Random Walks: WEEK 6

1 Rate of convergence: proofs

1.1 Reminder

Let $(X_n,\, n \geq 0)$ be a Markov chain with state space S and transition matrix P, and consider the following assumptions:

- X is ergodic (irreducible, aperiodic and positive-recurrent), so there exists a stationary distribution π and it is a limiting distribution as well.
- The state space S is finite, $|S| = N$.
- Detailed balance holds ($\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in S$).

Statement 1.1. Under these assumptions, we have seen that there exist numbers $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{N-1}$ and vectors $\phi^{(0)}, \phi^{(1)}, \ldots, \phi^{(N-1)} \in \mathbb{R}^N$ such that

$$P\phi^{(k)} = \lambda_k \phi^{(k)}, \quad k = 0, \ldots, N-1$$

and $\phi_j^{(k)} = \frac{u_j^{(k)}}{\sqrt{\pi_j}}$, where $u^{(0)}, \ldots, u^{(N-1)}$ is an orthonormal basis of $\mathbb{R}^N$ ($u^{(k)}$ are the eigenvectors of the symmetric matrix Q, where $q_{ij} = \sqrt{\pi_i}\, p_{ij}\, \frac{1}{\sqrt{\pi_j}}$). Note that the $\phi^{(k)}$ do not usually form an orthonormal basis of $\mathbb{R}^N$.

Facts

1. $\phi_j^{(0)} = 1 \quad \forall j \in S, \quad \lambda_0 = 1 \quad$ and $\quad |\lambda_k| \leq 1 \quad \forall k \in \{0, \ldots, N-1\}$
2. $\lambda_1 < +1$ and $\lambda_{N-1} > -1$

Definition 1.2. Let us define $\lambda_* = \max_{k \in \{1,\ldots,N-1\}} |\lambda_k| = \max\{\lambda_1, -\lambda_{N-1}\}$. The spectral gap is defined as $\gamma = 1 - \lambda_*$.

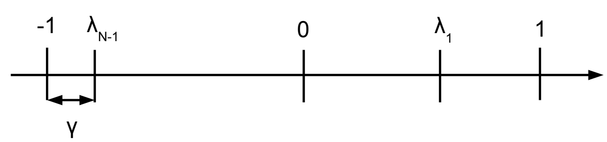


Figure 1: Spectral gap

Theorem 1.3. Under all the assumptions made above, we have

$$\|P_i^n - \pi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{j \in S} |p_{ij}(n) - \pi_j| \leq \frac{1}{2\sqrt{\pi_i}}\, \lambda_*^n \leq \frac{1}{2\sqrt{\pi_i}}\, e^{-\gamma n}, \quad \forall i \in S,\, n \geq 1$$

1.2 Proof of Fact 1

Let us first prove that $\phi_j^{(0)} = 1 \quad \forall j \in S$ and $\lambda_0 = 1$.

Consider $\phi_j^{(0)} = 1 \quad \forall j \in S$; we will prove that $(P\phi^{(0)})_i = \phi_i^{(0)}$ (so $\lambda_0 = 1$):

$$(P\phi^{(0)})_i = \sum_{j \in S} p_{ij} \underbrace{\phi_j^{(0)}}_{=1} = \sum_{j \in S} p_{ij} = 1 = \phi_i^{(0)}$$

Also, we know that $\phi_i^{(0)} = \frac{u_i^{(0)}}{\sqrt{\pi_i}}$, so $u_i^{(0)} = \sqrt{\pi_i}\,\phi_i^{(0)} = \sqrt{\pi_i}$. The norm of $u^{(0)}$ is therefore equal to 1:

$$\|u^{(0)}\|^2 = \sum_{i \in S} (u_i^{(0)})^2 = \sum_{i \in S} \pi_i = 1$$

Let us then prove that $|\lambda_k| \leq 1 \quad \forall k \in \{0, \ldots, N-1\}$.

