AN ABSTRACT OF THE DISSERTATION OF

Kyle B. Bradford for the degree of Doctor of Philosophy in Mathematics presented on May 15, 2013.
Title: Adiabatic and Stable Adiabatic Times

Abstract approved:

Yevgeniy Kovchegov

While the stability of time-homogeneous Markov chains have been extensively studied through the concept of mixing times, the stability of time-inhomogeneous Markov chains has not been studied as in depth. In this manuscript we will introduce special types of time-inhomogeneous Markov chains that are defined through an adiabatic transition. After doing this, we define the adiabatic and the stable adiabatic times as measures of stability these special time-inhomogeneous Markov chains. To construct an adiabatic transition one needs to make a transitioning convex combination of an initial and final probability transition matrix over the time interval $[0,1]$ for two time-homogeneous, discrete time, aperiodic and irreducible Markov chains. The adiabatic and stable adiabatic times depend on how this convex combinations transitions. In the most general setting, we suggested that as long as $\mathbf{P} : [0,1] \to \mathcal{P}_n^{ia}$ is a Lipschitz continuous function with respect to the $\| \cdot \|_1$ matrix norm, then the adiabatic time is bounded above by a function of the mixing time of the final probability transition matrix

$$t_{ad}(\mathbf{P}(0), \mathbf{P}(1), \epsilon) \leq \frac{L t_{mix}^2(\mathbf{P}_1, \epsilon)}{\epsilon}.$$

For the stable adiabatic time, the most general result we achieved was for non-linear adiabatic transitions $\mathbf{P}_{\phi(\mathbf{t})} = (1-\phi(t))\mathbf{P}_0 + \phi(t)\mathbf{P}_1$ where ϕ is a Lipschitz continuous functions that is piecewise defined over a finite partition of the interval $[0,1]$ so that on each subinterval ϕ is a bi-Lipschitz continuous function. In

this setting we asymptotically bounded the stable adiabatic time by the largest mixing of $\mathbf{P}_{\phi(t)}$ over all $t \in [0,1]$. We found that

$$t_{sad}\left(\mathbf{P}_0, \mathbf{P}_1, \epsilon\right) = \mathcal{O}\left(\frac{t_{mix}^4(\epsilon)}{\epsilon^3}\right).$$

We also have some additional results at bound the stable adiabatic time in this manuscript, but they are included to show the different attempts we took and highlight how important it is to pick the right variables to compare. We also provide examples to queueing and statistical mechanics.

Adiabatic and Stable Adiabatic Times

by
Kyle B. Bradford

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 15, 2013
Commencement June 2013

Doctor of Philosophy dissertation of Kyle B. Bradford presented on May 15, 2013.

APPROVED:

Major Professor, representing Mathematics

Chair of the Department of Mathematics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Kyle B. Bradford, Author

Table of Contents

Page

Table of Contents (continued)

Table of Contents (continued)

Preface

In this text I organize the primary topics of my graduate research at Oregon State University. This research was conducted from the Fall of 2009 to the Spring of 2013. Collaboration with my academic advisor, Dr. Yevgeniy Kovchegov, began after his first journal publication [16] on the topic of the adiabatic time. Together we made another journal publication [4], on the adiabatic time, before talks with with Dr. Thinh Nguyen and his graduate research assistant Leena Zacharias. Dr. Kovchegov, Dr. Nguyen and I made a journal submission on the topic of the stable adiabatic time and the four of us published a conference proceeding on adaptive queueing policies through an adiabatic approach [27]. I have plans to make another journal submission on the stable adiabatic time by the end of the year and we have expanded upon the conference proceeding to make yet another journal submission on adaptive queueing policies. With so many bodies of work, I found it prudent to use this dissertation to fully explain our motivation for this research and to outline all of our results in one location. I also want to bring attention to some ideas that did not make it into publication.

Because both the adiabatic time and the stable adiabatic time are measurements of stability of finite-state, time-inhomogeneous Markov chains, we first organize this text with a split between discrete-time and continuous-time Markov chains. Chapters 1, 2, 3 and 4 cover discrete-time Markov chains while Chapters 5, 6, 7 and 8 cover continuous-time Markov chains.

In Chapter 1 we review some basic concepts of finite-state, discrete-time Markov chains, define the mixing time of a Markov chain and outline the conditions necessary for the mixing time to exist. We also outline the relationship between the mixing time of a Markov chain and the spectral gap of its probability transition matrix before outlining a new relationship between the mixing time of a Markov chain and a singular value measurement of its probability transition matrix.

In Chapter 2 we find an upper bound of the adiabatic time by considering a function of a related mixing time and in Chapter 3 we find an asymptotic bound of the stable adiabatic time by considering a different function of a related mixing time. We split Chapter 2 into three parts depending on the kind of adiabatic evolution used in the creation of the time-inhomogeneous Markov chains and we similarly split Chapter 3 into two parts. Each type of adiabatic evolution demanded different proof techniques to find the upper bound of the adiabatic time and the asymptotic bound of stable adiabatic time. I included the different evolutions in these two chapters to give the reader a sense of how the proofs change in each setting. In Chapter 2 we also provide examples to show that our results are optimal.

Chapter 4 contains our attempts to find either an upper bound or an asymptotic bound of the stable adiabatic time with respect to a function of a related spectral gap. Our study is motivated in part by the Quantum Adiabatic theorem which characterizes the quantum adiabatic time for the evolution of a quantum system as a result of applying of a series of Hamiltonian operators, each is a linear combination of two pre-specified initial and final Hamilton operators. These linear combinations are similar to those of initial and final probability transition matrices described in Sections 2.1 and 3.1. The quantum adiabatic time of a quantum system specifies the rate at which Hamiltonian operators change so that the ground state of the system at any time s will always remain ϵ-close to that induced by the Hamilton operator at time s. The first Quantum Adiabatic theorem was stated in the 1920s by M. Born and V.A. Fock [9], and have been subsequently studied in [15] among others. Recently, the quantum adiabatic time plays an important role in the development of quantum adiabatic computing. Specifically, quantum adiabatic algorithms are constructed as a sequence of Hamilton operators applied to a quantum system in such a way that drives the system to the desirable state or output, see for example [17]. Thus, the quantum adiabatic time is a natural choice for characterizing the running times of adiabatic quantum algorithms.

In Section 4.1 we outline the Quantum Adiabatic theorem described in [2] and we establish that the quantum adiabatic time is of the order of the inverse cube of the smallest spectral gap over the entire transition of the energy function. This result gave us hope that we could asymptotically bound the stable adiabatic time by an inverse power of the smallest spectral gap over the entire transition. We were unable to find a bound for a general finite-state, discrete-time time-inhomogeneous Markov chain under a linear adiabatic evolution by the spectral gap, so we attempted different scenarios. The first scenario is outlined in Section 4.2. In this scenario we find a bound of the stable adiabatic time when the initial and final probability transition matrices have only two states. The second scenario is outlined in Section 4.4. In this scenario we attempt to find an asymptotic bound when the initial and final probability transition matrices are reversible. This attempt ultimately failed because the non-hermitian nature of Markov chains limits our ability to accurately discuss the spectrum of the probability transition matrices. In an attempt to solve this problem for reversible Markov chains (i.e. the probability transition matrices are self-adjoint in the $l^2(\mathbb{R}^n_+, \pi_t)$ space), we used spectral techniques described in [6]; however, the use of the local norm in the $l^2(\mathbb{R}^n_+, 1/\pi_t)$ space does not allow us to extract information in such a way so that the magnitude of the adiabatic time is solely in terms of the spectral gap. Finally, the third scenario is outlined in Section 4.6. In this scenario we again attempt to find an asymptotic bound when the initial and final probability matrices are birth-death matrices. Because birth-death

Markov chains are reversible, this is an attempt to restrict even further, and use known information about birth-death Markov chains to derive a result. This result failed as well and led us to end the endeavor of comparing the stable adiabatic time to the spectral gap and turn our attention to comparing the stable adiabatic time to the mixing time. I still find the attempts in this chapter alluring and I feel that they can be useful to somebody in the future, so I included them in this work.

The final chapters discuss continuous-time Markov chains. Chapter 5 we introduce the core concepts of finite-state, continuous-time Markov chains, define the mixing time of a time-homogeneous Markov chain, and discuss the conditions necessary for the mixing time to exist. This parallels the introductory style of Chapter 1.

In Chapter 6 we find both upper bounds and asymptotic upper bounds of the adiabatic time with respect to a related mixing time. We again split the results into scenarios of different adiabatic evolution. The results from this chapter mirror the results from Chapter 2.

The continuous-time, time-inhomogeneous Markov chains described in Sections 6.1 have been used to describe queueing models [27] for networks. Specifically, in the setting described in Chapter 7, the arrival rate of a packet at the queue is assumed to be unknown and is estimated progressively. An appropriate sending rate is then determined based on this estimation. As a result, the probability transition matrix at each discrete time describes a queuing policy (or sending rate) which varies with time based on the new statistics. The adiabatic time is then used to characterize the performance of the queuing model under uncertainty due to error in estimation. The stable adiabatic time has also found practical applications in network design. The recent work of Rajagoplan et al. [19] used the adiabatic time to design optimal medium access protocols in wireless networks. In Chapter 7 we discuss some of the results of these applications.

We apply our asymptotic bound of the adiabatic time to a statistical mechanical model in Chapter 8, namely the Ising model with Glauber dynamics. We consider a general adiabatic evolution between two Hamiltonian (energy) functions on different dimensional tori. Finally, in Chapter 9, I will briefly discuss my plans for conducting future research in this area.

At this point I feel that I cannot continue without acknowledging the people that helped this research take place. I have already mentioned the members of our research group that contributed to adiabatic time publications and I naturally owe these people a great deal of thanks for their contribution and inspiration. I want to extend further thanks to both Yevgeniy Kovchegov and Thinh Nguyen for their advice and guidance while helping me publish papers and start my career as a research mathematician. I want to recognize Oregon State University and their excellent mathematics, statistics, physics and engineering departments - I could not exist without their support. During my stay, I have worked with many great minds and instead of listing all of these people, I want to name a few people (listed alphabetically) that helped shape my views in probability: Max Brugger, Bob Burton, Zlatko Dimcovic, Bechir Hamdaoui, Mina Ossiander, Bob Smythe, Enrique Thomann and Ed Waymire. I have to thank all members of my family for helping me to survive as a student and I lastly have to thank my best friend, Shannon Baker, for her love and encouragement.

Kyle B. Bradford
Corvallis, OR, USA
http://bradfordmathematics.tumblr.com/

Chapter 1

Background on Discrete Markov Chains

This chapter is not necessary for advanced readers, but it contains information that might be useful when reading the later chapters. The first section will define a discrete-time Markov chain along with many properties of these chains. The second section will define the mixing time of an irreducible and aperiodic, discrete-time Markov chain. This section contains multiple propositions that outline special bounds on the mixing time.

1.1 Discrete Markov Chains

In this section we are going to consider the creation and development of discrete-time Markov chains. Markov chains have been studied for many years and these definitions and propositions are well-known to most readers, but I want to restate them here to give us a foundation which will lead to a clear understanding of the spectral structure of the probability transition matrices of the discrete-time Markov chains.

First consider a finite dimensional, $n \times n$, matrix with real entries

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

The space of all matrices of this form $\mathcal{M}_n(\mathbb{R})$ form a noncommutative ring with identity. A subset of this ring is the space of all $n \times n$ matrices with nonnegative (positive) entries. The following definition gives an apt name for matrices in this subset.

Definition 1 *A matrix* $\mathbf{P}$ *with real entries is called* <u>*nonnegative*</u> *(positive) if all its entries are nonnegative (positive). This is denoted* $\mathbf{P} \geq 0$ *(*$\mathbf{P} > 0$*).*

Next we consider a subset of the space of all $n \times n$ nonnegative matrices that is vital in the construction of discrete-time Markov chains. The following definition gives a name for matrices in this subset.

Definition 2 *A nonnegative matrix* $\mathbf{P}$ *is called* <u>*stochastic*</u> *if* $\sum_{j=1}^{n} p_{ij} = 1$ *for all* $1 \leq i \leq n$*.*

Before I describe how these matrices help in the construction of discrete-time Markov chains, I will stop to consider some of the spectral properties of stochastic matrices. The following two propositions have been known for generations, so much so that I could readily find them in my introductory graduate linear algebra text [10]. I included their proofs in Section 1.2 since they are so short.

Proposition 1 *If* $\mathbf{P}$ *is a stochastic matrix, then* 1 *is an eigenvalue of* $\mathbf{P}$*.*

Proposition 2 *If* λ *is an eigenvalue of the stochastic matrix* $\mathbf{P}$*, then* $|\lambda| \leq 1$*.*

We will return to these two propositions in a moment. We turn our attention to the creation of a discrete-time Markov chain. We start with a finite state space E. This can be an abstract collection of objects with no ordering, but we can enumerate this collection and impose an ordering so that we can consider $E = \{1, \cdots, n\}$. Given this state space, we make our definition.

Definition 3 *A discrete-time Markov chain is a random process in the set of all sequences $X : \mathbb{Z}^+ \to E$, where each sequence has a probability associated with it uniquely determined (up to initial distribution). This probability is governed by a sequence of stochastic matrices $\mathbf{P}[k]$ which give the conditional probability at time $k \in \mathbb{Z}^+$: $\mathbb{P}(X_{k+1} = j | X_k = i) = p[k]_{ij}$ where the $p[k]_{ij}$ is the ij entry of $\mathbf{P}[k]$.*

Because the entries of the stochastic matrices $\mathbf{P}[k]$ determine the conditional probabilities at time k we call it the probability transition matrix of the Markov chain at time k. There are two basic types of discrete-time Markov chains that we will now define and we will consider both types throughout this dissertation.

Definition 4 *A discrete-time Markov chain is said to be time-homogeneous if there exists a stochastic matrix $\mathbf{P}$ such that $\mathbf{P}[k] = \mathbf{P}$ for all $k \in \mathbb{Z}^+$.*

If there exists a pair $j, k \in \mathbb{Z}^+$ such that $\mathbf{P}[j] \neq \mathbf{P}[k]$ the Markov chain is called time-inhomogeneous.

Next we highlight two important properties of the probability transition matrices of discrete-time, time-homogeneous Markov chains. We will consider matrices with these two properties many times throughout the paper, so we want to know as much about these matrices as possible. We start by defining irreducibility, but before we can define this we must first consider the concept of accessibility and communication between states in our state space E.

Definition 5 *Let $\mathbf{1}_k$ be a column vector of length n with 1 in the k^{th} entry and 0 in every other entry.*

For a time-homogeneous Markov chain, a state $j \in E$ is said to be accessible from a state $i \in E$ if there exists $k \in \mathbb{Z}$ such that $\mathbf{1}_i^T \mathbf{P}^k \mathbf{1}_j > 0$.

A state $i \in E$ is said to communicate with a state $j \in E$ if both i is accessible from j and j is accessible from i.

Being able to communicate is an equivalence relationship, $\sim$ on E, so a communicating class is an element in the quotient space $E/\sim$. The structure of this quotient space tells us whether or not our probability transition matrix is irreducible.

Definition 6 *If $E/\sim$ consists of one element, then the matrix $\mathbf{P}$ is said to be irreducible.*

We finish by defining aperiodicity, but to do this we have to understand the period of a given state in our state space.

Definition 7 *The period of a state $i \in E$ for a time-homogeneous Markov chain is defined by $k = gcd\{m : \mathbb{P}(X_m = i|X_0 = i) > 0\}$.*

If the period of every state is one, then $\mathbf{P}$ is said to be aperiodic.

We briefly return to Propositions 1 and 2 before continuing. Remember that these propositions describe some spectral properties of stochastic matrices. Now we explore some further spectral properties of irreducible and aperiodic, often called primitive, matrices. We begin with the statement of the Perron-Frobenius Theorem for irreducible and aperiodic matrices as it was written in [6]. We will not include the proof to this theorem, but there is a nice proof of the Theorem of Frobenius in [11].

Theorem 1 *For an irreducible and aperiodic* $n \times n$ *stochastic matrix* $\mathbf{P}$*, the eigenvalue* 1 *has algebraic multiplicity one and all other eigenvalues have modulus less than* 1*.*

This theorem gives us two very important conditions to place on our stochastic matrix to guarantee the existence of a unique left-handed eigenvector for the matrix $\mathbf{P}$. Furthermore, this theorem leads to an explanation of the convergence of any probability vector to this unique left-handed eigenvector under repeated applications on the right by $\mathbf{P}$. For a more complete explanation, consider the following: for any vector

$$\mathbf{u} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}$$

such that $u_i \geq 0$ for $1 \leq i \leq n$ and $\sum_{i=1}^{n} u_i = 1$ we see that $\mathbf{uP}$ is also a vector such that each entry is nonnegative and its sum is one. This tells us that multiplication of row probability vectors on the right by $\mathbf{P}$ preserves the $l^1(\mathbb{R}^n)$ structure. Naturally, for every $n \in \mathbb{Z}^+$, $\mathbf{uP}^n$ is a vector such that each entry is nonnegative and its sum is one. What happens as $n \to \infty$?

For irreducible and aperiodic matrices, [6] shows that $\mathbf{P}^n$ tends to a matrix with every row equal to the unique-left handed eigenvector of $\mathbf{P}$. You could also show this through a Jordan Decomposition of your matrix $\mathbf{P}$. This would imply that any vector $\mathbf{u}$ multiplied on the right by this matrix many times would eventually approach the unique left-handed eigenvector of $\mathbf{P}$. For this reason, the vector was given a name and we end this section with its definition.

Definition 8 *For a discrete-time, time-homogeneous Markov chain with irreducible and aperiodic probability transition matrix* $\mathbf{P}$*, the unique left-handed eigenvector associated with the eigenvalue* 1*, denoted* π*, is called the* <u>*stationary distribution*</u> *of* $\mathbf{P}$*.*

1.2 Proofs

1.2.1 Proof of Proposition 1

If we multiply the vector

$$\mathbf{v}^T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

on the left by $\mathbf{P}$, then we find that

$$\begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} p_{1j} \\ \vdots \\ \sum_{j=1}^{n} p_{nj} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

This tells us that $\mathbf{P}\mathbf{v}^T = 1\mathbf{v}^T$.

1.2.2 Proof of Proposition 2

Let $\mathbf{v}^T$ be a right-handed eigenvector of $\mathbf{P}$ associated with λ.

Let

$$\mathbf{v}^T = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Let v_m be the entry of $\mathbf{v}^T$ with the largest modulus.

$$\begin{aligned}
|v_m||\lambda - p_{mm}| &= |\lambda v_m - p_{mm} v_m| \\
&= |\sum_{j=1}^{n} p_{mj} v_j - p_{mm} v_m| \\
&= |\sum_{j \neq m} p_{mj} v_j| \\
&\leq \sum_{j \neq m} p_{mj} |v_j| \\
&\leq \sum_{j \neq m} p_{mj} |v_m| \\
&= |v_m| \sum_{j \neq m} p_{mj} \\
&= |v_m| \left(1 - p_{mm}\right).
\end{aligned}$$

We now have that $|\lambda - p_{mm}| \leq 1 - p_{mm}$.

This will imply that

$$\begin{aligned}
|\lambda| &= |\lambda - p_{mm} + p_{mm}| \\
&\leq |\lambda - p_{mm}| + p_{mm} \\
&\leq (1 - p_{mm}) + p_{mm} \\
&\leq 1.
\end{aligned}$$

1.3 Mixing Time for Discrete Markov Chains

I ended Section 1.1 with a definition of the stationary distribution of a discrete-time, time-homogeneous Markov chain with a probability transition matrix that is irreducible and aperiodic. I described in rough terms why any initial probability distribution tends to the stationary distribution through consecutive applications of the probability transition matrix. A natural question arose from this: how does the structure of the irreducible and aperiodic matrix affect how quickly the Markov chain converges to its stationary distribution? To measure this stability one needs a norm to measure the size of the changes in these probability vectors, so I am going to use the total variation norm, denoted $\| \cdot \|_{TV}$ throughout this paper. One measurement of 'how quickly' the Markov chain converges was given a name many years ago, and I recall this definition.

