi) P^t\|_1 \leq (1 - n p_{\min}/2)^t \|v - \pi\|_1$. Sending t to $+\infty$, the right hand side approaches 0, so we have $\lim_{t \to +\infty} \|(v - \pi) P^t\|_1 = 0$. Since $\pi P = \pi$, we know $(v - \pi) P^t = v P^t - \pi$ and thus $\lim_{t \to +\infty} \|v P^t - \pi\|_1 = 0$. Therefore, $\lim_{t \to +\infty} v P^t = \pi$.

When the initial state is $X_0 = i$, the probability mass function of the initial distribution can be represented by the i-th basis vector e_i. Since P is the transition matrix, the probability $\Pr[X_t = j | X_0 = i]$ equals to the j-th coordinate of $e_i P^t$. Choosing v to be e_i, we have $\lim_{t \to +\infty} e_i P^t = \pi$. Taking the j-th coordinate of both sides, we have $\lim_{t \to +\infty} \Pr[X_t = j | X_0 = i] = \pi_j$.

2. **(11 pt.)** Let $n > 2$, and consider the Markov chain $\{X_t\}$ defined on the states $\{0, 1, \ldots, n\}$ consisting of a random walk with reflecting barriers at 0 and n:



That is, $\{X_t\}$ is defined by the following transition probabilities:

- For $i \in \{1, \ldots, n-1\}$, we have
$$\Pr[X_t = i+1 | X_{t-1} = i] = \Pr[X_t = i-1 | X_{t-1} = i] = \frac{1}{2}.$$
- At 0 and n, we have reflecting barriers:
$$\Pr[X_t = 1 | X_{t-1} = 0] = \Pr[X_t = n-1 | X_{t-1} = n] = 1.$$

(a) **(2 pt.)** Is this chain periodic or aperiodic? Is it irreducible? Justify your answers in one sentence each.

(b) **(5 pt.)** Consider the "lazy" version of $\{X_t\}$ that, at every timestep, flips a fair coin and with probability 1/2 stays in its current state, and with probability 1/2 transitions as prescribed above. Call this lazy version $\{\tilde{X}_t\}$. Define a coupling for $\tilde{X}_t$ that ensures that the two chains in your coupling "never cross without meeting." That is, if you are coupling $\{\tilde{X}_t\}$ and $\{\tilde{Y}_t\}$, you should ensure that if $\tilde{X}_0 \leq \tilde{Y}_0$, then it will hold that $\tilde{X}_t \leq \tilde{Y}_t$ for all t.

(c) **(4 pt.)** Show that $\{\tilde{X}_t\}$ has a unique stationary distribution, and that the mixing time of $\{\tilde{X}_t\}$ is bounded by $O(n^2)$.
[**HINT:** *To bound the mixing time, use the coupling you defined in part (b).*]
[**HINT:** *Recall Lemma 6 from Class 13, which says that if Z_t is a walk on $\{0, 1, 2, \ldots\}$ with a reflecting barrier at 0 (so $\Pr[Z_t = 1 | Z_{t-1} = 0] = 1$, and otherwise $Z_t = Z_{t-1} \pm 1$ with probability 1/2 each), then the expected amount of time before $Z_t = n$, given that $Z_0 \leq n$, is at most n^2.*]

SOLUTION:

(a) The chain is periodic, since for example you can only get from 1 back to 1 by taking an even number of steps. It is irreducible because you can get from any state to any other state.

(b) There are a number of different couplings with the desired property. Here, we describe one especially simple one. First, if $\tilde{X}_t = \tilde{Y}_t$, then both do the same thing according to $\{\tilde{X}_t\}$. Otherwise, flip a fair coin. If it is heads, let $\{\tilde{X}_t\}$ be lazy (e.g., not move), and let $\{\tilde{Y}_t\}$ take a step according to $\{X_t\}$. If it is tails, do it the other way around, so $\{\tilde{Y}_t\}$ is lazy and $\{\tilde{X}_t\}$ steps according to $\{X_t\}$. This way, the two chains never move at the same time, so they can never "cross" each other until they meet, at which point they are coupled forever.

(c) First, $\{\tilde{X}_t\}$ has a unique stationary distribution because it is aperiodic (since it has a self-loop) and irreducible (since you can get anywhere from anywhere else).

Now, consider the coupling from part (b). Let $T_{s,s'}$ be the time that these two chains couple, starting from s and s' respectively. Suppose WLOG that $\tilde{X}_0 = s < s' = \tilde{Y}_0$. Then, by the "non-crossing" property in (b), the two chains will have coupled when $\tilde{X}_t = n$ for the first time. Thus, $T_{s,s'}$ is at most the time for $\tilde{X}_t$ to reach n.

From the hint (with $Z_t \leftarrow \tilde{X}_t$), we recall from class that the expected amount of time for $\tilde{X}_t$ to reach n is at most n^2. By Markov's inequality, the probability that $\tilde{X}_t$ does not reach n by time $2en^2$ is at most $1/2e$. Thus,

$$\Pr[T_{s,s'} > 2en^2] \leq 1/2e.$$

By Proposition 9 from the Class 15 Lecture notes,

$$\Delta(2en^2) \leq \Pr[T_{s,s'} > 2en^2] \leq 1/2e,$$

so by definition

$$\tau_{mix} \leq 2en^2,$$

as desired.