Let $\phi^{(k)}$ be the eigenvector corresponding to λ_k. We define i to be such that $|\phi_i^{(k)}| \geq |\phi_j^{(k)}| \quad \forall j \in S$ ($|\phi_i^{(0)}| > 0$ because an eigenvector cannot be all-zero). We will use $P\phi^{(k)} = \lambda_k \phi^{(k)}$ in the following:

$$|\lambda_k \phi_i^{(k)}| = \left|(P\phi^{(k)})_i\right| = \left|\sum_{j \in S} p_{ij}\phi_j^{(k)}\right| \leq \sum_{j \in S} p_{ij} \underbrace{|\phi_j^{(k)}|}_{\leq |\phi_i^{(k)}|, \forall j \in S} \leq |\phi_i^{(k)}| \underbrace{\sum_{j \in S} p_{ij}}_{=1}$$

So we have $|\lambda_k||\phi_i^{(k)}| \leq |\phi_i^{(k)}|$, which implies that $|\lambda_k| \leq 1$, as $|\phi_i^{(k)}| > 0$. $\square$

1.3 Proof of Fact 2

We want to prove that $\lambda_1 < +1$ and $\lambda_{N-1} > -1$, which together imply that $\lambda_* < 1$.
By the assumptions made, we know that the chain is irreducible, aperiodic and finite, so $\exists n_0 > 1$ such that $p_{ij}(n) > 0$, $\forall i, j \in S, \forall n \geq n_0$.

$\boldsymbol{\lambda_1 < +1}$:

Assume ϕ is such that $P\phi = \phi$: we will prove that ϕ can only be a multiple of $\phi^{(0)}$, which implies that the eigenvalue $\lambda = 1$ has a unique eigenvector associated to it, so $\lambda_1 < 1$. Take i such that $|\phi_i| \geq |\phi_j|$, $\forall j \in S$, and let $n \geq n_0$.

$$\phi_i = (P\phi)_i = (P^n\phi)_i \overset{(*)}{=} \sum_{j \in S} p_{ij}(n)\,\phi_j$$

so

$$|\phi_i| = \left|\sum_{j \in S} p_{ij}(n)\,\phi_j\right| \leq \sum_{j \in S} p_{ij}(n)\,\underbrace{|\phi_j|}_{\leq |\phi_i|} \leq |\phi_i| \underbrace{\sum_{j \in S} p_{ij}(n)}_{1} = |\phi_i|$$

So we have $|\phi_i| \leq \sum_{j \in S} p_{ij}(n)|\phi_j| \leq |\phi_i|$. To have equality, we clearly need to have $|\phi_i| = |\phi_j|$, $\forall j \in S$ (because $p_{ij}(n) > 0$ for all i, j and $\sum_{j \in S} p_{ij}(n) = 1$ for all $i \in S$). Because $(*)$ is satisfied, we also have $\phi_i = \sum_{j \in S} p_{ij}(n)\phi_j$, which in turn implies that $\phi_j = \phi_i$ for all $j \in S$. So the vector ϕ is constant. $\square$

$\lambda_{\mathbf{N-1}} > -\mathbf{1}$:

Assume there exists $\phi \neq 0$ such that $P\phi = -\phi$: we will prove that this is impossible, showing therefore that no eigenvalue can take the value -1. Take i such that $|\phi_i| \geq |\phi_j|$, $\forall j \in S$ and let n odd be such that $n \geq n_0$.

Now, as $P^n\phi = P\phi = -\phi$, we have $-\phi_i \overset{(*)}{=} \sum_{j \in S} p_{ij}(n)\, \phi_j$ and $|\phi_i| \leq \sum_{j \in S} p_{ij}(n)\, |\phi_j| \leq |\phi_i|$. So, as above, we need to have $|\phi_j| = |\phi_i|$, for all $j \in S$ and then, thanks to $(*)$, $\phi_j = -\phi_i$, for all $j \in S$. This implies that $\phi_i = -\phi_i = 0$, and leads to $\phi_j = 0$ for all $j \in S$, which is impossible. □

1.4 Proof of the theorem

We will first use the Cauchy-Schwarz inequality which states that

$$\left|\sum_{j \in S} a_j b_j\right| \leq \left(\sum_{j \in S} a_j^2\right)^{1/2} \left(\sum_{j \in S} b_j^2\right)^{1/2}$$

so as to obtain

$$\begin{aligned} \|P_i^n - \pi\|_{\mathrm{TV}} &= \frac{1}{2} \sum_{j \in S} \underbrace{\left|\frac{p_{ij}(n) - \pi_j}{\sqrt{\pi_j}}\right|}_{a_j} \underbrace{\sqrt{\pi_j}}_{b_j} \leq \frac{1}{2} \left(\sum_{j \in S} \left(\frac{p_{ij}(n)}{\sqrt{\pi_j}} - \sqrt{\pi_j}\right)^2\right)^{1/2} \underbrace{\left(\sum_{j \in S} \pi_j\right)^{1/2}}_{1} \\ &= \frac{1}{2} \left(\sum_{j \in S} \left(\frac{p_{ij}(n)}{\sqrt{\pi_j}} - \sqrt{\pi_j}\right)^2\right)^{1/2} \end{aligned}$$