Definition 9 *Let* $\mathbf{P}$ *be the probability transition matrix for an irreducible, aperiodic, discrete-time, time-homogeneous Markov chain with stationary distribution* π. *Given an* $\epsilon > 0$, *the time* $t_{mix}(\mathbf{P}, \epsilon)$ *is called the* <u>*mixing time*</u> *if it is the least* $T \in \mathbb{N}$ *such that*

$$\max_{\nu} \|\nu \mathbf{P}^T - \pi\|_{TV} \leq \epsilon$$

where the maximum is taken over all probability distribution ν.

The mixing time is a topic of great interest in and of itself and it has been thoroughly researched. I want to highlight one aspect of this research that pertains to comparing the mixing time to the spectral gap of the probability transition matrix. First I will define what I mean by spectral gap.

Definition 10 *For a stochastic matrix* $\mathbf{P}$ *that is both irreducible and aperiodic, the* <u>*spectral gap*</u> *of the matrix is the difference of its largest two eigenvectors in magnitude. Specifically, if we denote* Δ *as the spectral gap of the matrix and we denote* $\lambda_2 \in (-1, 1)$ *as an eigenvalue of* $\mathbf{P}$ *such that for any other eigenvalue* λ, $|\lambda_2| \geq |\lambda|$, *then*

$$\Delta = 1 - |\lambda_2|.$$

The author of [18] provides upper and lower bounds on the mixing time with respect to the spectral gap for reversible Markov chains. Notice, however, that the upper bound is not entirely in terms of the spectral gap. Before I cite these two theorems, I will first define what it means for a Markov chain to be reversible.

Definition 11 *A Markov chain,* $\{X_m\}_{m\in\mathbb{Z}^+}$ *is* _reversible_ *if there exists a probability distribution* $\pi \in \mathbb{R}^n$ *such that for all* $m \in \mathbb{Z}^+$ *and all states* $i, j \in E$

$$\pi(i)\mathbb{P}\left(X_{m+1} = j | X_m = i\right) = \pi(j)\mathbb{P}\left(X_{m+1} = i | X_m = j\right).$$

Another way to describe reversibility is saying that the probability transition matrix is self-adjoint with respect to the the inner product $< \cdot, \cdot >_\pi$, but we will return to this definition is Section 4.4.

Theorem 2 *For a probability transition matrix,* $\mathbf{P}$*, of a time-homogeneous, discrete-time, reversible and irreducible Markov chain over a finite state space* E*, if we are given* $\epsilon > 0$ *and we let* $\pi_{min} := \min_{x\in E}\pi(x)$ *and* Δ *be the spectral gap of* $\mathbf{P}$*, then*

$$t_{mix}(\mathbf{P}, \epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{min}}\right)\frac{1}{\Delta}. \tag{1.1}$$

Theorem 3 *For a probability transition matrix,* $\mathbf{P}$*, of a time-homogeneous, discrete-time, reversible, irreducible and aperiodic Markov chain, if we are given* $\epsilon > 0$ *and let* Δ *be the spectral gap, then*

$$t_{mix}(\mathbf{P}, \epsilon) \geq \left(\frac{1}{\Delta} - 1\right)\log\left(\frac{1}{2\epsilon}\right). \tag{1.2}$$

It seems that the mixing time almost acts like the inverse of the spectral gap, but not quite. Chapter 4 addresses our attempts to asymptotically bound the stable adiabatic time with a power of the inverse spectral gap, but we unable to find a reasonable bound. Chapter 3 addresses our successful attempt to asymptotically bound the stable adiabatic time with a power of a related mixing time. The difference of these two attempts seems to be very slight if you consider the previous two theorems, but this difference turned out to be significant. To address this difference we are now going to develop a singular value relative of the spectral gap and show that this relative has a much more intrinsic relationship with the mixing time when compared to its spectral gap cousin.

The derivation of the following proposition was original work. We included it in [5] and we include the proof in Section 1.4. It is an important aspect of the reason behind exchanging the spectral gap with the mixing time when switching from the $l^2(\mathbb{R}^n)$ dynamics of the Quantum Adiabatic theorem to the $l^1(\mathbb{R}^n)$ dynamics of the Stable Adiabatic Theorem, see Chapter 4.

We know that irreducible, aperiodic time-homogeneous Markov chains governed by a probability transition matrix $\mathbf{P}$ has a unique stationary distribution, making the nullity of $(\mathbb{I}\lambda - \mathbf{P})$ equal to one when $\lambda = 1$. This would necessarily imply that the rank of $(\mathbb{I} - \mathbf{P})$ is $n-1$.

Let $\sigma_1 \geq \cdots \geq \sigma_{n-1} = \sigma$ be the positive singular values of $(\mathbb{I} - \mathbf{P})$ with respect to the Euclidean inner product, which we will denote $\|\cdot\|_2$ throughout this paper. We will similarly denote the $l^1(\mathbb{R}^n)$-norm by $\|\cdot\|_1$ throughout the paper. By definition our singular value decomposition gives us an orthonormal basis $\{\mathbf{v_1}, \cdots, \mathbf{v_n}\}$ such that $\mathbf{v_j}(\mathbb{I} - \mathbf{P})(\mathbb{I} - \mathbf{P})^T = \sigma_j^2 \mathbf{v_j}$ for $1 \leq j \leq n-1$ and $\mathbf{v_n}(\mathbb{I} - \mathbf{P})(\mathbb{I} - \mathbf{P})^T = \mathbf{0}$.

Clearly $\mathbf{v_n} = \pi / \|\pi\|_2$.

Proposition 3 *We have that for an time-homogeneous, discrete-time, n-state, irreducible and aperiodic Markov chain, if we are given $\epsilon > 0$, then if σ is the* *smallest singular value of* $\mathbb{I} - \mathbf{P}$

$$\frac{4 - n\epsilon}{4t_{mix}(\mathbf{P}, \epsilon/2)} \leq \sigma. \tag{1.3}$$

There have been many results bounding the inverse spectral gap for reversible Markov chains on weighted graphs, for example conductance bounds and weighted path upper bounds. In both [1] and [6] the authors introduce the necessary spectral structure to find these bounds. They also define a Dirichlet form to help derive the well-known Rayleigh Theorem and the Perron-Frobenius Theorem, which also describe bounds on the inverse spectral gap. Our work, however, does not employ these techniques directly, but these topics will be reviewed again in Chapter 4.

1.4 Proofs

1.4.1 Proof of Proposition 3

For $t \in \mathbb{N}$ define $\mathbf{M_{t-1}} = \mathbb{I} + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^{t-1}$.

Also define π to be the stationary distribution of $\mathbf{P}$.

Notice that $\mathbb{I} - \mathbf{P}^t = (\mathbb{I} - \mathbf{P})\mathbf{M_{t-1}}$.

For irreducible, aperiodic Markov chains we have that if $\lambda_1, \cdots \lambda_n$ are the eigenvalues of $\mathbf{P}$ such that $1 = \lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$, then

$$t, \frac{1 - \lambda_2^t}{1 - \lambda_2}, \cdots, \frac{1 - \lambda_n^t}{1 - \lambda_n}$$

are the eigenvalues of $\mathbf{M_{t-1}}$. Notice that $\mathbf{M_{t-1}}$ must be invertible because all eigenvalues are nonzero and also notice that t is the largest eigenvalue.

This implies that $\mathbb{I} - \mathbf{P} = (\mathbb{I} - \mathbf{P}^t)\mathbf{M_{t-1}}^{-1}$ and we see that

$$\sigma = \|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P})\|_2 = \|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\mathbf{M_{t-1}}^{-1}\|_2.$$

We see that if $\|\cdot\|_*$ is the standard matrix norm, then

$$\begin{aligned}
\|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\|_2 &= \|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\mathbf{M_{t-1}}^{-1}\mathbf{M_{t-1}}\|_2 \\
&\leq \|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\mathbf{M_{t-1}}^{-1}\|_2\|\mathbf{M_{t-1}}\|_* \\
&\leq t\|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\mathbf{M_{t-1}}^{-1}\|_2 \\
&\leq t\sigma.
\end{aligned}$$

If we let $\mathbf{u}$ be a vector such that for $1 \leq i \leq n$, $\mathbf{u}(i) = 0$ whenever $\mathbf{v_{n-1}}(i) \geq 0$ and $\mathbf{u}(i) = -\mathbf{v_{n-1}}(i)$ whenever $\mathbf{v_{n-1}}(i) < 0$, then we have that $\nu_{\mathbf{1}} = \mathbf{u}/\|\mathbf{u}\|_1$ and $\nu_{\mathbf{2}} = (\mathbf{v_{n-1}} + \mathbf{u})/\|\mathbf{v_{n-1}} + \mathbf{u}\|_1$ are probability distributions and

$$\begin{aligned}
\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t) = \mathbf{v_{n-1}} &+ (\|\mathbf{u}\|_1 - \|\mathbf{v_{n-1}} + \mathbf{u}\|_1)\pi \\
&- \left(\|\mathbf{v_{n-1}} + \mathbf{u}\|_1 \left(\nu_{\mathbf{2}} - \pi\right)\mathbf{P}^t - \|\mathbf{u}\|_1 \left(\nu_{\mathbf{1}} - \pi\right)\mathbf{P}^t\right).
\end{aligned}$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x}$ and $\mathbf{y}$ are probability measures, we see that

$$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_{TV} \leq \frac{\sqrt{n}}{2}\|\mathbf{x} - \mathbf{y}\|_2.$$

Through the triangle inequality we see that if we select $t = t_{mix}(\mathbf{P}, \epsilon/2)$, then

$$\| \left(\|\mathbf{v_{n-1}} + \mathbf{u}\|_1 \left(\nu_\mathbf{2} - \pi\right) \mathbf{P}^t - \|\mathbf{u}\|_1 \left(\nu_\mathbf{1} - \pi\right) \mathbf{P}^t \right) \|_2$$

$$\begin{aligned}
&\leq \|\mathbf{v_{n-1}} + \mathbf{u}\|_1 \| \left(\nu_\mathbf{2} - \pi\right) \mathbf{P}^t \|_2 \\
&\quad + \|\mathbf{u}\|_1 \| \left(\nu_\mathbf{1} - \pi\right) \mathbf{P}^t \|_2 \\
&\leq \frac{\sqrt{n}\|\mathbf{v_{n-1}} + \mathbf{u}\|_1 \| \left(\nu_\mathbf{2} - \pi\right) \mathbf{P}^t \|_{TV}}{2} \\
&\quad + \frac{\sqrt{n}\|\mathbf{u}\|_1 \| \left(\nu_\mathbf{1} - \pi\right) \mathbf{P}^t \|_{TV}}{2} \\
&\leq \frac{\sqrt{n}(\|\mathbf{v_{n-1}} + \mathbf{u}\|_1 + \|\mathbf{u}\|_1)\epsilon}{4} \\
&\leq \frac{\sqrt{n}\|\mathbf{v_{n-1}}\|_1 \epsilon}{4} \\
&\leq \frac{n\|\mathbf{v_{n-1}}\|_2 \epsilon}{4} \\
&\leq \frac{n\epsilon}{4}.
\end{aligned}$$

Because $\mathbf{v_{n-1}}$ and π are orthogonal, we see that

$$\begin{aligned}
\|\mathbf{v_{n-1}} + (\|\mathbf{u}\|_1 - \|\mathbf{v_{n-1}} - \mathbf{u}\|_1)\pi\|_2 &= \sqrt{1 + (\|\mathbf{u}\|_1 - \|\mathbf{v_{n-1}} + \mathbf{u}\|_1)^2 \left(\|\pi\|_2\right)^2} \\
&\geq 1.
\end{aligned}$$

Now through the reverse triangle inequality, meaning that for vectors $\mathbf{x}$ and $\mathbf{y}$, $\|\mathbf{x} - \mathbf{y}\|_2 \geq \|\mathbf{x}\|_2 - \|\mathbf{y}\|_2$, we see that if $t = t_{mix}(\mathbf{P}, \epsilon/2)$, then

$$\|\mathbf{v_{n-1}}(\mathbb{I} - \mathbf{P}^t)\| \geq 1 - \frac{n\epsilon}{4}.$$

This now implies that

$$\frac{4 - n\epsilon}{4t_{mix}(\mathbf{P}, \epsilon/2)} \leq \sigma.$$

Chapter 2

The Adiabatic Time Versus The Mixing Time For Discrete Markov Chains

This chapter first introduces three types of evolutions between two irreducible and aperiodic time-homogeneous Markov chains. For each type of evolution, a class of time-inhomogeneous Markov chains is created. We then turn our attention to the stability of these time-inhomogeneous Markov chains. We introduce a measurement called adiabatic time and bound this adiabatic time by a function of the mixing time of the final time-homogeneous Markov chain. We also provide an example to show that this bound is optimal.

2.1 Linear Evolution

This section introduces the notion of a linear evolution between the probability transition matrices of two discrete-time, time-homogeneous Markov chains. We then define our first type of discrete-time, time-inhomogeneous Markov chain and study a metric of stability called the adiabatic time. This nomenclature was first introduced in [16].

Definition 12 *Let $\mathbf{P_0}$ and $\mathbf{P_1}$ be the probability transition matrices for two discrete-time, time-homogeneous Markov chains. We call $\mathbf{P_0}$ the initial transition matrix and $\mathbf{P_1}$ the final transition matrix. We define a class of probability transition matrices based on a* <u>*linear evolution*</u> *between $\mathbf{P_0}$ and $\mathbf{P_1}$ to be $\{\mathbf{P_t}\}_{t\in[0,1]}$ so that*

$$\mathbf{P_t} = (1-t)\mathbf{P_0} + t\mathbf{P_1} \tag{2.1}$$

for each $t \in [0,1]$.

We define $\pi_{\mathbf{t}}$ to be the stationary distribution of $\mathbf{P_t}$ for each $t \in [0,1]$. Given $T \in \mathbb{N}$, the specific time-inhomogeneous Markov chain being considered in this section is the one such that the probability transition matrix at time k is $\mathbf{P_{\frac{k}{T}}}$ for $0 \leq k \leq T$. We consider the class of all time-inhomogeneous Markov chains of this type over all $T \in \mathbb{N}$. We will say that any Markov chain in this class is governed by an linear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$.

The adiabatic time is the smallest integer T guaranteeing that any distribution will evolve under consecutive applications of $\mathbf{P_{\frac{k}{T}}}$ for $1 \leq k \leq T$ to an epsilon-ball of the stationary distribution $\mathbf{P_1}$. We summarize this in the following definition.

Definition 13 *Given $\epsilon > 0$, a time $t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ is called the* <u>*adiabatic time*</u> *for a linear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ if it is the least $T^* \in \mathbb{N}$ such that*

$$\max_{\nu} \|\nu \mathbf{P_{\frac{1}{T}}} \mathbf{P_{\frac{2}{T}}} \cdots \mathbf{P_{\frac{T-1}{T}}} \mathbf{P_1} - \pi_1\|_{TV} \leq \epsilon \tag{2.2}$$

for all $T \geq T^*$ where the maximum is taken over all probability distributions ν.

It is crucial to note that we are using a linear evolution of discrete Markov chains for this definition, because we will later allow for evolutions that are not linear. We also remark that this definition only requires the uniqueness of the stationary distribution π_1.

Under this definition, we make a relationship between the adiabatic time for a linear adiabatic evolution between two discrete-time, time-homogeneous Markov chains and the mixing time of the final Markov chain. We attach the proof of the following corollary in Section 2.2.

Theorem 4 *Given a discrete-time, time-inhomogeneous Markov chain governed by a linear adiabatic evolution between two discrete-time, time-homogeneous, irreducible and aperiodic Markov chains with probability transition matrices* $\mathbf{P_0}$ *and* $\mathbf{P_1}$*, we have for* $\epsilon > 0$

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{2t^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}. \tag{2.3}$$

The following is a direct result of the previous theorem. This corollary was shown in [16].

Corollary 1 *Given a discrete-time, time-inhomogeneous Markov chain governed by a linear adiabatic evolution between two discrete-time, time-homogeneous, irreducible and aperiodic Markov chains with probability transition matrices* $\mathbf{P_0}$ *and* $\mathbf{P_1}$*, the asymptotic behavior of the adiabatic time as* $\epsilon \searrow 0$ *is*

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = O\left(\frac{t^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}\right). \tag{2.4}$$

We next give an example to show that the asymptotic bound from Corollary 1 is the best bound in this setting. That is to say that there exists at least one pair of Markov chains such that

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = \frac{Ct^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}$$

for some constant C. This was shown in [4] and you can reference a more detailed explanation of why this example is a lower bound in Section 2.2.

Example 1 *(The lower bound.) Let there be* $n+1$ *states,* $\{0, 1, 2, \ldots, n\}$.

$$\mathbf{P_0} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{P_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

The probability transition matrices mentioned above are not irreducible, but they have a unique stationary distribution, so that the adiabatic time can be defined in this case. Also note that we can perturb them to make them irreducible, and for these irreducible Markov chains, our adiabatic time will be a constant multiplied by the mixing time squared.

There are many practical applications for the linear adiabatic evolution between two time-homogeneous Markov chains, but there are many problems that are unaccessible by a mere linear evolution. In Sections 2.3 and 2.5 we will find a similar bound for more general types of evolution.

2.2 Proofs

2.2.1 Proof of Theorem 4

From the proof in [16], we notice that

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq K t_{mix}(\mathbf{P_1}, \epsilon/2)$$

where

$$1 + \left(\frac{\left(1 + \frac{1}{K-1}\right)^{K-1}}{e} \right)^{t_{mix}(\mathbf{P_1}, \epsilon/2)} \leq \epsilon/2.$$

After performing some basic algebra and taking the natural logarithm of either side of the equation, we see that

$$\begin{aligned}
\ln(1 - \epsilon/2) &\leq t_{mix}(\mathbf{P_1}, \epsilon/2) \left(\ln \left(\left(1 + \frac{1}{K-1}\right)^{K-1} \right) - 1 \right) \\
&\leq t_{mix}(\mathbf{P_1}, \epsilon/2) \left((K-1) \ln \left(1 + \frac{1}{K-1}\right) - 1 \right) \\
&\leq t_{mix}(\mathbf{P_1}, \epsilon/2) \left((K-1) \left(\sum_{j=1}^{\infty} (-1)^{j+1} \frac{1}{j(K-1)^j} \right) - 1 \right) \\
&\leq t_{mix}(\mathbf{P_1}, \epsilon/2) \left(\sum_{j=2}^{\infty} (-1)^{j+1} \frac{1}{j(K-1)^{j-1}} \right) \\
&\leq t_{mix}(\mathbf{P_1}, \epsilon/2) \left(\sum_{j=1}^{\infty} (-1)^{j+1} \frac{1}{j(K-1)^j} \left(\frac{-j}{j+1} \right) \right).
\end{aligned}$$

It is clear now that if we select K large enough so that

$$\begin{aligned}
\ln(1 - \epsilon/2) &\leq -t_{mix}(\mathbf{P_1}, \epsilon/2) \left(\sum_{j=1}^{\infty} (-1)^{j+1} \frac{1}{j(K-1)^j} \right) \\
&\leq -t_{mix}(\mathbf{P_1}, \epsilon/2) \ln \left(1 + \frac{1}{K-1}\right)
\end{aligned}$$

then K will be large enough to satisfy the previous inequality.

