Bounds on Relaxation Times using Dirichlet Forms on Reversible Ergodic Markov Chains

Arpon Basu

April 27, 2023

Contents

1 Introduction

This report will mainly explore the relaxation times of finite space reversible ergodic Markov chains, both in discrete time and continuous time. More specifically, we shall explore the convergence of these Markov chains to their equilibrium states, ie:- their (unique) limiting distributions.
To achieve the aforementioned goals, we introduce the notion of *Dirichlet forms* on Markov chains, which allow us to give an extremal characterization of the spectral properties of the Markov chain, ultimately enabling us to show that the speed with which a reversible ergodic Markov chain attains its equilibrium is controlled by the second largest eigenvalue ('τ_2') of a particular symmetric matrix associated to any reversible ergodic Markov chain.
Finally, we shall conclude the report by calculating the value of τ_2 for various common Markov chains.
The references for this report are [1, Chapter 3], [1, Chapter 5]. A freely available online copy can be found here.

2 Notation and Convention

Throughout this report, we'll deal with time-homogenous Markov chains only. We shall also canonically identify the time set $\mathbb{T}$ of our Markov chain with $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ (the discrete-time case) or with $\mathbb{R}_{\geq 0}$ (the continuous-time case). Our Markov chain shall be represented as $\mathbf{X} := (X_t : t \in \mathbb{T})$, where the state space of $\mathbf{X}$ is denoted by $\mathcal{R}$, **which we will take to be finite**. We shall also represent the transition function of our Markov chain with $\mathbf{P} = (p_{ij})_{i,j \in \mathcal{R}}$, for the discrete-time case.
Further, we shall only deal with irreducible ergodic reversible Markov chains, and there we shall denote the (unique) stationary distribution on $\mathbf{X}$ by π. Note that reversibility implies

$$\pi_i p_{ij} = \pi_j p_{ji} \; \forall \, i, j \in \mathcal{R}$$

by the so-called detailed balance equations. Also, the Chapman-Kolmogorov equations then imply

$$\pi_i p_{ij}^{(n)} = \pi_j p_{ji}^{(n)} \; \forall \, i, j \in \mathcal{R}, n \in \mathbb{N}$$

where $p_{ij}^{(n)} := \mathbb{P}(X_n = j | X_0 = i)$. Note that $p_{ij}^{(n)} = (\mathbf{P}^n)_{ij}$.
We shall denote by $\mathbb{E}_\rho[\cdot]$ the expectation of a random variable constituted by $\{X_t, t \in \mathbb{T}\}$, where the initial distribution, ie:- the distribution of X_0, is taken to be ρ. Further, if our Markov chain begins at some state i [1], then the corresponding expectation is denoted as $\mathbb{E}_i[\cdot]$.
We now define the main quantity of interest in this report, namely hitting times. Let $i \in \mathcal{R}$. Then the hitting time of i is defined as

$$T_i := \inf\{t \in \mathbb{Z}_{\geq 0} : X_t = i\}$$

[1] ie:- our initial distribution ρ is δ_i

Note that T_i depends on the initial distribution of the Markov chain.

2.1 A Brief Introduction to Continuous Time Markov chains

Continuous-time Markov Chains (CTMCs) are complex entities, which we shall not need in their full generality in this report. Instead, we define only the notions of our interest.
In particular, throughout this report, we shall be repeatedly switching our perspectives from the discrete-time case to the continuous-time case, and thus the CTMCs in which we are interested are those which can be generated from Discrete-Time Markov Chains (DTMCs) by continuizing them.
Thus, with that viewpoint in mind, let $\mathbf{P}$ be the transition matrix of our DTMC. Now, when we are at any state $r \in \mathcal{R}$, we *wait* for a time given by an exponential random variable with parameter 1, and then we jump to another state according to the probability distribution dictated by (the r^{th} row of) $\mathbf{P}$. Thus, our time parameter now takes values in $\mathbb{R}_{\geq 0}$, since exponential random variables are real-valued. Moreover, note what the trajectory of our CTMC looks like: Suppose we start from some state $r_0 \in \mathcal{R}$. Then the trajectory is

$$\mathbf{X}_t = \begin{cases} r_0, t \in [0, t_1) \\ r_1, t \in [t_1, t_1 + t_2) \\ r_2, t \in [t_1 + t_2, t_1 + t_2 + t_3) \\ \dots \end{cases}$$