Lemma 1.4.

$$\frac{p_{ij}(n)}{\sqrt{\pi_j}} - \sqrt{\pi_j} = \sqrt{\pi_j} \sum_{k=1}^{N-1} \lambda_k^n \phi_i^{(k)} \phi_j^{(k)}$$

Proof. Remember that $u^{(0)}, \ldots, u^{(N-1)}$ is an orthonormal basis of $\mathbb{R}^N$, so we can write for any $v \in \mathbb{R}^N$ $v = \sum_{k=0}^{N-1} (v^T u^{(k)}) u^{(k)}$ i.e. $v_j = \sum_{k=0}^{N-1} (v^T u^{(k)}) u_j^{(k)}$. For a fixed $i \in S$, take $v_j = \frac{p_{ij}(n)}{\sqrt{\pi_j}}$. We obtain

$$(v^T u^{(k)}) = \sum_{j \in S} \frac{p_{ij}(n)}{\sqrt{\pi_j}} u_j^{(k)} = \sum_{j \in S} p_{ij}(n) \phi_j^{(k)} = (P^n \phi^{(k)})_i = \lambda_k^n \phi_i^{(k)}$$

which in turn implies

$$v_j = \frac{p_{ij}(n)}{\sqrt{\pi_j}} = \sum_{k=0}^{N-1} \lambda_k^n \phi_i^{(k)} u_j^{(k)} = \sum_{k=0}^{N-1} \lambda_k^n \phi_i^{(k)} \phi_j^{(k)} \sqrt{\pi_j} = \underbrace{\lambda_0^n \phi_i^{(0)} \phi_j^{(0)}}_{1} \sqrt{\pi_j} + \sqrt{\pi_j} \sum_{k=1}^{N-1} \lambda_k^n \phi_i^{(k)} \phi_j^{(k)}$$

□

Let us continue with the proof of the theorem using this lemma.

$$\begin{aligned}
\|P_i^n - \pi\|_{\mathrm{TV}} &\leq \frac{1}{2}\left(\sum_{j\in S}\left(\frac{p_{ij}(n)}{\sqrt{\pi_j}} - \sqrt{\pi_j}\right)^2\right)^{1/2} = \frac{1}{2}\left(\sum_{j\in S}\left(\sqrt{\pi_j}\sum_{k=1}^{N-1}\lambda_k^n \phi_i^{(k)}\phi_j^{(k)}\right)^2\right)^{1/2} \\
&= \frac{1}{2}\left(\sum_{j\in S}\pi_j \sum_{k,l=1}^{N-1}\lambda_k^n\phi_i^{(k)}\phi_j^{(k)}\lambda_l^n\phi_i^{(l)}\phi_j^{(l)}\right)^{1/2} = \frac{1}{2}\left(\sum_{k,l=1}^{N-1}\lambda_k^n\phi_i^{(k)}\lambda_l^n\phi_i^{(l)}\sum_{j\in S}\pi_j\phi_j^{(k)}\phi_j^{(l)}\right)^{1/2} \\
&= \frac{1}{2}\left(\sum_{k=1}^{N-1}\lambda_k^{2n}(\phi_i^{(k)})^2\right)^{1/2}
\end{aligned}$$

where we have used the fact that $\sum_{j\in S}\pi_j\phi_j^{(k)}\phi_j^{(l)} = \sum_{j\in S}u_j^{(k)}u_j^{(l)} = (u^{(k)})^T u^{(l)} = \delta_{kl}$. Remembering now that $|\lambda_k| \leq \lambda_*$ for every $1 \leq k \leq N-1$, we obtain

$$\|P_i^n - \pi\|_{\mathrm{TV}} \leq \frac{1}{2}\,\lambda_*^n\,\left(\sum_{k=1}^{N-1}(\phi_i^{(k)})^2\right)^{1/2}$$

In order to compute the term in parentheses, remember again that $v_j = \sum_{k=0}^{N-1}(v^T u^{(k)})\,u_j^{(k)}$ for every $v \in \mathbb{R}^N$, so by choosing $v = e_i$, i.e., $v_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$, we obtain:

$$v^T u^{(k)} = u_i^{(k)} \quad \text{and} \quad \delta_{ij} = \sum_{k=0}^{N-1} u_i^{(k)} u_j^{(k)}$$