Exponentiating either side of the equation and performing the basic algebra required to solve for K we see that

$$\begin{aligned}
K &\geq 1 + \left(e^{\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)} - 1\right)^{-1} \\
&\geq 1 + \left(\sum_{j=0}^{\infty}\left(\frac{1}{j!}\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)^{j}\right) - 1\right)^{-1} \\
&\geq 1 + \left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)^{j-1}\right)^{-1} \\
&\geq 1 + \frac{t_{mix}(\mathbf{P_1},\epsilon/2)}{-\ln(1-\epsilon/2)}\left(\sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)^{j-1}\right)^{-1}.
\end{aligned}$$

Notice that the infinite sum that we have is the sum of positive terms and the first term in the sum is 1. This tells us that

$$1 \leq \sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)^{j-1}$$

therefore

$$1 \geq \left(\sum_{j=1}^{\infty}\frac{1}{j!}\left(\frac{-\ln(1-\epsilon/2)}{t_{mix}(\mathbf{P_1},\epsilon/2)}\right)^{j-1}\right)^{-1}.$$

This tells us that if we select K such that

$$K \geq 1 + \frac{t_{mix}(\mathbf{P_1},\epsilon/2)}{-\ln(1-\epsilon/2)}$$

then the above inequality will be satisfied.

Finally we can expand $\ln(1-\epsilon/2)$ to find that

$$K \geq 1 + \frac{2t_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon} \left(\sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{\epsilon}{2}\right)^{j-1} \right)^{-1}.$$

Again the infinite sum is the sum of positive terms and the first term in the sum is 1. This tells us that

$$1 \geq \left(\sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{\epsilon}{2}\right)^{j-1} \right)^{-1}.$$

We conclude that if we select K such that

$$K \geq \frac{2t_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}$$

then

$$1 + \left(\frac{\left(1 + \frac{1}{K-1}\right)^{K-1}}{e} \right)^{t_{mix}(\mathbf{P_1}, \epsilon/2)} \leq \epsilon/2.$$

Therefore, we see that

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{2t^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}.$$

2.2.2 Proof of Example 1

Recall that there are $n+1$ states, $\{0,1,2,\ldots,n\}$ and our probability matrices are

$$\mathbf{P_0} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{P_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Notice that $\pi_1 = (0, \cdots, 0, 1)$.

Also notice that $\|\nu\mathbf{P_1}^n - \pi_1\|_{TV} = 0$ for all distributions ν and

$$\|(1,0,\cdots,0)\mathbf{P_1}^{n-1} - \pi_1\|_{TV} = 1.$$

This implies that for $0 < \epsilon < 2$

$$\frac{t^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon} = \frac{n^2}{\epsilon}.$$

We also have that $\nu \mathbf{P_0} = (1, 0, \cdots, 0)$ for all distributions ν which implies that

$$\begin{aligned}
&\|\nu \mathbf{P}_{\frac{1}{T}} \cdot \mathbf{P}_{\frac{2}{T}} \cdots \mathbf{P}_{\frac{T-1}{T}} \cdot \mathbf{P_1} - \pi_1\|_{TV} \\
&\quad \geq \left\| \left(\sum_{j=0}^{T-1} \left(1 - \frac{j}{T}\right) \frac{T!}{j! \cdot T^{T-j}} \right) \left((1, 0, \cdots, 0) \mathbf{P_1}^{T-j} - \pi_1 \right) \right\|_{TV}.
\end{aligned}$$

Observe that $\nu \mathbf{P_1}^{T-j} - \pi_1 = \mathbf{0}$ for any $0 \leq j \leq T - n$. Therefore

$$\begin{aligned}
\|\nu \mathbf{P}_{\frac{1}{T}} \cdot \mathbf{P}_{\frac{2}{T}} \cdots \mathbf{P}_{\frac{T-1}{T}} \cdot \mathbf{P_1} - \pi_1\|_{TV} &\geq \sum_{j=T-n+1}^{T-1} \left(1 - \frac{j}{T}\right) \frac{T!}{j! \cdot T^{T-j}} \\
&\geq \sum_{j=T-n+1}^{T-1} \left(\frac{T!}{j! \cdot T^{T-j}} - \frac{T!}{(j-1)! \cdot T^{T-(j-1)}} \right) \\
&\geq 1 - \frac{T!}{(T-n)! \cdot T^n} \\
&\geq 1 - \frac{T-n+1}{T} \cdots \frac{T-1}{T}.
\end{aligned}$$

Now, because $\frac{T-n+1}{T} \cdots \frac{T-1}{T} \leq \left(\frac{T-\frac{n}{2}}{T}\right)^{\frac{n}{2}}$ for $n \geq 2$, we see that

$$\|\nu \mathbf{P}_{\frac{1}{T}} \cdot \mathbf{P}_{\frac{2}{T}} \cdots \mathbf{P}_{\frac{T-1}{T}} \cdot \mathbf{P_1} - \pi_1\|_{TV} \geq 1 - \left(\frac{T - \frac{n}{2}}{T}\right)^{\frac{n}{2}} \geq 1 - e^{-\left(\frac{n^2}{4T}\right)}.$$

Thus $\epsilon \geq \|\nu \mathbf{P}_{\frac{1}{T}} \cdot \mathbf{P}_{\frac{2}{T}} \cdots \mathbf{P}_{\frac{T-1}{T}} \cdot \mathbf{P_1} - \pi_1\|_{TV} \geq 1 - e^{-\left(\frac{n^2}{4T}\right)}$ implies

$$T \geq \frac{n^2}{-4 \log(1-\epsilon)} \approx \frac{1}{4} \cdot \frac{n^2}{\epsilon}.$$

This indeed tells us that

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = \mathcal{O}\left(\frac{t_{mix}^2(\mathbf{P_1}, \epsilon/2)}{\epsilon}\right).$$

2.3 Nonlinear Evolution

We now extend the results in Section 2.1 to cover more general evolutions between the probability transition matrices of two discrete-time, time-homogeneous Markov chains. In this section we will consider nonlinear evolutions rather than linear evolutions, but we find a familiar result. We will extend our results even further in the following section. We first develop our terminology for this section. Although our definitions look similar to those in the previous section, notice the different notation associated with the nonlinear evolution.

Definition 14 *Let* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *be the probability transition matrices for two irreducible, aperiodic, discrete-time, time-homogeneous Markov chains. We call* $\mathbf{P_0}$ *the initial transition matrix and* $\mathbf{P_1}$ *the final transition matrix. We define a class of probability transition matrices based on a* <u>*nonlinear evolution*</u> *between* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *to be*

$$\mathbf{P}_{\phi(\mathbf{t})} = (1 - \phi(t))\mathbf{P_0} + \phi(t)\mathbf{P_1} \tag{2.5}$$

where $\phi : [0,1] \to [0,1]$ *is continuous functions such that* $\phi(0) = 0$ *and* $\phi(1) = 1$.

We similarly define $\pi_{\phi(\mathbf{t})}$ to be the stationary distribution of $\mathbf{P}_{\phi(\mathbf{t})}$ for each $t \in [0,1]$. Given $T \in \mathbb{N}$, we now consider a time-inhomogeneous Markov chain such that the probability transition matrix at time k is $\mathbf{P}_{\phi(\frac{\mathbf{k}}{\mathbf{T}})}$ for $0 \leq k \leq T$. We consider the class of all time-inhomogeneous Markov chains of this type over all $T \in \mathbb{N}$. We will say that any Markov chain in this class is governed by a nonlinear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function ϕ.

For these types of adiabatic evolutions, the adiabatic time is now the smallest integer T guaranteeing that any distribution will evolve under consecutive applications of $\mathbf{P}_{\phi(\frac{\mathbf{k}}{\mathbf{T}})}$ for $1 \leq k \leq T$ to an ϵ-ball of the stationary distribution of $\mathbf{P_1}$. Our formal definition accounts for this nonlinear evolution.

Definition 15 *Given $\epsilon > 0$, a time $t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ is called the* <u>*adiabatic time*</u> *for a nonlinear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function ϕ if it is the least $T^* \in \mathbb{N}$ such that*

$$\max_{\nu} \|\nu \mathbf{P}_{\phi(\frac{1}{\mathbf{T}})} \mathbf{P}_{\phi(\frac{2}{\mathbf{T}})} \cdots \mathbf{P}_{\phi(\frac{\mathbf{T-1}}{\mathbf{T}})} \mathbf{P_1} - \pi_1\|_{TV} \leq \epsilon \tag{2.6}$$

for all $T \geq T^*$ where the maximum is taken over all probability distributions ν.

We want to compare this definition of the adiabatic time to the mixing time of the final Markov chain as we did in the previous section, however, the techniques used in the previous section will not help us derive this result. We will find a result on a dense subset of the continuous functions from the unit interval to itself. Our dense subset will be the space of Lipschitz continuous functions with finite Lipschitz constant. We will first remind the reader of what it means to be Lipschitz continuous.

Definition 16 *A function $\phi : [0, 1] \to [0, 1]$ is Lipschitz continuous with positive, real Lipschitz constant L if for $x, y \in [0, 1]$,*

$$\big|\phi(x) - \phi(y)\big| \leq L\big|x - y\big|. \tag{2.7}$$

Typically we understand the Lipschitz constant L to be the smallest such positive, real number under which the inequality holds. It should be clear that the space of Lipschitz continuous functions with finite Lipschitz constant mapping the unit interval to itself is dense in the space of continuous functions mapping the unit interval to itself. This is because continuous functions on the unit interval are uniformly continuous. For $\epsilon > 0$ and continuous function ϕ there exists $n \in \mathbb{N}$ (depending on the uniform continuity condition) such that the interpolation function

$$\phi_n(t) = \begin{cases} (1-nt)\phi(0) + nt\phi(\frac{1}{n}) & \text{if } 0 \leq t \leq \frac{1}{n} \\ (1-nt)\phi(\frac{1}{n}) + nt\phi(\frac{2}{n}) & \text{if } \frac{1}{n} \leq t \leq \frac{2}{n} \\ \vdots & \\ (1-nt)\phi(\frac{n-1}{n}) + nt\phi(1) & \text{if } \frac{n-1}{n} \leq t \leq 1 \end{cases}$$

has the following property:

$$\max_{t\in[0,1]} \left|\phi(t) - \phi_n(t)\right| < \epsilon.$$

Notice that ϕ_n is a Lipschitz continuous function with Lipschitz constant

$$L = \max_{1\leq k\leq n} \left|n\left(\phi\left(\frac{k}{n}\right) - \phi\left(\frac{k-1}{n}\right)\right)\right|.$$

To begin our comparison between the adiabatic time and the mixing time of the final probability transition matrix, we consider the case where ϕ is a Lipschitz continuous function with Lipschitz constant L such that $\phi(0) = 0$ and $\phi(1) = 1$. The proof of the following theorem is given in Section 2.4.

Theorem 5 *Given a time-inhomogeneous, discrete-time Markov chain governed by a nonlinear adiabatic evolution between the two irreducible and aperiodic* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *by the Lipschitz function* ϕ *with Lipschitz constant* L*, for* $\epsilon > 0$

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{4Lt^2_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}. \tag{2.8}$$

Our next goal is to expand the class of functions in the nonlinear evolution to all continuous functions on the unit interval. This has proven to be elusive, but there hope that even if we cannot do this we can expand the class to all

Hölder continous functions on the unit interval. We now turn our attention to defining a general evolution, rather than a nonlinear evolution. In Section 2.5 we will define a general evolution to account for even more adiabatic evolutions

2.4 Proofs

2.4.1 Proof of Theorem 5

Observe that

$$\nu \mathbf{P}_{\phi(\frac{1}{T})} \mathbf{P}_{\phi(\frac{2}{T})} \cdots \mathbf{P}_{\phi(\frac{T-1}{T})} \mathbf{P}_{\mathbf{1}} = \left[\prod_{j=N+1}^{T} \phi\left(j/T\right) \right] \nu_N {\mathbf{P}_{\mathbf{1}}}^{T-N} + \mathcal{E}$$

where $\nu_N = \nu \mathbf{P}_{\phi(\frac{1}{T})} \mathbf{P}_{\phi(\frac{2}{T})} \cdots \mathbf{P}_{\phi(\frac{N}{T})}$, and $\mathcal{E}$ is the rest of the terms, and both T and N are natural numbers with $N < T$.

By the triangle inequality, we have

$$\begin{aligned} \max_{\nu} \|\nu \mathbf{P}_{\phi(\frac{1}{T})} \mathbf{P}_{\phi(\frac{2}{T})} \cdots \mathbf{P}_{\phi(\frac{T-1}{T})} \mathbf{P}_{\mathbf{1}} - \pi_{\mathbf{1}}\|_{TV} & \\ \leq \max_{\nu} \|\nu {\mathbf{P}_{\mathbf{1}}}^{T-N} - \pi_{\mathbf{1}}\|_{TV} \cdot & \left[\prod_{j=N+1}^{T} \phi\left(j/T\right) \right] + S_N \end{aligned}$$

where $0 \leq S_N \leq 1 - \left[\prod_{j=N+1}^{T} \phi\left(j/T\right) \right]$.

Setting $T = Kt_{mix}(\mathbf{P}_{\mathbf{1}}, \epsilon/2)$ and $N = (K-1)t_{mix}(\mathbf{P}_{\mathbf{1}}, \epsilon/2)$, where $\epsilon > 0$ is small, we see that

$$\max_{\nu} \|\nu {\mathbf{P}_{\mathbf{1}}}^{T-N} - \pi_{\mathbf{1}}\|_{TV} \cdot \left[\prod_{j=N+1}^{T} \phi\left(j/T\right) \right] \leq \epsilon/2.$$

It now suffices to select K large enough so that $1 - \prod_{j=N+1}^{T} \phi\left(j/T\right) \leq \epsilon/2$ in order to show that $t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = Kt_{mix}(\mathbf{P_1}, \epsilon)$.

Assume that

$$K \geq \frac{4Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}$$

where L is the Lipschitz constant for the Lipschitz continuous function ϕ.

Because $1 \leq \sum_{j=1}^{\infty} (\epsilon/2)^{j-1}/j$, we see then that

$$\begin{aligned} K &\geq \frac{2Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{(\epsilon/2)\sum_{j=1}^{\infty} (\epsilon/2)^{j-1}/j} \\ &= \frac{2Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{\sum_{j=1}^{\infty} (\epsilon/2)^{j}/j} \\ &= \frac{2Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{-\log(1-\epsilon/2)}. \end{aligned}$$

After performing some algebra we find that

$$-\log(1-\epsilon/2) \geq \frac{2Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{K}.$$

First note that $2 > \pi^2/6$ where here we mean π to be the irrational number.

Remember that $\sum_{j=1}^{\infty} 1/j^2 = \pi^2/6$. We now find that

$$\begin{aligned} -\log(1-\epsilon/2) &\geq \frac{Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{K}\frac{\pi^2}{6} \\ &= \frac{Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{K}\sum_{j=1}^{\infty}\frac{1}{j^2} \\ &\geq \frac{Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{K}\sum_{j=1}^{\infty}\frac{1}{j(j+1)} \\ &= t_{mix}(\mathbf{P_1}, \epsilon/2)\left(\frac{L}{K}\right)\sum_{j=1}^{\infty}\frac{1}{j(j+1)}. \end{aligned}$$

Notice that when $\epsilon \geq 1$ the $t_{ad}\left(\mathbf{P_0}, \mathbf{P_1}, \epsilon\right) = 1$ because the total variation distance between two probability distributions is always going to be less than 1. We need only concern ourselves of the case when $\epsilon < 1$ and in this case it is easy to show that $K > L$.

This being the case we see that $1 \geq (L/K)^{j-1}$ for $j \in \mathbb{N}$.

This would now imply that

$$\begin{aligned}
-\log\left(1-\epsilon/2\right) &\geq t_{mix}(\mathbf{P_1}, \epsilon/2)\left(\frac{L}{K}\right)\sum_{j=1}^{\infty}\frac{1}{j(j+1)} \\
&\geq t_{mix}(\mathbf{P_1}, \epsilon/2)\left(\frac{L}{K}\right)\sum_{j=1}^{\infty}\frac{(L/K)^{j-1}}{j(j+1)} \\
&= t_{mix}(\mathbf{P_1}, \epsilon/2)\sum_{j=1}^{\infty}\frac{(L/K)^{j}}{j(j+1)} \\
&= t_{mix}(\mathbf{P_1}, \epsilon/2)\sum_{j=1}^{\infty}\left(\frac{L}{K}\right)^{j}\left[\frac{1}{j}-\frac{1}{j+1}\right] \\
&= t_{mix}(\mathbf{P_1}, \epsilon/2)\left[\sum_{j=1}^{\infty}\frac{(L/K)^{j}}{j}-\left(\frac{K}{L}\right)\sum_{j=1}^{\infty}\frac{(L/K)^{j+1}}{j+1}\right] \\
&= t_{mix}(\mathbf{P_1}, \epsilon/2)\left[\sum_{j=1}^{\infty}\frac{(L/K)^{j}}{j}+1-\left(\frac{K}{L}\right)\sum_{j=1}^{\infty}\frac{(L/K)^{j}}{j}\right].
\end{aligned}$$

Using the Taylor series representation of $\log\left(1-x\right)$ around $x = 0$, we see that we can write the previous inequality as

$$-\log\left(1-\epsilon/2\right) \geq t_{mix}(\mathbf{P_1}, \epsilon/2)\left[1+\left(\frac{K}{L}\right)\log\left(1-\frac{L}{K}\right)-\log\left(1-\frac{L}{K}\right)\right].$$

Performing some basic algebra, we find that

$$-\log\left(1-\epsilon/2\right) \geq -\frac{Kt_{mix}(\mathbf{P_1}, \epsilon/2)}{L}\left[\left(1-\frac{L}{K}\right)\left(1-\log\left(1-\frac{L}{K}\right)\right)-1\right].$$

We can now use the fact that $T = Kt_{mix}(\mathbf{P_1}, \epsilon/2)$ and $1/K = (1 - N/T)$ to write the previous expression as

$$-\log(1-\epsilon/2) \geq -\frac{T}{L}\left[(-1) - \left(1 - L + \frac{LN}{T}\right)\left(\log\left(1 - L + \frac{LN}{T}\right) - 1\right)\right].$$

We see that the right hand expression is an integral of a logarithm function. In fact

$$-\log(1-\epsilon/2) \geq \frac{T}{L}\int_{1-L+\frac{LN}{T}}^{1} -\log(x)\,dx.$$

Notice that $-\log(x)$ is strictly decreasing on the interval $[1 - L + \frac{LN}{T}, 1]$, so we can find a lower bound to the integral by taking a well known right hand sum from introductory integral calculus. Here we partition $[1 - L + \frac{LN}{T}, 1]$ into $T - N$ subintervals of length L/T and we have that

$$\int_{1-L+\frac{LN}{T}}^{1} -\log(x)\,dx \geq \sum_{j=1}^{T-N} -\left(\frac{L}{T}\right)\log\left(\left(1 - L + \frac{LN}{T}\right) + \frac{Lj}{T}\right).$$

We now have that

$$-\log(1-\epsilon/2) \geq -\sum_{j=1}^{T-N}\log\left(\left(1 - L + \frac{LN}{T}\right) + \frac{Lj}{T}\right).$$

With a change of indices on the summation we find that

$$\begin{aligned}
-\log(1-\epsilon/2) &\geq -\sum_{j=1}^{T-N}\log\left(\left(1 - L + \frac{LN}{T}\right) + \frac{Lj}{T}\right) \\
&= -\sum_{j=N+1}^{T}\log\left((1-L) + \frac{Lj}{T}\right) \\
&= -\sum_{j=N+1}^{T}\log\left(1 - L\left(1 - \frac{j}{T}\right)\right).
\end{aligned}$$

Multiplying both sides by negative one and then exponentiating either side, we see that

$$1 - \epsilon/2 \leq \prod_{j=N+1}^{T} \left(1 - L\left(1 - \frac{j}{T}\right)\right).$$

Now we notice that because ϕ is a Lipschitz continuous function with Lipschitz constant L

$$\begin{aligned}
\prod_{j=N+1}^{T} \left(1 - L\left(1 - \frac{j}{T}\right)\right) &= \prod_{j=N+1}^{T} \left(1 - L\big|1 - j/T\big|\right) \\
&\leq \prod_{j=N+1}^{T} \left(1 - \big|\phi\left(1\right) - \phi\left(j/T\right)\big|\right) \\
&\leq \prod_{j=N+1}^{T} \left(1 - \left(1 - \phi\left(j/T\right)\right)\right) \\
&= \prod_{j=N+1}^{T} \phi\left(j/T\right).
\end{aligned}$$

After some basic algebra, we can now state that

$$1 - \prod_{j=N+1}^{T} \phi\left(j/T\right) \leq \epsilon/2$$

when

$$K \geq \frac{4Lt_{mix}(\mathbf{P_1}, \epsilon/2)}{\epsilon}.$$

From this we now conclude that

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{4Lt_{mix}^2(\mathbf{P_1}, \epsilon/2)}{\epsilon}.$$

2.5 General Evolution

This section is devoted to extending our results from Sections 2.1 and 2.3. We begin by introducing terminology and we again find a familiar result. We will also discuss a more descriptive result at the end of this section. For the following definition we use the Matrix notation $\mathbf{P}(t) = (\mathbf{P_t}(i,j))_{i,j}$.