where $t_1, t_2, t_3, \dots$ are i.i.d Exp(1) random variables, and $(r_0, r_1, r_2, \dots)$ are a particular trajectory of the DTMC governed by $\mathbf{P}$. Indeed, the jump-and-hold description of a CTMC makes apparent the correspondence between a DTMC and a CTMC: One can take a particular trajectory of a DTMC, and separate its points with i.i.d exponential random variables to get a corresponding trajectory for the CTMC.
The description given above is known as the jump-and-hold description of a CTMC. However, for computational purposes, it is often convenient to work with another, equivalent, description of a CTMC, which is known as the infinitesimal description of a CTMC. That goes as follows: Suppose we have a CTMC which is generated in a jump-and-hold manner from a stochastic matrix $\mathbf{P}$. Then we associate a transition matrix $\mathbf{Q} := \mathbf{P} - \mathbf{I}$ to our CTMC, which has the interpretation that, for any $i \neq j, i, j \in \mathcal{R}$,

$$q_{ij} = \lim_{h \searrow 0} \frac{\mathbb{P}(X_h = j | X_0 = i)}{h}$$

or equivalently,

$$\mathbb{P}(X_h = j | X_0 = i) = q_{ij} h + o(h)$$

Also note that for $i = j$, $q_{ii} = -\sum_{j \neq i} q_{ij}$, thus making $\sum_{j \in \mathcal{R}} q_{ij} = 0$ for every $i \in \mathcal{R}$. Consequently, if π is the stationary distribution of a continuous-time Markov chain, then $\sum_{i \in \mathcal{R}} \pi_i q_{ij} = 0 \; \forall j \in \mathcal{R}$.

The reason the infinitesimal description is useful is that it can be shown that $\mathbb{P}(X_t = j|X_0 = i) =: \mathbb{P}_i(X_t = j) = q_{ij}^{(t)}$ is equal to $(\exp(\mathbf{Q}t))_{ij}$.
Yet another viewpoint of CTMCs, which is also equivalent to the jump-and-hold and infinitesimal descriptions is the observation that if $\mathbf{X}$ is a DTMC, then $Y_t \sim X_{\text{Poisson}(t)}$ is the corresponding CTMC corresponding to $\mathbf{Q} = \mathbf{P} - \mathbf{I}$. We shall employ this fruitfully in some of our derivations below.
Furthermore, let $\{\rho_j(t)\}_{j\in\mathcal{R}}$ denote the probability distribution on $\mathcal{R}$ at time t. Then

$$\Pr(X_{t+h} = j) = \sum_{i\in\mathcal{R}} \Pr(X_{t+h} = j|X_t = i) \Pr(X_t = i)$$

$$= \sum_{i\neq j} (q_{ij}h + o(h))\rho_i(t) + \Pr(X_{t+h} = j|X_t = j) \Pr(X_t = j)$$

Consequently,

$$\lim_{h\searrow 0} \frac{\Pr(X_{t+h} = j) - \Pr(X_t = j)}{h} = \sum_{i\neq j} q_{ij}\rho_i(t) + \underbrace{\Pr(X_t = j)}_{=\rho_j(t)} \lim_{h\searrow 0} \frac{\Pr(X_{t+h} = j|X_t = j) - 1}{h}$$

$$= \sum_{i\neq j} q_{ij}\rho_i(t) - \rho_j(t) \lim_{h\searrow 0} \frac{\Pr(X_{t+h} \neq j|X_t = j)}{h} = \sum_{i\neq j} q_{ij}\rho_i(t) - \rho_j(t) \underbrace{\sum_{k\neq j} q_{jk}}_{=-q_{jj}}$$

$$= \sum_{i\in\mathcal{R}} q_{ij}\rho_i(t)$$

But note that

$$\lim_{h\searrow 0} \frac{\mathbb{P}(X_{t+h} = j) - \mathbb{P}(X_t = j)}{h} = \frac{d\rho_j(t)}{dt}$$

Consequently, we arrive at the continuous version of the Chapman-Kolmogorov theorem, namely, for any $j \in \mathcal{R}$,

$$\frac{d\rho_j(t)}{dt} = \sum_{i\in\mathcal{R}} q_{ij}\rho_i(t)$$

Before concluding this section, it is very important to mention that although many properties of DTMCs have natural analogs in the continuous-time case, one property where there is a notable difference is *periodicity*: Indeed, note that for CTMCs, for any $i \in \mathcal{R}$, the set $\{t : p_{ii}^{(t)} > 0\}$ is a subset of $\mathbb{R}_{\geq 0}$ in general, and consequently we can't take the greatest common divisor of this set in the usual sense. Thus, the concept of periodicity **is not defined** in the continuous time paradigm.