For $i = j$, we get $\delta_{ii} = 1 = \sum_{k=0}^{N-1}(u_i^{(k)})^2 = \sum_{k=0}^{N-1}\pi_i\,(\phi_i^{(k)})^2$, so

$$\sum_{k=1}^{N-1}(\phi_i^{(k)})^2 = \sum_{k=0}^{N-1}(\phi_i^{(k)})^2 - \underbrace{(\phi_i^{(0)})^2}_{1} = \frac{1}{\pi_i} - 1 \leq \frac{1}{\pi_i}$$

which finally leads to the inequality

$$\|P_i^n - \pi\|_{\mathrm{TV}} \leq \frac{\lambda_*^n}{2\sqrt{\pi_i}}$$

and therefore completes the proof. □

1.5 Lazy random walks

Adding self-loops to a Markov chain makes it a priori "lazy". Surprisingly perhaps, this might in some cases speed up the convergence to equilibrium!

Adding self-loops of weight $\alpha \in (0,1)$ to every state has the following impact on the transition matrix: assuming P is the transition matrix of the initial Markov chain, the new transition matrix $\widetilde{P}$ becomes

$$\widetilde{P} = \alpha\, I + (1-\alpha)\, P$$

As a consequence:

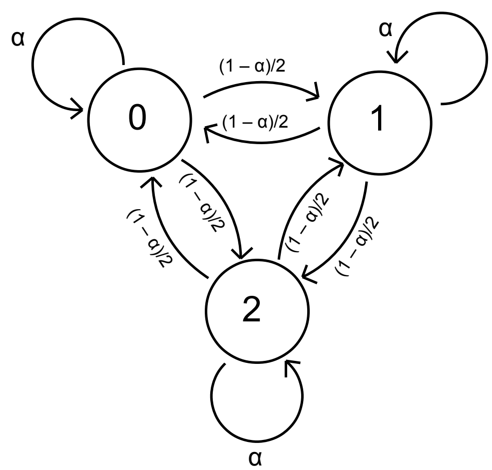
- The eigenvalues also change from λ_k to $\widetilde{\lambda}_k = \alpha + (1-\alpha)\lambda_k$, which sometimes reduces the value of $\lambda_* = \max_{1\leq k\leq N-1}|\lambda_k|$. The spectral gap being equal to $\gamma = 1 - \lambda_*$, we obtain that by reducing λ_*, we might increase the spectral gap as well as the convergence rate to equilibrium.

- Note that λ_0 stays the same: $\widetilde{\lambda}_0 = \alpha + (1-\alpha)\,\lambda_0 = 1$, as well as the stationary distribution π:

$$\pi\widetilde{P} = \pi\,(\alpha I + (1-\alpha)P) = \alpha\pi + (1-\alpha)\underbrace{\pi P}_{=\pi} = \pi$$

Example 1.5. Random walk on the circle with $N = 3$:

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \xrightarrow{\text{add } \alpha} \widetilde{P} = \begin{pmatrix} \alpha & \frac{1-\alpha}{2} & \frac{1-\alpha}{2} \\ \frac{1-\alpha}{2} & \alpha & \frac{1-\alpha}{2} \\ \frac{1-\alpha}{2} & \frac{1-\alpha}{2} & \alpha \end{pmatrix}$$



Example 1.6 (PageRank)**.** The principle behind Google's search engine algorithm is as follows. One can represent the web as a graph with the hyperlinks being the edges and the webpages being the vertices. We define the transition probabilities of a random walk on this graph as

$$p_{ij} = \frac{1}{d_i} \quad \forall j \text{ connected to } i$$

where d_i is the outgoing degree of webpage i.

The principle is that the most popular pages are the webpages visited the most often. If π is the stationary distribution of the above random walk, then π_i is a good indicator of the popularity of page i. We therefore need to solve $\pi = \pi P$. In practice however, due to the size of the state space, solving this linear system takes too long in real time. What PageRank does is to compute instead $\pi^{(0)}P^n$ for some initial distribution $\pi^{(0)}$ and a small value of n, which is meant to give a good approximation of the stationary distribution π. The quality of the approximation is of course directly linked to the rate of convergence to equilibrium. Adding self-loops of weight α to the graph seems to help in this case: the practical value chosen for α is around 15%.

NB: The above analysis unfortunately does not hold for the actual web: its graph being directed, detailed balance cannot hold! But the same phenomenon happens: adding self-loops speeds up convergence.