Definition 17 *Let $\mathbf{P_0}$ and $\mathbf{P_1}$ be two probability transition matrices where $\mathbf{P_0}$ is an initial transition matrix and $\mathbf{P_1}$ is a final transition matrix. We define a class of probability transition matrices based on a general evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ to be a class of matrices $\mathbf{P}(t)$ such that*

$$\mathbf{P}(t)(i,j) = (1-\phi_{i,j}(t))\mathbf{P_0}(i,j) + \phi_{i,j}(t)\mathbf{P_1}(i,j) \tag{2.9}$$

where $\phi_{i,j} : [0,1] \to [0,1]$ are continuous functions such that $\phi_{i,j}(0) = 0$ and $\phi_{i,j}(1) = 1$ for all $1 \le i,j \le n$ and $\sum_j \phi_{i,j}(t)\left(\mathbf{P_1}(i,j) - \mathbf{P_0}(i,j)\right) = 0$ for all $t \in [0,1]$ and each $1 \le i \le n$.

We can also describe a general evolution using matrix notation. Let $\mathcal{M}_n([0,1])$ be the collection of all $n \times n$ matrices with entries in $[0,1]$. Define

$$\mathcal{P}_n = \{\mathbf{P} \in \mathcal{M}_n([0,1]) : \mathbf{P1} = \mathbf{1}\}$$

where $\mathbf{1}$ is the n dimensional column vector with all entries 1 and define

$$\mathcal{P}_n^{ia} = \{\mathbf{P} \in \mathcal{P}_n : \mathcal{P} \text{ is irreducible and aperiodic}\}.$$

In this section we could also define the class of probability transition matrices based on a general evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ as a continuous function $\mathbf{P} : [0,1] \to \mathcal{P}_n^{ia}$ such that $\mathbf{P}(0) = \mathbf{P_0}$ and $\mathbf{P}(1) = \mathbf{P_1}$. We would also define a function $\pi : [0,1] \to \mathbb{R}^n$ such that $\pi(t)$ is the stationary distribution of $\mathbf{P}(t)$.

Given $T \in \mathbb{N}$, we now consider a time-inhomogeneous Markov chains such that the probability transition matrix at time k is $\mathbf{P}\left(\frac{k}{T}\right)$ for $0 \le k \le T$. We consider the class of all time-inhomogeneous Markov chains of this type over all $T \in \mathbb{N}$. We will say that any Markov chain in this class is governed by a general adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function $\mathbf{P}$.

For these types of adiabatic evolutions, the adiabatic time is now the smallest integer T guaranteeing that any distribution will evolve under consecutive applications of $\mathbf{P}\left(\frac{k}{T}\right)$ for $1 \le k \le T$ to an epsilon-ball of the stationary distribution of $\mathbf{P}(1)$. We summarize this in the following definition.

Definition 18 *For $\epsilon > 0$ the adiabatic time of a time-inhomogeneous, discrete-time Markov chain governed by an adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function $\mathbf{P}$, is defined as:*

$$\begin{aligned} t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = \inf\{T^* \in \mathbb{N} :& \\ &\max_{\nu} \|\nu \mathbf{P}(0)\, \mathbf{P}\left(\frac{1}{T}\right) \mathbf{P}\left(\frac{2}{T}\right) \cdots \mathbf{P}\left(\frac{T-1}{T}\right) \mathbf{P}(1) - \pi(1)\|_{TV} \\ &\le \epsilon \text{ for } T \in \mathbb{N}, T \ge T^*\}, \end{aligned} \tag{2.10}$$

where ν is a probability distribution.

We again want to compare this definition of the adiabatic time to the mixing time of the final Markov chain as we did in the two previous sections. We will follow an approach similar to the one outlined in Section 2.3. We will find a result on a dense subset of the space of continuous functions, now mapping the unit interval to $\mathcal{P}_n^{ia}$. This dense subset is the space of Lipschitz continuous functions with finite Lipschitz constant, again mapping the unit interval to $\mathcal{P}_n^{ia}$. It should be understood that this is a dense subspace if you follow a similar treatment as the one in Section 2.3. Our notion of Lipschitz continuous has now changed, however, so we define what it means to be Lipschitz continuous in this setting.

Definition 19 *A function $\mathbf{P} : [0,1] \to \mathcal{P}_n^{ia}$ is Lipschitz continuous with positive, real Lipschitz constant L, if for $x, y \in [0,1]$,*

$$\|\mathbf{P}(x) - \mathbf{P}(y)\| \leq L\big|x - y\big| \tag{2.11}$$

where $\|\cdot\|$ is some Matrix norm.

The matrix norm we will use in this section for a matrix $\mathbf{M}$ is $\|\mathbf{M}\|_1 = \max_\nu \|\nu M\|_1$ where the maximum is taken over all probability distributions ν and the vector norm $\|\cdot\|_1$ is the standard l^1-norm.

To begin our comparison between the adiabatic time and the mixing time of the final probability transition matrix, we now consider $\mathbf{P}$ to be a Lipschitz continuous function with Lipschitz constant L such that $\mathbf{P}(0) = \mathbf{P_0}$ and $\mathbf{P}(1) = \mathbf{P_1}$. The proof of the following Theorem is given in Section 2.6.

Theorem 6 *Given a time-inhomogeneous, discrete-time Markov chain governed by a general adiabatic evolution between the two irreducible and aperiodic $\mathbf{P_0}$ and $\mathbf{P_1}$ by the Lipschitz function $\mathbf{P}$ with Lipschitz constant L, for $\epsilon > 0$*

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{L t_{mix}^2(\mathbf{P_1}, \epsilon/2)}{\epsilon}. \tag{2.12}$$

We now wish to make a slight improvement which takes into account exactly how these general adiabatic evolutions are continuous. To do this we return to the notation that we developed at the beginning of the section from Definition 17. The following Theorem describes our more precise result and the proof is given in Section 2.6.

Theorem 7 *Suppose* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *are irreducible and aperiodic Markov chains. Consider a general adiabatic evolution between* $\mathbf{P_0}$ *and* $\mathbf{P_1}$*. Let* $\epsilon > 0$*. Letting* ϕ *be the piecewise minimum function of all of the* $\phi_{i,j}$ *functions, if* m *is the first positive integer such that* $\phi^{(m)}(1) \neq 0$ *then*

$$t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = \mathcal{O}\left(\frac{t_{mix}^{\frac{m+1}{m}}(\mathbf{P_1}, \epsilon/2)}{\epsilon^{\frac{1}{m}}}\right) \tag{2.13}$$

This is in fact the best bound in this new setting as shown through the example in Section 2.1; however, the proof is somewhat different. We include the following example and the proof of why this example shows our bound in Theorem 7 is optimal. This was shown in [4].

Example 2 *(The lower bound.) Let there be* $n+1$ *states,* $\{0, 1, 2, \ldots, n\}$.

$$\mathbf{P_0} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$\mathbf{P_1} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

2.6 Proofs

2.6.1 Proof of Theorem 6

Observe that

$$\begin{aligned}\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1)&\\ &= \nu_N\left(\mathbf{P}(1)+\left(\mathbf{P}\left(\frac{N+1}{T}\right)-\mathbf{P}(\mathbf{1})\right)\right)\mathbf{P}^*_{N+2}\\ &= \nu_N\mathbf{P}(1)\mathbf{P}^*_{N+2}+\nu_N\left(\mathbf{P}\left(\frac{N+1}{T}\right)-\mathbf{P}(\mathbf{1})\right)\mathbf{P}^*_{N+2}.\end{aligned}$$

where $\nu_N = \nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{N}{T}\right)$, $\mathbf{P}^*_k = \mathbf{P}\left(\frac{k}{T}\right)\cdots\mathbf{P}(1)$, and both T and N are natural numbers with $N < T$.

By continuing this process for $\mathbf{P}\left(\frac{N+2}{T}\right)$ and so on, we find that

$$\begin{aligned}\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1)&\\ &= \nu_N\left(\mathbf{P}(1)\right)^{T-N}\\ &+\sum_{k=0}^{T-N-2}\nu_N\left(\mathbf{P}(1)\right)^k\left(\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}(\mathbf{1})\right)\mathbf{P}^*_{N+2+k}.\end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned}\|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1)-\pi(1)\|_{TV}&\\ &\leq \|\nu_N\left(\mathbf{P}(1)\right)^{T-N}-\pi(1)\|_{TV}\\ &+\sum_{k=0}^{T-N-2}\|\nu_N\left(\mathbf{P}(1)\right)^k\left(\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}(1)\right)\mathbf{P}^*_{N+2+k}\|_{TV}.\end{aligned}$$

Because $2\|\mu-\nu\|_{TV} = \|\mu-\nu\|_1$ whenever μ and ν are probability distributions, we see that

$$\|\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1)-\pi(1)\|_{TV}\leq\|\nu_N(\mathbf{P}(1))^{T-N}-\pi(1)\|_{TV}$$
$$+\frac{1}{2}\sum_{k=0}^{T-N-2}\|\nu_k^*\mathbf{P}_{N+2+k}^*\|_1$$

where $\nu_k^*=\nu_N(\mathbf{P}(1))^k\left(\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}(1)\right).$

Notice that for $0\leq k\leq T-N-2$, because $\mathbf{P}_{N+2+k}^*$ is a probability distribution, we have that

$$\begin{aligned}
\|\nu_k^*\mathbf{P}_{N+2+k}^*\|_1 &= \sum_{j=1}^{n}\Big|\sum_{i=1}^{n}\nu_k^*(i)\,\mathbf{P}_{N+2+k}^*(i,j)\Big| \\
&\leq \sum_{j=1}^{n}\sum_{i=1}^{n}\left|\nu_k^*(i)\right|\mathbf{P}_{N+2+k}^*(i,j) \\
&= \sum_{i=1}^{n}\left|\nu_k^*(i)\right|\sum_{j=1}^{n}\mathbf{P}_{N+2+k}^*(i,j) \\
&= \sum_{i=1}^{n}\left|\nu_k^*(i)\right| \\
&= \|\nu_k^*\|_1.
\end{aligned}$$

We therefore see that

$$\begin{aligned}
&\|\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1)-\pi(1)\|_{TV} \\
&\qquad\leq\|\nu_N(\mathbf{P}(1))^{T-N}-\pi(1)\|_{TV} \\
&\qquad+\frac{1}{2}\sum_{k=0}^{T-N-2}\|\nu_N(\mathbf{P}(1))^k\left(\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}(1)\right)\|_1.
\end{aligned}$$

It is clear to see that $\nu_N(\mathbf{P}(1))^k$ is a probability vector for $0\leq k\leq T-N-2$, so we can see that

$$\begin{aligned}\max_{\nu}\|\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right)-\pi\left(1\right)\|_{TV}\\ &\leq\max_{\nu}\|\nu\left(\mathbf{P}\left(1\right)\right)^{T-N}-\pi\left(1\right)\|_{TV}\\ &+\frac{1}{2}\sum_{k=0}^{T-N-2}\max_{\nu}\|\nu\left(\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}\left(1\right)\right)\|_{1}\end{aligned}$$

where the maximum is taken over all probability vectors ν.

We observe that the terms in the sum of the right hand side of the inequality are now the matrix norm for the matrix $\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}\left(1\right)$ for $0\leq k\leq T-N-2$. This would imply that

$$\begin{aligned}\max_{\nu}\|\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right)-\pi\left(1\right)\|_{TV}\\ &\leq\max_{\nu}\|\nu\left(\mathbf{P}\left(1\right)\right)^{T-N}-\pi\left(1\right)\|_{TV}\\ &+\frac{1}{2}\sum_{k=0}^{T-N-2}\|\mathbf{P}\left(\frac{N+1+k}{T}\right)-\mathbf{P}\left(1\right)\|_{1}.\end{aligned}$$

At this point we consider $\epsilon>0$ and we use the fact that the space of Lipschitz continuous functions with finite Lipschitz constant from $[0,1]$ to $\mathcal{P}_n^{ia}$ is dense in the space of continuous functions from $[0,1]$ to $\mathcal{P}_n^{ia}$ to find a Lipschitz continuous function $\mathbf{P}^*:[0,1]\to\mathcal{P}_n^{ia}$ with Lipschitz constant L such that $\mathbf{P}^*\left(0\right)=\mathbf{P}_0$, $\mathbf{P}^*\left(1\right)=\mathbf{P}_1$ and

$$\|\mathbf{P}(t)-\mathbf{P}^*(t)\|_1\leq\frac{\epsilon}{2t_{mix}\left(\mathbf{P}\left(1\right),\epsilon/2\right)}$$

for all $t\in[0,1]$.

We can use our previous inequality to write

$$\begin{aligned}
\max_{\nu} \|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1) - \pi(1)\|_{TV} \\
&\leq \max_{\nu} \|\nu \left(\mathbf{P}(1)\right)^{T-N} - \pi(1)\|_{TV} \\
&+ \frac{1}{2}\sum_{k=0}^{T-N-2} \|\mathbf{P}^{*}\left(\frac{N+1+k}{T}\right) - \mathbf{P}^{*}(1)\|_{1} \\
&+ \frac{1}{2}\sum_{k=0}^{T-N-2} \|\mathbf{P}\left(\frac{N+1+k}{T}\right) - \mathbf{P}^{*}\left(\frac{N+1+k}{T}\right)\|_{1}
\end{aligned}$$

Because $\mathbf{P}^{*} : [0,1] \to \mathcal{P}_{n}^{ia}$ is a Lipschitz continuous function with Lipschitz constant L, we see then that

$$\begin{aligned}
\max_{\nu} \|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1) - \pi(1)\|_{TV} \\
&\leq \max_{\nu} \|\nu \left(\mathbf{P}(1)\right)^{T-N} - \pi(1)\|_{TV} \\
&+ \frac{L}{2}\sum_{k=0}^{T-N-2} \left|\frac{N+1+k}{T} - 1\right| \\
&+ \frac{1}{2}\sum_{k=0}^{T-N-2} \frac{\epsilon}{2t_{mix}\left(\mathbf{P}(1), \epsilon/2\right)}.
\end{aligned}$$

After relabeling our sum, we can easily see that

$$\begin{aligned}
\max_{\nu} \|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}(1) - \pi(1)\|_{TV} \\
&\leq \max_{\nu} \|\nu \left(\mathbf{P}(1)\right)^{T-N} - \pi(1)\|_{TV} \\
&+ \frac{L}{4T}(T-N-1)(T-N) \\
&+ \frac{\epsilon}{4t_{mix}\left(\mathbf{P}(1), \epsilon/2\right)}(T-N-1).
\end{aligned}$$

Setting $T = Kt_{mix}(\mathbf{P}(1), \epsilon/2)$ and $N = (K-1)t_{mix}(\mathbf{P}(1), \epsilon/2)$, where $\epsilon > 0$

is small, we see that

$$\begin{aligned}
\max_{\nu} \|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots \mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right) - \pi\left(1\right)\|_{TV} \\
\leq \frac{3\epsilon}{4} + \frac{L}{4K}\left(t_{mix}\left(\mathbf{P}\left(1\right), \epsilon/2\right) - 1\right) \\
\leq \frac{3\epsilon}{4} + \frac{L}{4K} t_{mix}\left(\mathbf{P}\left(1\right), \epsilon/2\right).
\end{aligned}$$

Selecting

$$K = \frac{L t_{mix}\left(\mathbf{P_1}, \epsilon/2\right)}{\epsilon}$$

we find that

$$\max_{\nu} \|\nu \mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots \mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right) - \pi\left(1\right)\|_{TV} \leq \epsilon.$$

This implies that

$$t_{ad}\left(\mathbf{P_0}, \mathbf{P_1}, \epsilon\right) \leq \frac{L t_{mix}^2\left(\mathbf{P_1}, \epsilon/2\right)}{\epsilon}.$$

2.6.2 Proof of Theorem 7

Recall that for each $s \in [0,1]$ we have $\phi(s) = \min_{i,j}\{\phi_{i,j}(s)\}$.

We see that

$$\mathbf{P}(s)(i,j) = (1-\phi_{i,j}(s))\mathbf{P_0}(i,j) + (\phi_{i,j}(s) - \phi(s))\mathbf{P_1}(i,j) + \phi(s)\mathbf{P_1}(i,j)$$

will give us a well defined transition matrix $\hat{\mathbf{P}}$ such that

$$(1-\phi(s))\hat{\mathbf{P}}(i,j) = (1-\phi_{i,j}(s))\mathbf{P_0}(i,j) + (\phi_{i,j}(s) - \phi(s))\mathbf{P_1}(i,j).$$

We will thus have that

$$\mathbf{P}(s) = (1-\phi(s))\hat{\mathbf{P}} + \phi(s)\mathbf{P_1}.$$

Observe that

$$\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right) = \left[\prod_{j=N+1}^{T}\phi(\frac{j}{T})\right]\nu_N{\mathbf{P_1}}^{T-N} + \mathcal{E}$$

where $\nu_N = \nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{N-1}{T}\right)\mathbf{P}\left(\frac{N}{T}\right)$, and $\mathcal{E}$ is the rest of the terms, and both T and N are natural numbers with $N < T$.

By the triangle inequality, we have

$$\begin{aligned}\max_{\nu}\|\nu\mathbf{P}\left(\frac{1}{T}\right)\mathbf{P}\left(\frac{2}{T}\right)\cdots\mathbf{P}\left(\frac{T-1}{T}\right)\mathbf{P}\left(1\right) - \pi\left(1\right)\|_{TV} \\ \leq \max_{\nu}\|\nu{\mathbf{P_1}}^{T-N} - \pi_1\|_{TV}\left[\prod_{j=N+1}^{T}\phi(\frac{j}{T})\right] + S_N\end{aligned}$$

where $0 \leq S_N \leq 1 - \left[\prod_{j=N+1}^{T}\phi(\frac{j}{T})\right]$.

Supposing we set $T - N = t_{mix}(\mathbf{P_1}, \epsilon/2)$ where $\epsilon > 0$ is small, we have that

$$\max_{\nu}\|\nu{\mathbf{P_1}}^{T-N} - \pi_1\|_{TV}\left[\prod_{j=N+1}^{T}\phi(\frac{j}{T})\right] \leq \epsilon/2.$$

Setting $1 - \left[\prod_{j=N+1}^{T}\phi(\frac{j}{T})\right] \leq \epsilon/2$ we obtain

$$\log\left(1-\epsilon/2\right) \leq \sum_{j=N+1}^{T}\log\phi(\frac{j}{T}).$$

We plug in the approximation of the minimum function ϕ around $x = 1$

$$\phi(x) = 1 + \frac{\phi^{(m)}(1)(x-1)^m}{m!} + \mathcal{O}\left(|x-1|^{m+1}\right)$$

obtaining

$$\begin{aligned} &-\log(1-\epsilon/2) \\ &\quad \geq -\sum_{j=N+1}^{T} \log\left(1 + \frac{(-1)^m \phi^{(m)}(1)(T-j)^m}{T^m \cdot m!} + \mathcal{O}\left((1-j/T)^{m+1}\right)\right). \end{aligned}$$

Therefore

$$-\log\left(1-\epsilon/2\right) \geq \frac{(-1)^{m+1}\phi^{(m)}(1)}{T^m \cdot m!} \sum_{j=1}^{T-N-1} j^m + \mathcal{O}\left(\frac{(T-N)^{m+2}}{T^{m+1}}\right).$$

Observe that $(-1)^{m+1}\phi^{(m)}(1) \geq 0$ as $\phi : [0,1] \to [0,1]$ and $\phi(1) = 1$.