3 Spectral Representation of Reversible Markov Chains

Let $|\mathcal{R}| =: n$. Note that in this case, the ergodicity of our Markov chain is ensured simply if we assume that our Markov chain is irreducible, so we shall assume irreducibility. We also assume reversibility.
We can define the quantity

$$s_{ij} := p_{ij}\sqrt{\frac{\pi_i}{\pi_j}}$$

for all $i, j \in \mathcal{R}$. Note that since our Markov chain is irreducible, π is strictly positive everywhere on $\mathcal{R}$ and thus the above equation is consistently defined. Furthermore, by the detailed balance equations, since we have $\pi_i p_{ij} = \pi_j p_{ji}$, we get that $s_{ij} = s_{ji}$, ie:- the matrix $\mathbf{S} := (s_{ij})_{i,j\in\mathcal{R}}$ is a real symmetric matrix. Consequently, we can diagonalize $\mathbf{S}$ to obtain

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathsf{T}$$

where $\mathbf{U}$ is an orthonormal matrix, and $\mathbf{\Lambda}$ is the diagonal matrix containing the eigenvalues of $\mathbf{S}$ in decreasing order, ie:- $\mathbf{\Lambda}_{ii} = \lambda_i$, and $\lambda_1 \geq \ldots \geq \lambda_n$. Now, we observe that the eigenvalues of $\mathbf{S}$ are the same as the eigenvalues of $\mathbf{P}$: Indeed, if we have $\mathbf{P}x = \lambda x$ for some $\lambda \in \mathbb{R}, x \in \mathbb{R}^n$, then we have $\mathbf{S}\tilde{x} = \lambda\tilde{x}$, where $\tilde{x}_j = \sqrt{\pi_j}\cdot x_j$, and thus for any eigenvalue λ we have the bijection $x \mapsto \mathbf{\Pi}x$, where $\mathbf{\Pi}_{ij} = \sqrt{\pi_i}\delta_{ij}$, between the eigenspaces of $\mathbf{P}, \mathbf{S}$ corresponding to the eigenvalue λ. Also, keep in mind that the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{S}$.
But note that since our Markov chain $\mathbf{X}$ is irreducible, the transition matrix $\mathbf{P}$ is a stochastic irreducible matrix. Consequently, by the Perron-Frobenius theorem, we have $1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_n \geq -1$.
Further note that

$$\mathbf{S}^t = \mathbf{U}\mathbf{\Lambda}^t\mathbf{U}^\mathsf{T} \implies s_{ij}^{(t)} = \sum_{k=1}^{n} \lambda_k^t u_{ik} u_{jk}$$

Further, the Chapman-Kolmogorov equations yield $p_{ij}^{(t)} = s_{ij}^{(t)}\sqrt{\frac{\pi_j}{\pi_i}}$. Thus, we can combine these two equations to yield the spectral representation formula, which goes as

Theorem 3.1 (Spectral Representation formula)**.** *We have*

$$\Pr(X_t = j | X_0 = i) = \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=1}^{n} \lambda_k^t u_{ik} u_{jk}$$

This result can be easily *continuized* to obtain a corresponding result for continuous-time Markov chains. Indeed, if $\mathbf{Q}$ is the transition matrix of our CTMC $\mathbf{X}$, then $\mathbf{P} = \mathbf{Q} + \mathbf{I}$ is the transition matrix for some DTMC $\mathbf{Y}$. Consequently, if the DTMC has an eigenvalue λ, then the corresponding CTMC has the eigenvalue