Because

$$\sum_{j=1}^{n-1} j^k = \sum_{j=0}^{k} \frac{B_j}{(k+1)-j} \left(\begin{array}{c} k \\ j \end{array}\right) n^{(k+1)-j}$$

we see that

$$\begin{aligned} \sum_{j=1}^{t_{mix}(\mathbf{P_1},\epsilon/2)-1} j^m = \sum_{k=0}^{m} \frac{B_k}{(m+1)-k} \left(\begin{array}{c} m \\ k \end{array}\right) \\ \cdot\, t_{mix}\left(\mathbf{P_1}, \epsilon/\mathbf{2}\right)^{(m+1)-k} + \mathcal{O}\left(\frac{(T-N)^{m+2}}{T^{m+1}}\right) \end{aligned}$$

where B_k is the k^{th} Bernoulli number, and therefore

$$\begin{aligned}
\epsilon &> -\log(1-\epsilon/2) \\
&\geq \frac{(-1)^{m+1}\phi^{(m)}(1)}{T^m \cdot m!} \sum_{k=0}^{m} \frac{B_k}{(m+1)-k} \begin{pmatrix} m \\ k \end{pmatrix} \\
&\qquad\qquad \cdot t_{mix}\left(\mathbf{P_1}, \epsilon/2\right)^{(m+1)-k} + \mathcal{O}\left(\frac{(T-N)^{m+2}}{T^{m+1}}\right)
\end{aligned}$$

In order for the right hand side of the above equation to be $-\log(1-\epsilon/2)$ close to zero, it is sufficient for T to be of order

$$\mathcal{O}\left(\frac{t_{mix}^{\frac{m+1}{m}}\left(\mathbf{P_1}, \epsilon/2\right)}{\epsilon^{\frac{1}{m}}}\right).$$

2.6.3 Proof of Example 2

Recall that in this general adiabatic setting

$$\mathbf{P}(t)(i,j) = \left(1-\phi_{i,j}(t)\right)\mathbf{P_0}(i,j) + \phi_{i,j}(t)\mathbf{P_1}(i,j).$$

Suppose $\phi_{i,j}(t) = \phi(t)$ for all $1 \leq i, j \leq n$ and suppose $m \geq 1$ is the smallest integer such that $\phi^{(m)}(1) \neq 0$. Then noticing that for any distribution ν, $\nu\mathbf{P_0} = \mathbf{e}_1 = (1, 0, \cdots, 0)$, we have that

$$\begin{aligned}
&\|\nu\mathbf{P}\left(\frac{1}{T}\right)\cdots\mathbf{P}\left(1\right) - \pi(1)\|_{TV} \\
&\qquad \geq \left\|\mathbf{e}_1\left(\sum_{l=0}^{T-1}(1-\phi(l/T))\prod_{j=l+1}^{T}\phi(j/T)\right)\mathbf{P}^{T-l}(1) - \pi(1)\right\|_{TV}
\end{aligned}$$

and therefore

$$\begin{aligned}\|\nu\mathbf{P}\left(\frac{1}{T}\right)\cdots\mathbf{P}(1)-\pi(1)\|_{TV} &\geq \sum_{l=T-n+1}^{T-1}(1-\phi(l/T))\prod_{j=l+1}^{T}\phi(j/T)\\ &= \sum_{l=T-n+1}^{T-1}\left(\prod_{j=l+1}^{T}\phi(j/T)-\prod_{j=l}^{T}\phi(j/T)\right)\\ &= 1-\prod_{j=T-n+1}^{T}\phi(j/T).\end{aligned}$$

The minimum function $\phi(x)=1+\frac{\phi^{(m)}(1)(x-1)^m}{m!}+\mathcal{O}\left(|x-1|^{m+1}\right)$ and

$$\begin{aligned}&\|\nu\mathbf{P}\left(\frac{1}{T}\right)\cdots\mathbf{P}(1)-\pi(1)\|_{TV}\\ &\quad\geq 1-\prod_{j=T-n+1}^{T}\left(1+\frac{(-1)^m\phi^{(m)}(1)(T-j)^m}{T^m m!}+\mathcal{O}\left((1-j/T)^{m+1}\right)\right)\\ &\quad= 1-e^{\sum_{j=T-n+1}^{T}\log\left(1+\frac{(-1)^m\phi^{(m)}(1)(T-j)^m}{T^m m!}+\mathcal{O}\left((1-j/T)^{m+1}\right)\right)}\\ &\quad\geq 1-e^{\frac{(-1)^m\phi^{(m)}(1)}{T^m}\sum_{j=1}^{n-1}j^m+\mathcal{O}\left((n/T)^{m+1}\right)}\end{aligned}$$

as $\log(1+x)\leq x$.

It is a well known fact that

$$\sum_{j=1}^{n-1}j^k=\sum_{j=0}^{k}\frac{B_j}{(k+1)-j}\begin{pmatrix}k\\ j\end{pmatrix}n^{(k+1)-j},$$

where B_j is the jth Bernoulli number.

Suppose $\epsilon \geq \|\nu \mathbf{P}\left(\frac{1}{T}\right) \cdots \mathbf{P}(1) - \pi(1)\|_{TV}$, then

$$\begin{aligned}
\epsilon &\approx -\log(1-\epsilon) \\
&\geq \frac{(-1)^{m+1}\phi^{(m)}(1)}{T^m} \sum_{j=0}^{m} \frac{B_j}{(m+1)-j} \left(\begin{array}{c} m \\ \\ j \end{array} \right) n^{(m+1)-j} + \mathcal{O}\left((n/T)^{m+1}\right)
\end{aligned}$$

Thus confirming that the order of the adiabatic time in Theorem 7 is optimal

$$t_{ad}\left(\mathbf{P}(0), \mathbf{P}(1), \epsilon\right) = \mathcal{O}\left(\frac{t_{mix}^{\frac{m+1}{m}}\left(\mathbf{P}(1), \epsilon/2\right)}{\epsilon^{\frac{1}{m}}} \right).$$

Chapter 3

The Stable Adiabatic Time Versus The Mixing Time For Discrete Markov Chains

This chapter will consider both linear and nonlinear evolutions between two irreducible and aperiodic time-homogeneous Markov chains. We consider the corresponding time-inhomogeneous Markov chains from Chapter 2: Section 3.1 considers linear adiabatic evolutions and Section 3.3 considers nonlinear adiabatic evolutions. We now seek to to find a stricter form of stability of these Markov chains. We introduce a measurement called the stable adiabatic time and asymptotically bound this measurement by a function of the largest mixing time over the entire adiabatic evolution.

3.1 Linear Evolution

This section is going to define a new metric of stability for the time-inhomogeneous Markov chains introduced in Section 2.1. Recall that Definition 13 suggests for $\epsilon > 0$ and any $T \geq t_{ad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ that any probability distribution will evolve under consecutive applications of $\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}$ to an ϵ-ball around $\pi_{\mathbf{1}}$ in the space of probability distributions with respect to the total variation norm. We desire a stronger notion of stability in this paper to match the description of the quantum adiabatic theorem mentioned in [9]. We want to select T large enough so that starting at $\pi_{\mathbf{0}}$, the distribution will evolve under consecutive applications of $\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}$ within an ϵ-corridor of $\pi_{\frac{\mathbf{k}}{\mathbf{T}}}$ for $1 \leq k \leq T$. This leads us to the following definition.

Definition 20 *Given $\epsilon > 0$, a time $t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ is called the <u>stable adiabatic time</u> for a linear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ if it is the least such $T \in \mathbb{N}$ such that*

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} < \epsilon \tag{3.1}$$

for $1 \leq k \leq T$.

The goal of this section is to find an asymptotic bound for the stable adiabatic time with respect to the maximum mixing time over all the probability transition matrices in the linear adiabatic evolution. For $\epsilon > 0$ we let

$$t_{mix}(\epsilon) = \sup_{s \in [0,1]} \{t_{mix}(\mathbf{P_s}, \epsilon)\} \tag{3.2}$$

and we seek our bound in terms of this $t_{mix}(\epsilon)$.

The following theorem gives us insight into the nature of the stable adiabatic time. Its proof is given in Section 3.2.

Theorem 8 *Given a time-inhomogeneous, discrete-time Markov chain governed by a linear adiabatic evolution between the irreducible and aperiodic* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *and given* $\delta \in (0, 1]$, *for any* $\epsilon > 0$,

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{1}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon$$

for

$$T \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\delta}, \tag{3.3}$$

and $\delta \leq k/T \leq 1$.

If we can find $T \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\delta}$ that satisfies $\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{1}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon$ for $k/T \in [0, \delta]$, then this value of T would be an upper bound for the stable adiabatic time. Notice that we can choose $\delta \in (0, 1]$ to be as small as we like. If we find a bound of $\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{1}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}$ in terms of $\|\pi_{\mathbf{0}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}$ then we can use the continuity of $\pi_{\mathbf{s}}$ at $s = 0$ to find this value of T. We devote the following propositions to this endeavor and their proofs are shown in Section 3.2.

Proposition 4 *For* $1 \leq k \leq T$

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{1}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{(k+1)^2}{2T}. \tag{3.4}$$

Now we can use the continuity of $\pi_{\mathbf{s}}$ at $s = 0$ to find an appropriate bound for $\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{1}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV}$ for $0 \leq k/T \leq \delta$. We devote the following proposition to the discovery of how $\pi_{\mathbf{s}}$ is continuous at $s = 0$. The spectral structure of $\mathbf{P_0}$ is crucial to this development. Its proof is given in Section 3.2

Proposition 5 $\pi_{\mathbf{s}}$ *is continuous with respect to the total variation norm at* $s = 0$. *In particular, for* $\epsilon > 0$ *if we let* σ *be the smallest nonzero singular value of* $\mathbb{I} - \mathbf{P_0}$, *then if*

$$\delta = \frac{\epsilon\sigma}{2n^{3/2}} \tag{3.5}$$

we have for all $s \leq \delta$, $\|\pi_{\mathbf{s}} - \pi_{\mathbf{0}}\|_{TV} \leq \epsilon$.

Now we can use Proposition 3 along with the fact that $t_{mix}(\mathbf{P_0}, \epsilon) \leq t_{mix}(\epsilon)$ to derive the following Corollary to Proposition 5.

Corollary 2 *$\pi_{\mathbf{s}}$ is continuous with respect to the total variation norm at $s = 0$. In particular, for $0 < \epsilon < 1/\sqrt{n}$ if*

$$\delta = \frac{\epsilon(1 - \sqrt{n}\epsilon)}{4n^{3/2} t_{mix}(\epsilon/2)} \tag{3.6}$$

we have for all $s \leq \delta$, $\|\pi_{\mathbf{s}} - \pi_{\mathbf{0}}\|_{TV} \leq \epsilon/2$.

We now have all the necessary tools to find a bound for the stable adiabatic time. The proof is in Section 3.2.

Theorem 9 *Given a time-inhomogeneous, discrete-time Markov chain governed by a linear adiabatic evolution between the irreducible and aperiodic $\mathbf{P_0}$ and $\mathbf{P_1}$, for any $\epsilon > 0$,*

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = O\left(\frac{t^4_{mix}(\epsilon/2)}{\epsilon^3}\right). \tag{3.7}$$

We will see in Section 4.1 that this result somewhat reaffirms what has been shown in the Quantum Adiabatic Theorem in [3], but a main difference is that the inverse spectral gap bound for the quantum system is replaced with a mixing time bound in our result. Our result also has an extra multiple of $1/\epsilon$. Notice that the inverse spectral gap was a natural choice for the Quantum Adiabatic Theorem due to the Hamiltonian matrix being self-adjoint. For general, not necessarily reversible, Markov Chains, the Adiabatic Theorem is expressed using mixing times.

3.2 Proofs

3.2.1 Proof of Theorem 8

To develop the tools for this theorem, we consider the following treatment of our probability transition matrices. If we are given $s \in (0, 1]$, then we see that

$$\mathbf{P_t} = \left(1 - \frac{t}{s}\right)\mathbf{P_0} + \frac{t}{s}\mathbf{P_s}$$

for all $t \in [0, s]$.

Defining $\mathbf{P_t^{(s)}} = \mathbf{P_{st}}$, we see that

$$\mathbf{P_t^{(s)}} = (1 - t)\mathbf{P_0^{(s)}} + t\mathbf{P_1^{(s)}}$$

for all $t \in [0, 1]$. We also define $\pi_{\mathbf{t}}{}^{(s)} = \pi_{\mathbf{st}}$.

We see that $\{\mathbf{P_t^{(s)}}\}_{t\in[0,1]}$ is a class of probability transition matrices where $\mathbf{P_0} = \mathbf{P_0^{(s)}}$ and $\mathbf{P_s} = \mathbf{P_1^{(s)}}$.

Since the time-homogeneous Markov chains determined by $\mathbf{P_0}$ and $\mathbf{P_s}$ are irreducible and aperiodic, we can consider a time-inhomogeneous, discrete-time Markov chain governed by adiabatic evolution between these two time-homogeneous Markov chains.

Now let $\epsilon > 0$ and $\delta \in (0, 1]$.

For $s \in [\delta, 1]$ we have that $T^* = t_{ad}(\mathbf{P_0^{(s)}}, \mathbf{P_1^{(s)}}, \epsilon)$ is the adiabatic time between $\mathbf{P_0^{(s)}}$ and $\mathbf{P_1^{(s)}}$.

This tells us that

$$\max_{\nu} \|\nu \mathbf{P}^{(\mathbf{s})}_{\frac{\mathbf{1}}{\mathbf{T}^*}} \mathbf{P}^{(\mathbf{s})}_{\frac{\mathbf{2}}{\mathbf{T}^*}} \cdots \mathbf{P}^{(\mathbf{s})}_{\mathbf{1}} - \pi_{\mathbf{1}}{}^{(s)}\|_{TV} \leq \epsilon.$$

Because $\pi_{\mathbf{0}}{}^{(s)}$ is a specific distribution, we have that

$$\begin{aligned}
\epsilon &\geq \|\pi_{\mathbf{0}}{}^{(s)} \mathbf{P}^{(\mathbf{s})}_{\frac{\mathbf{1}}{\mathbf{T}^*}} \mathbf{P}^{(\mathbf{s})}_{\frac{\mathbf{2}}{\mathbf{T}^*}} \cdots \mathbf{P}^{(\mathbf{s})}_{\mathbf{1}} - \pi_{\mathbf{1}}{}^{(s)}\|_{TV} \\
&= \|\pi_{\mathbf{0}}{}^{(s)} \mathbf{P}^{(\mathbf{s})}_{\frac{(\mathbf{1}/\mathbf{s})}{(\mathbf{T}^*/\mathbf{s})}} \mathbf{P}^{(\mathbf{s})}_{\frac{(\mathbf{2}/\mathbf{s})}{(\mathbf{T}^*/\mathbf{s})}} \cdots \mathbf{P}^{(\mathbf{s})}_{\frac{(\mathbf{T}^*/\mathbf{s})}{(\mathbf{T}^*/\mathbf{s})}} - \pi^{(\mathbf{s})}_{\mathbf{1}}\|_{TV} \\
&= \|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{(\mathbf{T}^*/\mathbf{s})}} \mathbf{P}_{\frac{\mathbf{2}}{(\mathbf{T}^*/\mathbf{s})}} \cdots \mathbf{P}_{\frac{\mathbf{s}(\mathbf{T}^*/\mathbf{s})}{(\mathbf{T}^*/\mathbf{s})}} - \pi_{\frac{\mathbf{s}(\mathbf{T}^*/\mathbf{s})}{(\mathbf{T}^*/\mathbf{s})}}\|_{TV}.
\end{aligned}$$

Clearly if $T = t_{ad}(\mathbf{P}^{(\mathbf{s})}_{\mathbf{0}}, \mathbf{P}^{(\mathbf{s})}_{\mathbf{1}}, \epsilon)/s$, then

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \mathbf{P}_{\frac{\mathbf{2}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{sT}}{\mathbf{T}}} - \pi_{\frac{\mathbf{sT}}{\mathbf{T}}}\|_{TV} \leq \epsilon.$$

We showed in Theorem 4 that for $\epsilon > 0$

$$t_{ad}(\mathbf{P}^{(\mathbf{s})}_{\mathbf{0}}, \mathbf{P}^{(\mathbf{s})}_{\mathbf{1}}, \epsilon) \leq \frac{2t^2_{mix}(\mathbf{P}_{\mathbf{s}}, \epsilon/2)}{\epsilon}.$$

It follows that for $\epsilon > 0$

$$t_{ad}(\mathbf{P}^{(\mathbf{s})}_{\mathbf{0}}, \mathbf{P}^{(\mathbf{s})}_{\mathbf{1}}, \epsilon) \leq \frac{2t^2_{mix}(\epsilon/2)}{\epsilon}.$$

For $\epsilon > 0$ if we let T any integer such that

$$T \geq \frac{2t^2_{mix}(\epsilon/2)}{\epsilon\delta}$$

we have

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}}\mathbf{P}_{\frac{\mathbf{2}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}\leq\epsilon$$

for all $\delta \leq k/T \leq 1$.

3.2.2 Proof of Proposition 4

Because $\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{j}}{\mathbf{T}}} = \pi_{\mathbf{0}} + \frac{j}{T}\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})$ for $1 \leq j \leq k$ we notice that

$$\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{j}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}=\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}+\frac{j}{T}\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}$$

for $1 \leq j \leq k-1$ and

$$\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}=(\pi_{\mathbf{0}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}})+\frac{k}{T}\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0}).$$

Using the convention $\mathbf{P_{j+1}}\cdots\mathbf{P_k}=\mathbb{I}$ when $j \geq k$, we would see that

$$\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}=(\pi_{\mathbf{0}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}})+\sum_{j=1}^{k}\frac{j}{T}\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}.$$

Taking the total variation norm to either side of the inequality, using the triangle inequality and pulling out constants, we see that

$$\begin{aligned}\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}&=\|(\pi_{\mathbf{0}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}})+\sum_{j=1}^{k}\frac{j}{T}\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}\\&\leq\|\pi_{\mathbf{0}}-\pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}+\sum_{j=1}^{k}\frac{j}{T}\|\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV}.\end{aligned}$$

Notice that for $1 \leq j \leq k-1$

$$\pi_{\mathbf{0}}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}=\pi_{\mathbf{0}}\mathbf{P_1}\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}-\pi_{\mathbf{0}}\mathbf{P_0}\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}}\cdots\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}$$

is the difference between two probability distributions and

$$\pi_{\mathbf{0}}(\mathbf{P}_{\mathbf{1}} - \mathbf{P}_{\mathbf{0}}) = \pi_{\mathbf{0}}\mathbf{P}_{\mathbf{1}} - \pi_{\mathbf{0}}\mathbf{P}_{\mathbf{0}}$$

is also the difference between two probability distributions.

Because we are taking the total variation norm to the difference of two probability distributions we see that $\|\cdot\|_{TV} = \frac{1}{2}\|\cdot\|_1$ where $\|\cdot\|_1$ is the l_1-norm.

We have that for probability distributions μ and ν, $\|\mu - \nu\|_{TV} = \frac{1}{2}\|\mu - \nu\|_1 \leq \frac{1}{2}(\|\mu\|_1 + \|\nu\|_1) \leq 1$.

This tells us that $\|\pi_{\mathbf{0}}(\mathbf{P}_{\mathbf{1}} - \mathbf{P}_{\mathbf{0}})\mathbf{P}_{\frac{\mathbf{j+1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq 1$ for $1 \leq j \leq k$.

We see then that

$$\begin{aligned}
\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} &\leq \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \sum_{j=1}^{k} \frac{j}{T} \\
&= \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{1}{T}\sum_{j=1}^{k} j \\
&= \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{k(k+1)}{2T} \\
&\leq \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{(k+1)^2}{2T}.
\end{aligned}$$

3.2.3 Proof of Proposition 5

We begin with the creation of an orthonormal basis of eigenvectors associated with $(\mathbb{I} - \mathbf{P}_{\mathbf{0}})(\mathbb{I} - \mathbf{P}_{\mathbf{0}})^T$ by a singular value decomposition similar to the process we mentioned in Proposition 3.

Here we let $\sigma_1 \geq \cdots \geq \sigma_{n-1} = \sigma$ be the positive singular values of $(\mathbb{I} - \mathbf{P_0})$ with respect to the Euclidean inner product. This implies that there exists an orthonormal basis $\{\mathbf{v_1}, \cdots, \mathbf{v_n}\}$ such that $\mathbf{v_j}(\mathbb{I} - \mathbf{P_0})(\mathbb{I} - \mathbf{P_0})^T = \sigma_j^2 \mathbf{v_j}$ for $1 \leq j \leq n-1$ and $\mathbf{v_n}(\mathbb{I} - \mathbf{P_0})(\mathbb{I} - \mathbf{P_0})^T = \mathbf{0}$.

Here $\mathbf{v_n} = \pi_\mathbf{0}/\|\pi_\mathbf{0}\|_2$.

To show continuity at $s = 0$ let $\epsilon > 0$ and first notice that for any $s \in [0, 1]$, $(\pi_\mathbf{s} - \pi_\mathbf{0})(\mathbb{I} - \mathbf{P_0}) = s\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})$.

Using the Euclidean norm, we see that if $\mathbf{P_0} \neq \mathbf{P_1}$ and $s \neq 0$, then

$$\frac{\|(\pi_\mathbf{s} - \pi_\mathbf{0})(\mathbb{I} - \mathbf{P_0})\|_2}{\|\pi_\mathbf{s} - \pi_\mathbf{0}\|_2} = s\frac{\|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\|\pi_\mathbf{s} - \pi_\mathbf{0}\|_2}.$$

Throughout this proof we will use $< \cdot, \cdot >$ as the Euclidean inner product.

For $1 \leq j \leq n$ let $c_j = < \pi_\mathbf{s} - \pi_\mathbf{0}, \mathbf{v_j} >$. Then we see that $\pi_\mathbf{s} - \pi_\mathbf{0} = \sum_{j=1}^{n} c_j \mathbf{v_j}$.

We have that

$$\begin{aligned}
\frac{\|(\pi_{\mathbf{s}}-\pi_{\mathbf{0}})(\mathbb{I}-\mathbf{P_0})\|_2^2}{\|\pi_{\mathbf{s}}-\pi_{\mathbf{0}}\|_2^2} &= \frac{<(\pi_{\mathbf{s}}-\pi_{\mathbf{0}})(\mathbb{I}-\mathbf{P_0}),(\pi_{\mathbf{s}}-\pi_{\mathbf{0}})(\mathbb{I}-\mathbf{P_0})>}{<\pi_{\mathbf{s}}-\pi_{\mathbf{0}},\pi_{\mathbf{s}}-\pi_{\mathbf{0}}>} \\
&= \frac{<\pi_{\mathbf{s}}-\pi_{\mathbf{0}},(\pi_{\mathbf{s}}-\pi_{\mathbf{0}})(\mathbb{I}-\mathbf{P_0})(\mathbb{I}-\mathbf{P_0})^T>}{<\pi_{\mathbf{s}}-\pi_{\mathbf{0}},\pi_{\mathbf{s}}-\pi_{\mathbf{0}}>} \\
&= \frac{<\sum_{j=1}^n c_j\mathbf{v_j},\sum_{j=1}^{n-1}\sigma_j^2 c_j\mathbf{v_j}>}{<\sum_{j=1}^n c_j\mathbf{v_j},\sum_{j=1}^n c_j\mathbf{v_j}>} \\
&= \frac{\sum_{j=1}^{n-1}\sigma_j^2 c_j^2}{\sum_{j=1}^n c_j^2} \\
&\geq \sigma_{n-1}^2\frac{\sum_{j=1}^{n-1} c_j^2}{\sum_{j=1}^n c_j^2} \\
&= \sigma_{n-1}^2\left(1-\frac{c_n^2}{\sum_{j=1}^n c_j^2}\right) \\
&= \sigma_{n-1}^2\left(1-\left(\frac{<\pi_{\mathbf{s}}-\pi_{\mathbf{0}},\mathbf{v_n}>}{\|\pi_{\mathbf{s}}-\pi_{\mathbf{0}}\|_2}\right)^2\right).
\end{aligned}$$

If we let $\mathbf{w}(s)=(\pi_{\mathbf{s}}-\pi_{\mathbf{0}})/\|\pi_{\mathbf{s}}-\pi_{\mathbf{0}}\|_2$ then we see that

$$\sigma_{n-1}^2\left(1-(<\mathbf{w}(s),\mathbf{v_n}>)^2\right)\leq s^2\frac{\|\pi_{\mathbf{s}}(\mathbf{P_1}-\mathbf{P_0})\|_2^2}{\|\pi_{\mathbf{s}}-\pi_{\mathbf{0}}\|_2^2}.$$

Because $\mathbf{w}(s)$ and $\mathbf{v_n}$ are unit vectors, we can use the fact that

$$\|\mathbf{w}(s)\|_2^2-2<\mathbf{w}(s),\mathbf{v_n}>+\|\mathbf{v_n}\|_2^2=\|\mathbf{w}(s)-\mathbf{v_n}\|_2^2$$

to show that

$$1-<\mathbf{w}(s),\mathbf{v_n}>=\frac{1}{2}\|\mathbf{w}(s)-\mathbf{v_n}\|_2^2$$

and we can use the fact that

$$\|\mathbf{w}(s)\|_2^2+2<\mathbf{w}(s),\mathbf{v_n}>+\|\mathbf{v_n}\|_2^2=\|\mathbf{w}(s)+\mathbf{v_n}\|_2^2$$

to show that

$$1+ < \mathbf{w}(s), \mathbf{v_n} >= \frac{1}{2}\|\mathbf{w}(s) + \mathbf{v_n}\|_2^2.$$

From this we see that $1 - (< \mathbf{w}(s), \mathbf{v_n} >)^2 = \|\mathbf{w}(s) - \mathbf{v_n}\|_2^2 \cdot \|\mathbf{w}(s) + \mathbf{v_n}\|_2^2/4$. Plugging this into our previous equation, we can see that

$$\frac{\sigma_{n-1}^2}{4}\|\mathbf{w}(s) - \mathbf{v_n}\|_2^2 \cdot \|\mathbf{w}(s) + \mathbf{v_n}\|_2^2 \leq s^2 \frac{\|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2^2}{\|\pi_\mathbf{s} - \pi_\mathbf{0}\|_2^2}.$$

After performing some basic algebra we see that

$$\|\pi_\mathbf{s} - \pi_\mathbf{0}\|_2 \leq \frac{2s\|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1}\|\mathbf{w}(s) - \mathbf{v_n}\|_2 \cdot \|\mathbf{w}(s) + \mathbf{v_n}\|_2}.$$

Notice that $< \mathbf{w}(s), \mathbf{1} > /\sqrt{n} = 0$ and $< \mathbf{v_n}, \mathbf{1} > /\sqrt{n} = 1/\left(\sqrt{n}\|\pi_\mathbf{0}\|_2\right)$ for all $s \in [0, 1]$. Because these are the scalar components of the projections of $\mathbf{w}(s)$ and $\mathbf{v_n}$ onto $\mathbf{1}$ respectively, we see that the minimum possible value for $\|\mathbf{w}(s) - \mathbf{v_n}\|_2$ and $\|\mathbf{w}(s) + \mathbf{v_n}\|_2$ is at least $1/\left(\sqrt{n}\|\pi_\mathbf{0}\|_2\right)$.

We now have that

$$\begin{aligned}
\|\pi_\mathbf{s} - \pi_\mathbf{0}\|_2 &\leq \frac{2sn\|\pi_\mathbf{0}\|_2^2 \cdot \|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1}} \\
&\leq \frac{2sn\|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1}} \\
&= \frac{2sn\|\pi_\mathbf{s}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma}.
\end{aligned}$$

Again for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x}$ and $\mathbf{y}$ are probability measures, we see that

$$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_{TV} \leq \frac{\sqrt{n}}{2}\|\mathbf{x} - \mathbf{y}\|_2.$$

This will imply that

$$\|\pi_{\mathbf{s}} - \pi_{\mathbf{0}}\|_{TV} \leq \frac{2sn^{3/2}\|\pi_{\mathbf{s}}(\mathbf{P_1} - \mathbf{P_0})\|_{TV}}{\sigma_{n-1}}.$$

Because $\pi_{\mathbf{s}}(\mathbf{P_1} - \mathbf{P_0}) = \pi_{\mathbf{s}}\mathbf{P_1} - \pi_{\mathbf{s}}\mathbf{P_0}$ is the difference of two probability distributions, we see that $\|\cdot\|_{TV} = \frac{1}{2}\|\cdot\|_1$ where $\|\cdot\|_1$ is the l_1-norm. This implies that

$$\|\pi_{\mathbf{s}}(\mathbf{P_1} - \mathbf{P_0})\|_{TV} = \frac{1}{2}\|\pi_{\mathbf{s}}\mathbf{P_1} - \pi_{\mathbf{s}}\mathbf{P_0}\|_1 \leq \frac{1}{2}\left(\|\pi_{\mathbf{s}}\mathbf{P_1}\|_1 + \|\pi_{\mathbf{s}}\mathbf{P_0}\|_1\right) \leq 1.$$

This shows that

$$\|\pi_{\mathbf{s}} - \pi_{\mathbf{0}}\|_{TV} \leq \frac{2sn^{3/2}}{\sigma}.$$

Clearly if $\epsilon > 0$, then

$$s \leq \delta = \frac{\epsilon\sigma}{2n^{3/2}}$$

implies $\|\pi_{\mathbf{s}} - \pi_{\mathbf{0}}\|_{TV} \leq \epsilon$.

This shows that $\pi_{\mathbf{s}}$ is continuous at $s = 0$.

3.2.4 Proof of Theorem 9

We first provide a sketch of the proof followed by the technical details. Our proof is based on the results of Theorem 8, Proposition 4, and Corollary 2. Specifically, we divide our proof into two cases. In the first case, we will show how to select T and δ in order to satisfy the two conditions in Theorem 8, namely:

$$T \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\delta},$$

and

$$\delta \leq k/T \leq 1.$$

Therefore, by Theorem 8, we have

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon.$$

However, the selected T is not yet $t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ since this only holds for k such that

$$\delta \leq k/T \leq 1.$$

In the second case, we will use the results of Proposition 4 and Corollary 2 to show that for the same selected T and δ,

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon,$$

even in the case when

$$k/T \leq \delta < 1.$$

Therefore, we conclude that the selected T is a sufficient condition for $t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$.

We now proceed with the details of the proof, starting with the first case. Let $\epsilon > 0$. For this fixed ϵ, we choose T be an integer such that

$$\begin{aligned} T &\geq \frac{4t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{4t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{1}{\epsilon} \\ &= \left(\frac{2t_{mix}^2(\epsilon/2)}{\epsilon\sqrt{\epsilon}} + \frac{1}{\sqrt{\epsilon}}\right)^2. \end{aligned}$$

This implies

$$\sqrt{T} \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\sqrt{\epsilon}} + \frac{1}{\sqrt{\epsilon}}.$$

Multiplying either side by $\sqrt{\epsilon}$ and subtracting 1 from either side, we obtain

$$\sqrt{\epsilon}\sqrt{T} - 1 \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon}.$$

Notice that $\sqrt{\epsilon}\sqrt{T} - 1 > 0$ because

$$\frac{2t_{mix}^2(\epsilon/2)}{\epsilon} > 0.$$

Dividing either side of the above inequality by $\sqrt{\epsilon}\sqrt{T} - 1$, we obtain

$$1 \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\left(\sqrt{\epsilon}\sqrt{T} - 1\right)}.$$

Multiplying either side by T, we obtain

$$T \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\left(\sqrt{\frac{\epsilon}{T}} - \frac{1}{T}\right)}.$$

Now, let

$$\delta = \sqrt{\frac{\epsilon}{T}} - \frac{1}{T},$$

then clearly

$$T \geq \frac{2t_{mix}^2(\epsilon/2)}{\epsilon\delta}.$$

Next, let k be an integer such that

$$\delta = \sqrt{\frac{\epsilon}{T}} - \frac{1}{T} \leq \frac{k}{T} \leq 1.$$

Then by Theorem 8, we conclude that

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon.$$

Now in the second (complementary) case, i.e., when $k/T \leq \delta < 1$, we will show that for the same selected $\delta = \sqrt{\frac{\epsilon}{T}} - \frac{1}{T}$, and T, it is still true that:

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon,$$

Let k be an integer such that

$$0 \leq \frac{k}{T} \leq \sqrt{\frac{\epsilon}{T}} - \frac{1}{T} = \delta.$$

Then,

$$\frac{k+1}{T} \leq \sqrt{\frac{\epsilon}{T}}.$$

Using Proposition 4, we have

$$\begin{aligned}
\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} &\leq \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{(k+1)^2}{2T} \\
&= \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{T}{2}\left(\frac{k+1}{T}\right)^2 \\
&\leq \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{T}{2}\left(\sqrt{\frac{\epsilon}{T}}\right)^2 \\
&= \|\pi_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\mathbf{0}}\|_{TV} + \frac{\epsilon}{2}.
\end{aligned}$$

Next, from Corollary 2, as long as $\epsilon < 1/\sqrt{n}$ and $\sqrt{\frac{\epsilon}{T}} - \frac{1}{T} \leq \frac{\epsilon(1-\sqrt{n}\epsilon)}{4n^{3/2} t_{mix}(\epsilon/2)}$, we have

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{\mathbf{k}}{\mathbf{T}}}\|_{TV} \leq \epsilon$$

for

$$0 \leq \frac{k}{T} \leq \sqrt{\frac{\epsilon}{T}} - \frac{1}{T}.$$

It should be clear that as $\epsilon \to 0$,

$$\sqrt{\frac{\epsilon}{T}} - \frac{1}{T} \leq \frac{\epsilon(1-\sqrt{n}\epsilon)}{4n^{3/2}t_{mix}(\epsilon/2)}$$

when

$$T \geq \frac{4t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{4t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{1}{\epsilon}.$$

This tells us that as $\epsilon \to 0$,

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{4t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{4t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{1}{\epsilon}.$$

We conclude that

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = O\left(\frac{t_{mix}^4(\epsilon/2)}{\epsilon^3}\right).$$

3.3 Nonlinear Evolution

We devote this section to the creation of a stable adiabatic time for the time-inhomogeneous Markov chains introduced in Section 2.3. We will again find an asymptotic bound of the stable adiabatic time as $\epsilon \to 0$ with respect to an inverse power of ϵ multiplied by a power of the largest mixing time over the entire adiabatic evolution for a subclass of continuous functions.

Recall how we defined a nonlinear evolution between the probability transition matrices of two irreducible, aperiodic, discrete-time, time-homogeneous Markov chains in Definition 14. Denoting the initial and the final matrices $\mathbf{P_0}$ and $\mathbf{P_1}$ respectively, we defined for $t \in [0,1]$ a class of probability transition matrices $\{\mathbf{P}_{\phi(t)}\}_{t\in[0,1]}$ such that

$$\mathbf{P}_t = (1-\phi(t))\mathbf{P_0} + \phi(t)\mathbf{P_1}$$

where $\phi : [0,1] \to [0,1]$ is a continuous function such that $\phi(0) = 0$ and $\phi(1) = 1$. We similarly define $\pi_{\phi(t)}$ as the stationary distribution of $\mathbf{P}_{\phi(t)}$.

We used these matrices to define a class of time-inhomogeneous Markov chains. Any Markov chain in this class was said to be governed by a nonlinear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function ϕ. We defined the adiabatic time in Section 2.3 as the smallest integer T guaranteeing that any distribution will evolve under consecutive applications of $\mathbf{P}_{\phi(\frac{k}{T})}$ for $1 \leq k \leq T$ to an ϵ-ball of the stationary distribution of $\mathbf{P_1}$. As we did in Section 3.1, we are now going to demand that $\mathbf{P}_{\phi(\frac{k}{T})}$ remain in an ϵ-corridor of $\pi_{\phi(\frac{k}{T})}$ for all $1 \leq k \leq T$. We now define this new metric of stability.

Definition 21 *Given $\epsilon > 0$, a time $t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon)$ is called the <u>stable adiabatic time</u> for a nonlinear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ by the function ϕ if it is the least such $T \in \mathbb{N}$ such that*

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\phi(\frac{1}{T})} \cdots \mathbf{P}_{\phi(\frac{k}{T})} - \pi_{\phi(\frac{k}{T})}\|_{TV} < \epsilon \tag{3.8}$$

for $1 \leq k \leq T$.

Again, our goal is to seek an asymptotic bound for the stable adiabatic time only now it is with respect to the maximum mixing time over all the transition probability matrices in the nonlinear adiabatic evolution. For $\epsilon > 0$ we again let

$$t_{mix}(\epsilon) = \sup_{s \in [0,1]} \{t_{mix}(\mathbf{P_s}, \epsilon)\}$$

and we seek our bound in terms of this $t_{mix}(\epsilon)$.

To find this result for any continuous ϕ proved to be difficult. As we did in Section 2.3, we attempted to restrict the types of functions ϕ that we consider so that they are contained in a dense subset of the continuous functions on $[0,1]$, namely the Lipschitz continuous functions with finite Lipschitz constant.

To find an asymptotic bound for Lipschitz continuous functions also proved to be difficult, so we restrict even further by considering bi-Lipschitz continuous functions ϕ with Lipschitz constant L, so we first define what it means to be bi-Lipschitz continuous.

Definition 22 *Assume $a, b \in \mathbb{R}$ with $a < b$. A function $\phi : [a, b] \to \mathbb{R}$ is bi-Lipschitz continuous if there exists a positive constant L, called the Lipschitz constant, so that*

$$\frac{1}{L}\left|x - y\right| \leq \left|\phi\left(x\right) - \phi\left(y\right)\right| \leq L\left|x - y\right| \tag{3.9}$$

for $x, y \in [a, b]$.

The following proposition outlines an important property of bi-Lipschitz continuous functions that we will use throughout this section. The proof of this proposition is given in Section 3.4.

Proposition 6 *If $\phi : [a, b] \to \mathbb{R}$ is a bi-Lipschitz continuous function with Lipschitz constant L, then ϕ is either strictly increasing or strictly decreasing on $[a, b]$.*

We can find an asymptotic bound for the stable adiabatic time if ϕ is a bi-Lipschitz continuous function with finite Lipschitz constant L. We do this using nearly the same process of that in Section 3.1, however, each theorem has a slightly different process. Instead of considering only bi-Lipschitz continuous functions on the interval $[0, 1]$ we will consider bi-Lipschitz continuous functions on the interval $[a, b]$ where $0 \leq a < b \leq 1$. We do this because we ultimately want to consider Lipschitz continuous functions on $[0, 1]$ that are not bi-Lipschitz and this generalization helps us in this cause. Notice that if we pick $a = 0$ and $b = 1$ we will have found an asymptotic result for the stable adiabatic time when ϕ is bi-Lipschitz continuous. We start by considering a general continuous function $\phi : [0,1] \to [0,1]$ with $\phi(0) = 0$ and $\phi(1) = 1$ such that ϕ is bi-Lipschitz continuous on a subinterval $[a,b]$ with $0 \leq a < b \leq 1$. The following theorem finds a bound for $\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$ when $\delta^* \leq a + k/T \leq b$ for some $\delta^* \in (a,b]$ when ϕ is strictly increasing on $[a,b]$. Its proof is given in Section 3.4. We will return to the case when ϕ is a continuous function on $[0,1]$ such that it is bi-Lipschitz continuous and strictly decreasing on $[a,b]$ later.