$1 - \lambda$.
Another way to obtain the above result is to use the Poisson method for continuizing DTMCs. Thus to extend our result in the continuous time domain, we simply set

$$\mathbb{P}(X_t = j | X_0 = i) = \mathbb{P}_i(X_t = j) = \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=1}^{n} \lambda_k^t u_{ik} u_{jk}$$

$$= \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=1}^{n} u_{ik} u_{jk} \sum_{\nu=0}^{\infty} \lambda_k^{\nu} \frac{e^{-t} t^{\nu}}{\nu!} = \sqrt{\frac{\pi_j}{\pi_i}} \sum_{k=1}^{n} e^{-(1-\lambda_k)t} u_{ik} u_{jk}$$

Thus from this point onwards, we'll freely interchange between the spectral representations in the discrete and continuous time cases, such that $\lambda^{(c)} = 1 - \lambda^{(d)}$, where c, d denote continuous and discrete respectively.
We also define the *relaxation time* τ_2 of our Markov chain at this point, which is

$$\tau_2 := \begin{cases} 1/\lambda_2, \text{for continuous-time Markov chains} \\ 1/(1-\lambda_2), \text{for discrete-time Markov chains} \end{cases}$$

As we shall see throughout this report, τ_2 will denote how fast the Markov chain converges towards its stationary distribution π.

4 Extremal Characterizations

4.1 Definition of $\mathcal{E}(g, g)$

Let $\mathbf{X}$ be an ergodic reversible Markov chain on a finite state space $\mathcal{R}$ with stationary distribution π. We define

$$\mathcal{E}(g, g) := \frac{1}{2} \sum_{i,j \in \mathcal{R}} \pi_i p_{ij} (g(j) - g(i))^2$$

for any function $g : \mathcal{R} \mapsto \mathbb{R}$, where we replace p_{ij} by q_{ij} in the continuous time analogue.

Lemma 4.1. *For the discrete-time case, we have*

$$\mathcal{E}(g, g) = \frac{1}{2} \mathbb{E}_\pi \left[(g(X_1) - g(X_0))^2 \right] = \mathbb{E}_\pi \left[g(X_0)(g(X_0) - g(X_1)) \right]$$

Proof. Note that

$$\mathbb{E}_\pi \left[(g(X_1) - g(X_0))^2 \right] = \sum_{i,j \in \mathcal{R}} (g(j) - g(i))^2 \mathbb{P}(X_1 = j, X_0 = i)$$

$$= \sum_{i,j \in \mathcal{R}} (g(j) - g(i))^2 \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(X_0 = i) = \sum_{i,j \in \mathcal{R}} (g(j) - g(i))^2 p_{ij} \pi_j$$

Furthermore,

$$\frac{1}{2}\mathbb{E}_\pi\left[(g(X_1)-g(X_0))^2\right]=\frac{1}{2}\mathbb{E}_\pi\left[(g(X_1)^2+g(X_0)^2-2g(X_0)g(X_1))\right]$$

Since we are taking expectation over a stationary distribution, we have $\mathbb{E}\left[g(X_1)^2\right]=\mathbb{E}\left[g(X_0)^2\right]$, yielding

$$\frac{1}{2}\mathbb{E}_\pi\left[(g(X_1)^2+g(X_0)^2-2g(X_0)g(X_1))\right]=\mathbb{E}_\pi\left[(g(X_0)^2-g(X_0)g(X_1))\right]$$

□

Lemma 4.2. *For the continuous-time case, we have*

$$\mathcal{E}(g,g):=-\sum_{i,j\in\mathcal{R}}\pi_i q_{ij}g(i)g(j)$$

Proof. Note that

$$\mathbb{E}_\pi\left[(g(X_h)-g(X_0))^2\right]=\sum_{i,j\in\mathcal{R}}(g(j)-g(i))^2\mathbb{P}(X_h=j,X_0=i)$$

$$=\sum_{i,j\in\mathcal{R}}(g(j)-g(i))^2\mathbb{P}(X_h=j|X_0=i)\pi_i$$

Thus

$$\lim_{h\searrow 0}\frac{1}{h}\mathbb{E}_\pi\left[(g(X_h)-g(X_0))^2\right]=\sum_{i,j\in\mathcal{R}}\pi_i(g(j)-g(i))^2\lim_{h\searrow 0}\frac{\mathbb{P}(X_h=j|X_0=i)}{h}$$