Theorem 10 *Suppose $a, b \in \mathbb{R}$ such that $0 \leq a < b \leq 1$. Given a time-inhomogeneous, discrete-time Markov chain governed by a nonlinear adiabatic evolution between the irreducible and aperiodic $\mathbf{P}_0$ and $\mathbf{P}_1$ by the continuous function ϕ such that ϕ is a strictly increasing, bi-Lipschitz continuous function on the interval $[a,b]$ having Lipschitz constant L and given $\delta^* \in (a,b]$, for any $\epsilon > 0$,*

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon$$

for

$$T \geq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon(\delta^* - a)}, \tag{3.10}$$

and $\delta^* \leq a + k/T \leq b$.

If we can find $T \geq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon(\delta^* - a)}$ that satisfies

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon$$

for $a + k/T \in [a, \delta^*]$, then this value of T would guarantee that the nonlinear adiabatic transition between a and $k_{max}/T \leq b$ is within an ϵ-corridor of $\pi_{\phi(t)}$. Notice that we can choose $\delta^* \in (a,b]$ to be as small as we like. If we find a bound of

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$$

in terms of $\|\pi_{\phi(a)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$ then we can use the continuity of $\pi_{\phi(\mathbf{s})}$ at $s = a$ to find this value of T. We devote the following propositions to this endeavor and their proofs are shown in Section 3.4.

Proposition 7 *For $1 \leq k \leq T$ and $\phi : [a, b] \to \mathbb{R}$ a strictly increasing, bi-Lipschitz continuous function with Lipschitz constant L,*

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \|\pi_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi(a)}\|_{TV} + L\frac{(k+1)^2}{2T}. \tag{3.11}$$

Now we can use the continuity of $\pi_{\phi(s)}$ at $s = a$ to find an appropriate bound for $\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$ for $a \leq a + k/T \leq \delta^*$ where δ^* is the unique number in $(a, b]$ such that $\phi(\delta^*) = \delta$, $\phi(t) < \delta$ for $t < \delta^*$ and $\phi(t) > \delta$ for $t > \delta^*$. We devote the following proposition to the discovery of how $\pi_{\phi(s)}$ is continuous at $s \in [a, b]$ when ϕ is a bi-Lipschitz continuous function with Lipschitz constant L. The spectral structure of $\mathbf{P_s}$ is crucial to this development. The proof is in Section 3.4.

Proposition 8 *For $a, b \in \mathbb{R}$ such that $0 \leq a < b \leq 1$ we have that π_ϕ is continuous with respect to the total variation norm on $[a, b]$ when ϕ is a bi-Lipschitz continuous function with Lipschitz constant L. In particular, for $\epsilon > 0$ if we let σ be the smallest nonzero singular value of $\mathbb{I} - \mathbf{P_s}$ over all $s^*, s \in [a, b]$, then if*

$$\delta^* = \frac{\epsilon\sigma}{2Ln^{3/2}} \tag{3.12}$$

we have for all $\left|s^* - s\right| \leq \delta^*$, $\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} \leq \epsilon$.

Now we can use Proposition 3 along with the fact that $t_{mix}(\mathbf{P_0}, \epsilon) \leq t_{mix}(\epsilon)$ to derive the following corollary to Proposition 8.

Corollary 3 *π_ϕ is continuous with respect to the total variation norm on $[a, b]$ when ϕ is a bi-Lipschitz continuous function with Lipschitz constant L. In particular, for $0 < \epsilon < 1/\sqrt{n}$ if*

$$\delta^* = \frac{\epsilon(1 - \sqrt{n}\epsilon)}{4Ln^{3/2} t_{mix}(\epsilon/2)} \tag{3.13}$$

we have for all $\left|s^* - s\right| \leq \delta^*$, $\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} \leq \epsilon/2$.

We now have all the necessary tools to find an asymptotic value of T with respect to ϵ that ensures

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon$$

for all $1 \leq k \leq T$ when ϕ is a strictly increasing bi-Lipschitz continuous function with Lipschitz constant L. The proof is in Section 3.4.

Theorem 11 *Suppose $a, b \in \mathbb{R}$ such that $0 \leq a < b \leq 1$. Given a time-inhomogeneous, discrete-time Markov chain governed by a nonlinear adiabatic evolution between the irreducible and aperiodic $\mathbf{P}_0$ and $\mathbf{P}_1$ by the continuous function ϕ such that ϕ is a strictly increasing, bi-Lipschitz continuous function on the interval $[a, b]$ having Lipschitz constant L we see that as $\epsilon \to 0$, if*

$$T \geq \frac{16L^5 t^4_{mix}(\epsilon/2)}{\epsilon^3} + \frac{8L^3 t^2_{mix}(\epsilon/2)}{\epsilon^2} + \frac{L}{\epsilon} \tag{3.14}$$

then we have

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} < \epsilon$$

for $a \leq a + k/T \leq b$.

Again, notice that if we select $a = 0$ and $b = 1$, we have found a bound on the adiabatic time when ϕ is a bi-Lipschitz continuous function on $[0, 1]$. We summarize this in the following corollary.

Corollary 4 *Given a time-inhomogeneous, discrete-time Markov chain governed by adiabatic evolution between two time-homogeneous, discrete-time, n-state, irreducible and aperiodic Markov chains with probability transition matrices* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *by the bi-Lipschitz function* ϕ *with Lipschitz constant* L*, for any* $\epsilon > 0$*,*

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = O\left(\frac{t_{mix}^4(\epsilon/2)}{\epsilon^3}\right). \tag{3.15}$$

Now we want to use this information to find a bound for

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$$

for $a \leq a + k/T \leq b$ when ϕ is a continuous function $\phi : [0, 1] \to [0, 1]$ with $\phi(0) = 0$ and $\phi(1) = 1$ such that ϕ is bi-Lipschitz continuous and strictly decreasing on a subinterval $[a, b]$ with $0 \leq a < b \leq 1$.

Consider the Markov chain that is governed by a nonlinear adiabatic evolution between $\mathbf{P_1}$ and $\mathbf{P_0}$ by the function ψ where $\psi(t) = 1 - \phi(t)$ for $t \in [0, 1]$.

Notice now that

$$\begin{aligned}
\mathbf{P}^*_{\psi(t)} &= (1 - \psi(t))\mathbf{P}_1 + \psi(t)\mathbf{P}_0 \\
&= \phi(t)\mathbf{P}_1 + (1 - \phi(t))\mathbf{P}_0 \\
&= \mathbf{P}_{\phi(t)}
\end{aligned}$$

and $\pi^*_{\psi(t)} = \pi_{\phi(t)}$.

Naturally this would imply that

$$\|\pi^*_{\psi(a)}\mathbf{P}^*_{\psi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}^*_{\psi\left(a+\frac{k}{T}\right)} - \pi^*_{\psi\left(a+\frac{k}{T}\right)}\|_{TV}$$
$$= \|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}$$

only now ψ is a bi-Lipschitz continuous function on $[a, b]$ with Lipschitz constant L that is strictly increasing. This would imply that Theorem 11 would also be applicable to functions where ϕ is strictly decreasing on $[a, b]$. I summarize this in the following corollary.

Corollary 5 *Suppose $a, b \in \mathbb{R}$ such that $0 \leq a < b \leq 1$. Given a time-inhomogeneous, discrete-time Markov chain governed by a nonlinear adiabatic evolution between the irreducible and aperiodic $\mathbf{P}_0$ and $\mathbf{P}_1$ by the continuous function ϕ such that ϕ is a bi-Lipschitz continuous function on the interval $[a, b]$ having Lipschitz constant L we see that as $\epsilon \to 0$, if*

$$T \geq \frac{16L^5 t^4_{mix}(\epsilon/2)}{\epsilon^3} + \frac{8L^3 t^2_{mix}(\epsilon/2)}{\epsilon^2} + \frac{L}{\epsilon}$$

then we have

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} < \epsilon$$

for $a \leq a + k/T \leq b$.

Our ultimate goal is to find a similar result for Lipschitz continuous functions with finite Lipschitz constant, but we will work incrementally toward this goal. I want to highlight the following property of bi-Lipschitz functions. This property will be important when defining the next class of functions for which we have a stable adiabatic theorem. The proof of the following proposition is given in Section 3.4.

Proposition 9 *Let $\phi_1 : [a,b] \to \mathbb{R}$ be a bi-Lipschitz continuous function with Lipschitz constant L_1 and let $\phi_2 : [b,c] \to \mathbb{R}$ be a bi-Lipschitz continuous function with Lipschitz constant L_2 such that $\phi_1(b) = \phi_2(b)$. Suppose that either both ϕ_1 and ϕ_2 are strictly increasing or both ϕ_1 and ϕ_2 are strictly decreasing. Define a function $\phi : [a,c] \to \mathbb{R}$ such that*

$$\phi(t) = \begin{cases} \phi_1(t) & \textit{if } t \in [a,b] \\ \phi_2(t) & \textit{if } t \in [b,c]. \end{cases}$$

Then we have that ϕ is a bi-Lipschitz continuous function with Lipschitz constant $\max\{L_1, L_2\}$.

Let $M \in \mathbb{N}$. Let $\mathcal{P}_M$ be a partition of $[0,1]$ into $2M-1$ intervals. By creating a partition, we pick numbers $0 = a_0 < a_1 < \cdots < a_{2M-2} < a_{2M-1} = 1$ to divide the set $[0,1]$. We will denote $I_i = [a_{i-1}, a_i]$. Now we want to consider continuous functions $\Phi = \phi_{M,\mathcal{P}_M} : [0,1] \to [0,1]$ so that there exists a partition $\mathcal{P}_M$ of the interval $[0,1]$ so that $\Phi = \Phi_{M,\mathcal{P}_M}$ is a piecewise defined function

$$\Phi(t) = \begin{cases} \phi_1(t) & \text{if } t \in I_1 \\ \phi_2(t) & \text{if } t \in I_2 \\ \vdots & \\ \phi_{2M-2}(t) & \text{if } t \in I_{2M-2} \\ \phi_{2M-1}(t) & \text{if } t \in I_{2M-1} \end{cases}$$

where $\phi_i : I_i \to [0,1]$ is a bi-Lipschitz continuous function with Lipschitz constant L_i and $\phi_i(a_i) = \phi_{i+1}(a_i)$.

Proposition 9 tells us that a minimal such partition exists, so we will assume that $\mathcal{P}_M$ is the minimal partition. We also know that since $\Phi(0) = 0$ and

$\Phi(1) = 1$, then Φ is strictly increasing on I_{2i-1} for $1 \leq i \leq M$ and strictly decreasing on I_{2i} for $1 \leq i \leq M-1$. We will use the results from Corollary 5 and Corollary 3 to find an asymptotic bound of the stable adiabatic time with respect to the largest mixing time for a time-inhomogeneous, discrete-time Markov chain governed by adiabatic evolution between two time-homogeneous, discrete-time, n-state, irreducible and aperiodic Markov chains with probability transition matrices $\mathbf{P_0}$ and $\mathbf{P_1}$ by the Lipschitz function Φ with Lipschitz constant $\max\{L_1, \cdots, L_{2M-1}\}$.

Theorem 12 *Given a time-inhomogeneous, discrete-time Markov chain governed by adiabatic evolution between two time-homogeneous, discrete-time, n-state, irreducible and aperiodic Markov chains with probability transition matrices* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *by the Lipschitz function* Φ *defined above with Lipschitz constant* $\max\{L_1, \cdots, L_{2M-1}\}$*, for any* $\epsilon > 0$*,*

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) = O\left(\frac{t_{mix}^4(\epsilon/2)}{\epsilon^3}\right). \tag{3.16}$$

Now to complete the process we must consider general Lipschitz continuous functions ϕ, however, this proved more difficult that I had hoped. We would like to take the Lipschitz constant to large so that depends on $1/\epsilon$ and find the subintervals of $[0, 1]$ where $\left|\phi(x) - \phi(y)\right| < \epsilon$ and make a bound of the stable adiabatic time or these intervals by suggesting that the adiabatic transition doesn't change our initial matrix that much in these intervals. The proof is likely to come in the near future, but getting the precise details ironed out for this manuscript was not possible.

3.4 Proofs

3.4.1 Proof of Proposition 6

Let $x, y \in [a, b]$ so that $x < y$ and assume that $\phi(x) = \phi(y)$. From the definition of bi-Lipschitz we see that

$$\frac{1}{L}\big|x - y\big| \leq \big|\phi\,(x) - \phi\,(y)\,\big|.$$

If $\phi(x) = \phi(y)$ then we see that the previous equation implies

$$\big|x - y\big| = 0.$$

But this can only happen if $x = y$. This contradicts our declaration that $x < y$. This would imply that $\phi(x) \neq \phi(y)$.

There exists a point $z \in [a, b]$ such that $\phi(x) < \phi(z)$ for all $x \in [a, b]$ because ϕ is a continuous function over a compact set. Assume that $z \in (a, b)$. Because ϕ is continuous and $\phi(z)$ is maximal, there exists some $z_1 < z$ and $z_2 > z$ such that $\phi(z_1) = \phi(z_2)$ and we showed this cannot happen. Therefore the maximum element must be $\phi(a)$ or $\phi(b)$. Through a similar argument, we see that the minimum element must also be $\phi(a)$ or $\phi(b)$. We also know that the minimum element and the maximal element cannot be the same, because that would imply that the function is constant, and not bi-Lipschitz.

First consider the case where $\phi(a)$ is minimal and $\phi(b)$ is maximal. For $x, y \in [a, b]$ such that $x < y$ assume that $\phi(x) > \phi(y)$. This would imply that $y \neq b$ because if $\phi(x) > \phi(b)$, then this contradicts our assertion that $\phi(b)$ is maximal. Because ϕ is continuous on the interval $[y, b]$ and $\phi(x) \in (\phi(y), \phi(b)]$, we can use the intermediate value theorem to claim that there exists a constant $c \in [y, b]$ so that $\phi(c) = \phi(x)$. But we now have two numbers $c, x \in [a, b]$ with $x < c$ and $\phi(x) = \phi(c)$. This is a contradiction, so we now know that $\phi(x) < \phi(y)$. We have that ϕ is strictly increasing

Next consider the case where $\phi(b)$ is minimal and $\phi(a)$ is maximal. For $x, y \in [a, b]$ such that $x < y$ assume that $\phi(x) < \phi(y)$. This would imply that $x \neq a$ because if $\phi(a) < \phi(y)$, then this contradicts our assertion that $\phi(a)$ is maximal. Because ϕ is continuous on the interval $[a, x]$ and $\phi(y) \in (\phi(x), \phi(a)]$, we can use the intermediate value theorem again to claim that there exists a constant $c \in [a, x]$ so that $\phi(c) = \phi(y)$. But we again have two numbers $c, y \in [a, b]$ with $c < y$ and $\phi(c) = \phi(y)$. This is a contradiction, so we know that $\phi(x) > \phi(y)$. We have that ϕ is strictly decreasing.

3.4.2 Proof of Theorem 10

We first pick a number $s \in (\phi(a), \phi(b)]$. Because ϕ is a strictly increasing, continuous function on $[a, b]$ with $\phi(a)$ the minimum value of ϕ on $[a, b]$ and $\phi(b)$ the maximum value of ϕ on $[a, b]$, we know from the intermediate value theorem that there exists $s^* \in (a, b]$ such that $\phi(s^*) = s$. We know that there is only one such $s^* \in (a, b]$ because ϕ is strictly increasing. We would also see that $\phi(t) < s$ for $t \in [a, s^*)$ and $\phi(t) > s$ for $t \in (s^*, b]$.

We next consider the following treatment of our probability transition matrices:

$$\mathbf{P}_{\phi(t)} = \left(1 - \frac{\phi(t) - \phi(a)}{s - \phi(a)}\right) \mathbf{P}_{\phi(a)} + \frac{\phi(t) - \phi(a)}{s - \phi(a)} \mathbf{P}_s$$

for all $t \in [a, s^*]$.

Defining $\mathbf{P}^{(\mathbf{s})}(t) = \mathbf{P}_{\phi(a+t(s^*-a))}$, we see that

$$\begin{aligned}\mathbf{P}^{(\mathbf{s})}(t) = &\left(1 - \frac{\phi(a+t(s^*-a)) - \phi(a)}{s-\phi(a)}\right)\mathbf{P}^{(\mathbf{s})}(0) \\ &+ \frac{\phi(a+t(s^*-a)) - \phi(a)}{s-\phi(a)}\mathbf{P}^{(\mathbf{s})}(1)\end{aligned}$$

for all $t \in [0,1]$. We also define $\pi^{(s)}(t) = \pi_{\phi(a+t(s^*-a))}$.

We see that $\{\mathbf{P}^{(\mathbf{s})}(t)\}_{t\in[0,1]}$ is a class of probability transition matrices where $\mathbf{P}^{(\mathbf{s})}(0) = \mathbf{P}_{\phi(a)}$ and $\mathbf{P}^{(\mathbf{s})}(1) = \mathbf{P}_s$, however, the transition function is no longer the bi-Lipschitz continuous function ϕ. The function of transition is now a function depending on s, $\psi_s : [0,1] \to [0,1]$ such that $\psi_s(0) = 0$, $\psi_s(1) = 1$ and

$$\psi_s(t) = \frac{\phi(a+t(s^*-a)) - \phi(a)}{s-\phi(a)}.$$

For $x, y \in [0,1]$ notice that

$$\begin{aligned}\left|\psi_s(x) - \psi_s(y)\right| &= \left|\frac{\phi(a+x(s^*-a)) - \phi(a)}{s-\phi(a)} - \frac{\phi(a+y(s^*-a)) - \phi(a)}{s-\phi(a)}\right| \\ &= \frac{1}{s-\phi(a)}\left|\phi(a+x(s^*-a)) - \phi(a+y(s^*-a))\right| \\ &\leq \frac{L}{s-\phi(a)}\left|x(s^*-a) - y(s^*-a)\right| \\ &= \frac{L(s^*-a)}{s-\phi(a)}\left|x-y\right|.\end{aligned}$$

Because ϕ is bi-Lipschitz continuous on $[a,b]$ we see that $(s^*-a)/(s-\phi(a)) = (s^*-a)/(\phi(s^*)-\phi(a)) \leq L$.

We see that for any $s \in (0,1]$, ψ_s is a Lipschitz continuous function with Lipschitz constant L^2.

Since the time-homogeneous Markov chains determined by $\mathbf{P^{(s)}}(0)$ and $\mathbf{P^{(s)}}(1)$ are irreducible and aperiodic, we can consider a time-inhomogeneous, discrete-time Markov chain governed by adiabatic evolution between these two time-homogeneous Markov chains by the Lipschitz continuous function ψ_s.

Now let $\epsilon > 0$ and $\delta \in (\phi(a), \phi(b)]$.

For $s \in [\delta, \phi(b)]$ we have that $T^* = t_{ad}(\mathbf{P^{(s)}}(0), \mathbf{P^{(s)}}(1), \epsilon)$ is the adiabatic time of a time-inhomogeneous Markov chain governed by the adiabatic evolution between $\mathbf{P^{(s)}}(0)$ and $\mathbf{P^{(s)}}(1)$ by the Lipschitz function ψ_s.