$$=\sum_{i,j\in\mathcal{R}}\pi_i(g(j)-g(i))^2q_{ij}=2\mathcal{E}(g,g)$$

On the other hand,

$$\frac{1}{2}\lim_{h\searrow 0}\frac{\mathbb{E}_\pi\left[(g(X_h)-g(X_0))^2\right]}{h}=-\lim_{h\searrow 0}\frac{\mathbb{E}_\pi\left[g(X_0)(g(X_h)-g(X_0))\right]}{h}$$

as in the discrete-time case. But the above expression equals

$$-\sum_{i,j\in\mathcal{R}}g(i)(g(j)-g(i))\mathbb{P}(X_0=i)\lim_{h\searrow 0}\frac{\mathbb{P}(X_h=j|X_0=i)}{h}$$

$$=-\sum_{i,j\in\mathcal{R}}g(i)(g(j)-g(i))\pi_i q_{ij}=-\sum_{i,j\in\mathcal{R}}g(i)g(j)\pi_i q_{ij}+\sum_{i,j\in\mathcal{R}}g(i)^2\pi_i q_{ij}$$

But

$$\sum_{i,j\in\mathcal{R}}g(i)^2\pi_i q_{ij}=\sum_{i\in\mathcal{R}}\left(g(i)^2\pi_i\sum_{j\in\mathcal{R}}q_{ij}\right)=0$$

as desired. □

Finally, we come to the reason why $\mathcal{E}$ was defined in the first place: When we want to quantify *how fast* a changing probability distribution (the canonical example of which is the continuous-time Markov chain) is tending to its stationary distribution, the $\mathcal{E}$ function arises naturally.

Definition 1. *Let μ be a probability distribution on $\mathcal{R}$. Then we define*

$$\|\mu - \pi\|_2^2 := \left(\sum_{i\in\mathcal{R}} \frac{\mu_i^2}{\pi_i}\right) - 1$$

Theorem 4.3. *Let $\rho(t)$ be the distribution over $\mathcal{R}$ at time t. Then*

$$\frac{d\|\rho(t) - \pi\|_2^2}{dt} = -2\mathcal{E}(f(t), f(t))$$

where $f_j(t) = \rho_j(t)/\pi_j$.

Proof. Note that

$$\frac{d}{dt}\|\rho(t) - \pi\|_2^2 = \sum_{j\in\mathcal{R}} \frac{1}{\pi_j}\frac{d}{dt}\rho_j(t)^2 = \sum_{j\in\mathcal{R}} \frac{2}{\pi_j}\rho_j(t)\sum_{i\in\mathcal{R}}\rho_i(t)q_{ij}$$

$$= 2\sum_{i,j\in\mathcal{R}} \frac{\rho_j(t)}{\pi_j}\frac{\rho_i(t)}{\pi_i}\pi_i q_{ij} = -2\mathcal{E}(f(t), f(t))$$

where the last equality follows from Lemma 4.2. □

4.2 Extremal Characterization of the Relaxation Time

Let λ_2 be the second largest eigenvalue of the spectral matrix $\mathbf{S}$ as defined in the earlier sections. Then we define the relaxation time τ_2 to be $1/(1 - \lambda_2)$. Informally speaking, τ_2 signifies how fast a Markov chain converges to its stationary distribution starting from an arbitrary one. Then we have a very neat characterization of τ_2 in terms of $\mathcal{E}$.

Theorem 4.4 (Extremal characterization of τ)**.**

$$\tau_2 = \sup_{\substack{g\not\equiv 0 \\ \sum_{i\in\mathcal{R}} g(i)\pi_i = 0}} \frac{\|g\|_2^2}{\mathcal{E}(g, g)}$$

where $\|g\|_2^2$ is defined to be $\sum_{i\in\mathcal{R}} \pi_i g(i)^2 = \mathbb{E}_\pi\left[g(X_0)^2\right]$.

Proof. Let $\mathbf{A}$ be any real symmetric matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \ldots$[2]. Then from the Rayleigh characterization of eigenvalues, we have

$$\mu_2 = \sup_{\substack{x\neq 0 \\ \langle x,v\rangle = 0}} \frac{x^\mathsf{T}\mathbf{A}x}{\|x\|_2^2} = \sup_{\substack{x\neq 0 \\ \langle x,v\rangle = 0}} \frac{\sum x_i a_{ij} x_j}{\sum x_i^2}$$

[2]Note that we don't make any assumptions on the signs of μ_1, μ_2, and so on. Thus the order $\mu_1 \geq \mu_2 \geq \ldots$ is the usual order on $\mathbb{R}$, not the order of their magnitudes

where v is the eigenvector of $\mathbf{A}$ corresponding to μ_1.
We apply this to our spectral matrix $\mathbf{S} := (s_{ij})_{i,j\in\mathcal{R}}$[3]. Note that the eigenvector corresponding to the highest eigenvalue μ_1 was v, where $v_i = \sqrt{\pi_i}$. Then setting $x_i = \sqrt{\pi_i} g(i)$ yields

$$\mu_2 = \sup_{\sum \sqrt{\pi_i} x_i = 0} \frac{\sum x_i s_{ij} x_j}{\sum x_i^2} = \sup_{\sum \pi_i g(i)=0} \frac{\sum x_i s_{ij} x_j}{\sum \pi_i g(i)^2} = \sup_{\sum \pi_i g(i)=0} \frac{\sum \pi_i p_{ij} g(i) g(j)}{\|g\|_2^2}$$

$$= \sup_{\sum \pi_i g(i)=0} \frac{\mathbb{E}_\pi \left[g(X_0)g(X_1)\right]}{\|g\|_2^2} = \sup_{\sum \pi_i g(i)=0} \frac{\mathbb{E}_\pi \left[g(X_0)^2\right] - \mathcal{E}(g,g)}{\|g\|_2^2}$$

where the last equality follows from Lemma 4.1.
But $\mathbb{E}_\pi \left[g(X_0)^2\right] = \|g\|_2^2$, following which the desired result easily follows. □

This has a very useful corollary, which talks about what happens when we "short" a set of vertices in $\mathcal{R}$.

Corollary 4.4.1. *Suppose we collapse a subset $A \subseteq \mathcal{R}$ into a singleton $\{a\}$. Let τ_2^A be the relaxation time of this collapsed chain. Then $\tau_2^A \leq \tau_2$.*

Proof. For any function g on $(\mathcal{R} \setminus A) \cup \{a\}$, we can extend it to a function on $\mathcal{R}$ by setting $g(\alpha) = g(a)$ for every $\alpha \in A$. This doesn't change $\|g\|, \sum \pi_i g(i)$ and $\mathcal{E}(g,g)$, and thus τ_2 is at least τ_2^A since for every candidate function in the supremum for τ_2^A, the same value is also attained in the corresponding supremum expression for τ_2. □

We can also use the above extremal characterization along with the interpretation of $\mathcal{E}$ as the rate of convergence of a Markov chain to obtain this very pleasing result.

Theorem 4.5. *Let $\rho(t)$ be the distribution of our continuous-time Markov chain (whose initial distribution is assumed to be arbitrary). Then*

$$\|\rho(t) - \pi\|_2 \leq e^{-t/\tau_2} \|\rho(0) - \pi\|_2$$

Proof. From Theorem 4.3, we have

$$\frac{d}{dt} \|\rho(t) - \pi\|_2^2 = -2\mathcal{E}(f(t), f(t))$$

where $f = \rho/\pi$ [4] is as defined in the theorem.
But

$$-2\mathcal{E}(f(t), f(t)) = -2\mathcal{E}(f(t) - 1, f(t) - 1) \leq -2 \frac{\|f(t) - 1\|_2^2}{\tau_2}$$

where the last inequality follows from Theorem 4.4. Thus

$$\frac{d}{dt} \|\rho(t) - \pi\|_2^2 \leq -2 \frac{\|f(t) - 1\|_2^2}{\tau_2} = \frac{-2}{\tau_2} \|\rho(t) - \pi\|_2^2$$

Integration then yields our desired result. □

[3]remember that $s_{ij} = p_{ij}\sqrt{\pi_i/\pi_j}$
[4]the division is taken to be componentwise

5 Relaxation Times of a Few Common Markov Chains

We calculate the relaxation times τ_2 for various common Markov chains. As established in the report, the magnitude of τ_2 gives us an indication of how fast the Markov chain attains equilibrium.

5.1 On-off chain

Consider the on-off chain, which is a Markov chain with two states, say α, β such that $p_{\alpha,\beta} = 1 - p_{\alpha,\alpha} = p \in (0,1)$, and $p_{\beta,\alpha} = 1 - p_{\beta,\beta} = q \in (0,1)$.
The transition matrix is given by $\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$. This chain is irreducible, ergodic, and aperiodic. Its stationary distribution is given by π where $\pi_\alpha = \frac{q}{p+q} = 1 - \pi_\beta$. Clearly, π satisfies the detailed balance equations, and consequently, the on-off chain is reversible. Calculating τ_2 yields $\tau_2 = \frac{1}{p+q}$.
Note that if both p, q are close to 0, then τ_2 is very large: Indeed, note that p, q are the probabilities that the walk goes from one state to the other state. Consequently, p, q being small means that our on-off chain is "sticky", and it will take more time to mix, which tallies with a larger value of τ_2.

5.2 Random Walks on Graphs

A natural source of interesting finite space ergodic reversible chains is random walks on weighted graphs. However, we face a problem when we try to analyze random walks on graphs: Not all graph random walks are aperiodic. Indeed, consider a bipartite graph where our initial distribution is (concentrated on) some particular vertex. Then a random walk on this graph, where from every vertex of our graph we move to its neighbors, is **not ergodic** since we are on different partitions of the bipartite graph on consecutive time instants.
Typically, in the Markov chain literature, two workarounds to the above dilemma are presented: The first is to analyze *lazy* random walks instead, where the walk stays at its current location with some non-zero probability ε. This destroys the periodicity of the walk since for any $i \in \mathcal{R}$, if $p_{ii}^{(n)} > 0$ for some $n \in \mathbb{N}$, then $p_{ii}^{(n+1)} \geq \varepsilon p_{ii}^{(n)} > 0$.
In fact, once we have analyzed such a lazy random walk, many of the properties of the original random walk can be recovered in the limit $\varepsilon \to 0$.
The second workaround is to move to continuous time Markov chains on the same graph. As mentioned before, the concept of periodicity doesn't make any sense in the continuous-time régime. Moreover, the spectral properties of the graph carry over to the continuous-time regime without any hassle, and consequently, we can analyze the relaxation times of our graphs without worrying about periodicity.
Since we have developed an entire framework of results for the continuous time case, this is the approach that we'll choose.

5.2.1 The Cycle Graph

Consider the cycle graph C_n on n vertices, where at any given point, we move to its left/right neighbor with equal probability. This corresponds to a stochastic matrix $\mathbf{P}$ of the form

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \dots & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & \dots & 0 \\ 0 & \frac{1}{2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & 0 & 0 & \dots & 0 \end{bmatrix}$$

ie:- $p_{ij} = \frac{1}{2}$ if and only if $i = j \pm 1 \mod n$. The chain is irreducible. Further, as discussed above, in the continuous time régime, we needn't worry about periodicity.
In this case, the stationary distribution π turns out to be identically equal to $\frac{1}{n}$ for every $i \in \{0, 1, \dots, n-1\}$, which is unsurprising given the regularity of the graph.
Moreover, this π satisfies the detailed balance equations too, so we have reversibility. Moreover, note that since π is the same for all points in its sample space, $\mathbf{S} = \mathbf{P}$.
Finally, we note that $\mathbf{P}$ is a *circulant* matrix, which has very special spectral properties. Since we don't have to completely work it out, we directly quote

$$\boxed{\tau_2 = \frac{1}{1 - \cos(\frac{2\pi}{n})} \sim \frac{n^2}{2\pi^2}}$$

At a qualitative level, it tells us that the Markov chain needs $\mathcal{O}(n^2)$ time to "mix thoroughly". One can't help but note the similarity with symmetric 1D random walks, where it takes $\mathcal{O}(n^2)$ time (in expectation) to cover a distance of $\mathcal{O}(n)$.

5.2.2 Some Other Special Graphs

We present here the values of τ_2 for random walks on some other special graphs [5].

Graphs	τ_2
n-path	$\sim (2/\pi^2)n^2$
Complete graph on n vertices	$(n-1)/n$
Star with n vertices	1
d-dimensional hypercube	$d/2$

[5] it is not difficult to verify that a random walk on a connected undirected graph is *always* irreducible and reversible

References

[1] Aldous and Fill. *Reversible Markov Chains and Random Walks on Graphs.* 2014.