This tells us that

$$\max_{\nu} \|\nu \mathbf{P^{(s)}}\left(\frac{1}{T^*}\right)\mathbf{P^{(s)}}\left(\frac{2}{T^*}\right)\cdots \mathbf{P^{(s)}}(1) - \pi^{(s)}(1)\|_{TV} \leq \epsilon.$$

Because $\pi^{(s)}(0)$ is a specific distribution, we have that

$$\begin{aligned}
\epsilon &\geq \|\pi^{(s)}(0)\mathbf{P}^{(s)}\left(\frac{1}{T^*}\right)\mathbf{P}^{(s)}\left(\frac{2}{T^*}\right)\cdots \mathbf{P^{(s)}}(1) - \pi^{(s)}(1)\|_{TV} \\
&= \|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{(s^*-a)}{T^*}\right)}\mathbf{P}_{\phi\left(a+\frac{2(s^*-a)}{T^*}\right)}\cdots \mathbf{P}_{\phi\left(a+\frac{T^*(s^*-a)}{T^*}\right)} - \pi_{\phi\left(a+\frac{T^*(s^*-a)}{T^*}\right)}\|_{TV}.
\end{aligned}$$

Clearly if $T = t_{ad}(\mathbf{P^{(s)}}(0), \mathbf{P^{(s)}}(1), \epsilon)/(s^* - a)$, then

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\mathbf{P}_{\phi\left(a+\frac{2}{T}\right)}\cdots \mathbf{P}_{\phi(s^*)} - \pi_{\phi(s^*)}\|_{TV} \leq \epsilon.$$

Because ψ_s is a Lipschitz continuous function with Lipschitz constant L^2 we can reference Theorem 5 to see that for $\epsilon > 0$

$$t_{ad}(\mathbf{P^{(s)}}(0), \mathbf{P^{(s)}}(1), \epsilon) \leq \frac{4L^2 t_{mix}^2(\mathbf{P}_s, \epsilon/2)}{\epsilon}.$$

It follows that for $\epsilon > 0$

$$t_{ad}(\mathbf{P^{(s)}}(0), \mathbf{P^{(s)}}(1), \epsilon) \leq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon}.$$

This would imply that if T is any integer such that

$$T \geq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon(\delta^* - a)}$$

we have

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \mathbf{P}_{\phi\left(a+\frac{2}{T}\right)} \cdots \mathbf{P}_{\phi(s^*)} - \pi_{\phi(s^*)}\|_{TV} \leq \epsilon$$

for all $\delta^* \leq s^* \leq b$.

In particular, if we consider the values of $1 \leq k \leq T$ such that

$$\delta^* \leq a + k/T \leq b$$

we see that

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon.$$

3.4.3 Proof of Proposition 7

Because $\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{j}{T}\right)} = \pi_{\phi(a)} + \left(\phi\left(a+\frac{j}{T}\right) - \phi(a)\right)\pi_{\phi(a)}(\mathbf{P_1} - \mathbf{P_0})$ for $1 \leq j \leq k$ we notice that

$$\begin{aligned}
&\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{j}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)} \\
&\quad = \pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)} \\
&\qquad + \left(\phi\left(a+\frac{j}{T}\right) - \phi(a)\right)\pi_{\phi(a)}(\mathbf{P_1} - \mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}.
\end{aligned}$$

for $1 \leq j \leq k-1$ and

$$\begin{aligned}
&\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)} \\
&\qquad = (\pi_{\phi(a)} - \pi_{\phi\left(a+\frac{k}{T}\right)}) + \left(\phi\left(a+\frac{k}{T}\right) - \phi(a)\right)\pi_{\phi(a)}(\mathbf{P_1} - \mathbf{P_0}).
\end{aligned}$$

Using the convention $\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} = \mathbb{I}$ when $j \geq k$, we would see that

$$\begin{aligned}
&\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)} \\
&\quad = \left(\pi_{\phi(a)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\right) \\
&\quad + \sum_{j=1}^{k}\left(\phi\left(a+\frac{j}{T}\right) - \phi(a)\right)\pi_{\phi(a)}(\mathbf{P_1} - \mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}.
\end{aligned}$$

Taking the total variation norm to either side of the inequality, using the triangle inequality and pulling out constants, we see that

$$\begin{aligned}
&\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\\
&=\|\left(\pi_{\phi(a)}-\pi_{\phi\left(a+\frac{k}{T}\right)}\right)\\
&+\sum_{j=1}^{k}\left(\phi\left(a+\frac{j}{T}\right)-\phi(a)\right)\pi_{\phi(a)}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\\
&\leq\|\pi_{\phi(a)}-\pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\\
&+\sum_{j=1}^{k}\left(\phi\left(a+\frac{j}{T}\right)-\phi(a)\right)\|\pi_{\phi(a)}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}.
\end{aligned}$$

Notice that for $1\leq j\leq k-1$

$$\begin{aligned}
\pi_{\phi(a)}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}&=\pi_{\phi(a)}\mathbf{P_1}\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}\\
&-\pi_{\phi(a)}\mathbf{P_0}\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}
\end{aligned}$$

is the difference between two probability distributions and

$$\pi_{\phi(a)}(\mathbf{P_1}-\mathbf{P_0})=\pi_{\phi(a)}\mathbf{P_1}-\pi_{\phi(a)}\mathbf{P_0}$$

is also the difference between two probability distributions.

Because we are taking the total variation norm to the difference of two probability distributions we see that $\|\cdot\|_{TV}=\frac{1}{2}\|\cdot\|_1$ where $\|\cdot\|_1$ is the l_1-norm.

We have that for probability distributions μ and ν, $\|\mu-\nu\|_{TV}=\frac{1}{2}\|\mu-\nu\|_1\leq\frac{1}{2}(\|\mu\|_1+\|\nu\|_1)\leq 1$.

This tells us that $\|\pi_{\phi(a)}(\mathbf{P_1}-\mathbf{P_0})\mathbf{P}_{\phi\left(a+\frac{(j+1)}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\leq 1$ for $1\leq j\leq k$.

We see then that

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\leq\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\mathbf{0}}\|_{TV}$$
$$+\sum_{j=1}^{k}\left(\phi\left(a+\frac{j}{T}\right)-\phi(a)\right).$$

Remember that ϕ is a bi-Lipschitz continuous function with Lipschitz constant L, we see that

$$\begin{aligned}
&\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV}\\
&\qquad\leq\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi(a)}\|_{TV}\\
&\qquad\quad+\sum_{j=1}^{k}\left(\phi\left(a+\frac{j}{T}\right)-\phi(a)\right)\\
&\qquad\leq\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi(a)}\|_{TV}+\sum_{j=1}^{k}L\left|a+\frac{j}{T}-a\right|\\
&\qquad=\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi(a)}\|_{TV}+\frac{L}{T}\sum_{j=1}^{k}j\\
&\qquad=\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi(a)}\|_{TV}+L\frac{k(k+1)}{2T}\\
&\qquad\leq\|\pi_{\phi\left(a+\frac{k}{T}\right)}-\pi_{\phi(a)}\|_{TV}+L\frac{(k+1)^2}{2T}.
\end{aligned}$$

3.4.4 Proof of Proposition 8

We begin with the creation of an orthonormal basis of eigenvectors associated with $(\mathbb{I}-\mathbf{P}_{\phi(t)})(\mathbb{I}-\mathbf{P}_{\phi(t)})^T$ by a singular value decomposition similar to the process we mentioned in Proposition 3, where $t\in[a,b]$.
Here we let $\sigma_{1,t}\geq\cdots\geq\sigma_{n-1,t}=\sigma_t$ be the positive singular values of $(\mathbb{I}-\mathbf{P}_{\phi(t)})$ with respect to the Euclidean inner product. This implies that there exists an orthonormal basis $\{\mathbf{v_{1,t}},\cdots,\mathbf{v_{n,t}}\}$ such that $\mathbf{v_{j,t}}(\mathbb{I}-\mathbf{P}_{\phi(t)})(\mathbb{I}-\mathbf{P}_{\phi(t)})^T=\sigma_{j,t}^2\mathbf{v_{j,t}}$

for $1 \leq j \leq n-1$ and $\mathbf{v_{n,t}}(\mathbb{I} - \mathbf{P}_{\phi(t)})(\mathbb{I} - \mathbf{P}_{\phi(t)})^T = \mathbf{0}$.

Here $\mathbf{v_{n,t}} = \pi_{\phi(t)}/\|\pi_{\phi(t)}\|_2$.

To show continuity of $\pi_{\phi(s)}$ on $[a,b]$ let $\epsilon > 0$ and first notice that for any $s^*, s \in [a,b]$, $(\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)}) = (\phi(s^*) - \phi(s))\, \pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})$.

Using the Euclidean norm, we see that if $\mathbf{P_0} \neq \mathbf{P_1}$ and $s^* \neq s$, then

$$\frac{\|(\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)})\|_2}{\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2} = (\phi(s^*) - \phi(s))\, \frac{\|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2}.$$

Throughout this proof we will use $< \cdot, \cdot >$ as the Euclidean inner product.

For $1 \leq j \leq n$ let $c_{j,s} =< \pi_{\phi(s^*)} - \pi_{\phi(s)}, \mathbf{v_{j,s}} >$. Then we see that $\pi_{\phi(s^*)} - \pi_{\phi(s)} = \sum_{j=1}^{n} c_{j,s}\mathbf{v_{j,s}}$.

We have that

$$
\begin{aligned}
&\frac{\|(\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)})\|_2^2}{\|\pi_{\phi(s^s)} - \pi_{\phi(s)}\|_2^2} \\
&\quad = \frac{< (\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)}), (\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)}) >}{< \pi_{\phi(s^*)} - \pi_{\phi(s)}, \pi_{\phi(s^*)} - \pi_{\phi(s)} >} \\
&\quad = \frac{< \pi_{\phi(s^*)} - \pi_{\phi(s)}, (\pi_{\phi(s^*)} - \pi_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)})(\mathbb{I} - \mathbf{P}_{\phi(s)})^T >}{< \pi_{\phi(s^*)} - \pi_{\phi(s)}, \pi_{\phi(s^*)} - \pi_{\phi(s)} >} \\
&\quad = \frac{< \sum_{j=1}^{n} c_{j,s}\mathbf{v_{j,s}}, \sum_{j=1}^{n-1} \sigma_{j,s}^2 c_{j,s}\mathbf{v_{j,s}} >}{< \sum_{j=1}^{n} c_{j,s}\mathbf{v_{j,s}}, \sum_{j=1}^{n} c_{j,s}\mathbf{v_{j,s}} >} \\
&\quad = \frac{\sum_{j=1}^{n-1} \sigma_{j,s}^2 c_{j,s}^2}{\sum_{j=1}^{n} c_{j,s}^2} \\
&\quad \geq \sigma_{n-1,s}^2 \frac{\sum_{j=1}^{n-1} c_{j,s}^2}{\sum_{j=1}^{n} c_{j,s}^2} \\
&\quad = \sigma_{n-1,s}^2 \left(1 - \frac{c_{n,s}^2}{\sum_{j=1}^{n} c_{j,s}^2}\right) \\
&\quad = \sigma_{n-1,s}^2 \left(1 - \left(\frac{< \pi_{\phi(s^*)} - \pi_{\phi(s)}, \mathbf{v_{n,s}} >}{\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2}\right)^2\right).
\end{aligned}
$$

If we let $\mathbf{w}(s^*, s) = (\pi_{\phi(s^*)} - \pi_{\phi(s)})/\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2$ then we see that

$$
\sigma_{n-1,s}^2 \left(1 - (< \mathbf{w}(s^*, s), \mathbf{v_{n,s}} >)^2\right) \leq (\phi(s^*) - \phi(s))^* \frac{\|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2^2}{\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2^2}.
$$

Because $\mathbf{w}(s^*, s)$ and $\mathbf{v_{n,s}}$ are unit vectors, we can use the fact that

$$
\|\mathbf{w}(s^*, s)\|_2^2 - 2 < \mathbf{w}(s^*, s), \mathbf{v_{n,s}} > + \|\mathbf{v_{n,s}}\|_2^2 = \|\mathbf{w}(s^*, s) - \mathbf{v_{n,s}}\|_2^2
$$

to show that

$$
1 - < \mathbf{w}(s^*, s), \mathbf{v_{n,s}} > = \frac{1}{2}\|\mathbf{w}(s^*, s) - \mathbf{v_{n,s}}\|_2^2
$$

and we can use the fact that

$$
\|\mathbf{w}(s^*, s)\|_2^2 + 2 < \mathbf{w}(s^*, s), \mathbf{v_{n,s}} > + \|\mathbf{v_{n,s}}\|_2^2 = \|\mathbf{w}(s^*, s) + \mathbf{v_{n,s}}\|_2^2
$$

to show that

$$1+ < \mathbf{w}(s^*, s), \mathbf{v_{n,s}} >= \frac{1}{2}\|\mathbf{w}(s^*, s) + \mathbf{v_{n,s}}\|_2^2.$$

From this we see that $1 - (< \mathbf{w}(s^*, s), \mathbf{v_{n,s}} >)^2 = \|\mathbf{w}(s^*, s) - \mathbf{v_{n,s}}\|_2^2 \cdot \|\mathbf{w}(s^*, s) + \mathbf{v_{n,s}}\|_2^2/4$.

Plugging this into our previous equation, we can see that

$$\frac{\sigma_{n-1,s}^2}{4}\|\mathbf{w}(s^*, s)-\mathbf{v_{n,s}}\|_2^2 \cdot \|\mathbf{w}(s^*, s)+\mathbf{v_{n,s}}\|_2^2 \leq (\phi(s^*) - \phi(s))^2 \frac{\|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2^2}{\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2^2}.$$

After performing some basic algebra we see that

$$\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2 \leq \frac{2\big|\phi(s^*) - \phi(s)\big| \cdot \|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1,s}\|\mathbf{w}(s^*, s) - \mathbf{v_{n,s}}\|_2 \cdot \|\mathbf{w}(s^*, s) + \mathbf{v_{n,s}}\|_2}.$$

Notice that $< \mathbf{w}(s^*, s), \mathbf{1} > /\sqrt{n} = 0$ and $< \mathbf{v_{n,s}}, \mathbf{1} > /\sqrt{n} = 1/\left(\sqrt{n}\|\pi_{\phi(s)}\|_2\right)$ for all $s^*, s \in [a, b]$. Because these are the scalar components of the projections of $\mathbf{w}(s^*, s)$ and $\mathbf{v_{n,s}}$ onto $\mathbf{1}$ respectively, we see that the minimum possible value for $\|\mathbf{w}(s^*, s) - \mathbf{v_{n,s}}\|_2$ and $\|\mathbf{w}(s^*, s) + \mathbf{v_{n,s}}\|_2$ is at least $1/\left(\sqrt{n}\|\pi_{\phi(s)}\|_2\right)$.

Letting $\sigma = \inf_{t\in[a,b]}\{\sigma_{n-1,s}\}$, we now have that

$$\begin{aligned}\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_2 &\leq \frac{2n\big|\phi(s^*) - \phi(s)\big| \cdot \|\pi_{\phi(s)}\|_2^2 \cdot \|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1,s}} \\ &\leq \frac{2n\big|\phi(s^*) - \phi(s)\big| \cdot \|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma_{n-1,s}} \\ &= \frac{2n\big|\phi(s^*) - \phi(s)\big| \cdot \|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_2}{\sigma}.\end{aligned}$$

Again for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x}$ and $\mathbf{y}$ are probability measures, we see that

$$\frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_{TV} \leq \frac{\sqrt{n}}{2}\|\mathbf{x} - \mathbf{y}\|_2.$$

This will imply that

$$\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} \leq \frac{2n^{3/2}\big|\phi(s^*) - \phi(s)\big| \cdot \|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_{TV}}{\sigma}.$$

Because $\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0}) = \pi_{\phi(s^*)}\mathbf{P_1} - \pi_{\phi(s^*)}\mathbf{P_0}$ is the difference of two probability distributions, we see that $\|\cdot\|_{TV} = \frac{1}{2}\|\cdot\|_1$ where $\|\cdot\|_1$ is the l_1-norm. This implies that

$$\begin{aligned}
\|\pi_{\phi(s^*)}(\mathbf{P_1} - \mathbf{P_0})\|_{TV} &= \frac{1}{2}\|\pi_{\phi(s^*)}\mathbf{P_1} - \pi_{\phi(s^*)}\mathbf{P_0}\|_1 \\
&\leq \frac{1}{2}\left(\|\pi_{\phi(s^*)}\mathbf{P_1}\|_1 + \|\pi_{\phi(s^*)}\mathbf{P_0}\|_1\right) \\
&\leq 1.
\end{aligned}$$

This shows that

$$\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} \leq \frac{2n^{3/2}\big|\phi(s^*) - \phi(s)\big|}{\sigma}.$$

Remember that ϕ is a bi-Lipschitz continuous function with Lipschitz constant L, we see that

$$\begin{aligned}
\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} &\leq \frac{2n^{3/2}\big|\phi\left(s^*\right) - \phi\left(s\right)\big|}{\sigma} \\
&\leq \frac{2n^{3/2}L\big|s^* - s\big|}{\sigma}.
\end{aligned}$$

Clearly if $\epsilon > 0$, then

$$\big|s^* - s\big| \leq \delta = \frac{\epsilon\sigma}{2Ln^{3/2}}$$

implies $\|\pi_{\phi(s^*)} - \pi_{\phi(s)}\|_{TV} \leq \epsilon$.

This shows that $\pi_{\phi(s)}$ is continuous on $[a, b]$.

3.4.5 Proof of Theorem 11

We first provide a sketch of the proof followed by the technical details. Our proof is based on the results of Theorem 10, Proposition 7, and Corollary 3. Specifically, we divide our proof into three cases. In the first case, we will show how to select T and δ in order to satisfy the two conditions in Theorem 10, namely:

$$T \geq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon(\delta^* - a)},$$

and

$$\delta^* \leq a + k/T \leq b.$$

Therefore, by Theorem 10, we have

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon.$$

However, the selected T is not large enough to bound this for all $1 \leq k \leq k_{max}$ since this only holds for k such that

$$\delta^* \leq a + k/T \leq b.$$

In the second case, we will use the results of Proposition 7 and Corollary 3 to show that for the same selected T and δ,

$$\|\pi_{\phi(a)} \mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon,$$

even in the case when

$$a \leq a + k/T \leq \delta^* < b.$$

In the final case, we will do something to fix the remaining problem. Therefore, we conclude that the selected T is a sufficient condition to remain in our ϵ-corridor for $1 \leq k \leq T$.

We now proceed with the details of the proof, starting with the first case. Let $\epsilon > 0$. Let L be the Lipschitz constant associated with the strictly increasing, bi-Lipschitz function ϕ. For this fixed ϵ, we choose T be an integer such that

$$\begin{aligned} T &\geq \frac{16L^5 t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{8L^3 t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{L}{\epsilon} \\ &= \left(\frac{4L^2\sqrt{L} t_{mix}^2(\epsilon/2)}{\epsilon\sqrt{\epsilon}} + \sqrt{\frac{L}{\epsilon}} \right)^2 . \end{aligned}$$

This implies

$$\sqrt{T} \geq \frac{4L^2\sqrt{L} t_{mix}^2(\epsilon/2)}{\epsilon\sqrt{\epsilon}} + \sqrt{\frac{L}{\epsilon}}.$$

Multiplying either side by $\sqrt{\epsilon/L)}$ and subtracting 1 from either side we obtain the following after switching the direction of the inequality

$$\begin{aligned} \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon} &\leq \sqrt{\frac{\epsilon T}{L}} - 1 \\ &\leq T\left(\sqrt{\frac{\epsilon}{LT}} - \frac{1}{T}\right) \end{aligned}$$

Now, letting

$$\delta^* - a = \sqrt{\frac{\epsilon}{LT}} - \frac{1}{T}$$

and dividing both sides by $\delta^* - a$, we clearly have

$$T \geq \frac{4L^2 t_{mix}^2(\epsilon/2)}{\epsilon(\delta^* - a)}.$$

Next, let k be an integer such that

$$a < \delta^* = a + \sqrt{\frac{\epsilon}{LT}} - \frac{1}{T} \leq a + \frac{k}{T} \leq b.$$

Then by Theorem 10, we conclude that

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon.$$

Now in the second (complementary) case, i.e., when $a + k/T \leq \delta^* < b$, we will show that for the same selected $\delta^* = \sqrt{\frac{\epsilon}{LT}} - \frac{1}{T}$, and T, it is still true that:

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \leq \epsilon,$$

Let k be an integer such that

$$a \leq a + \frac{k}{T} \leq a + \sqrt{\frac{\epsilon}{LT}} - \frac{1}{T} = \delta^*.$$

Then,

$$\frac{k+1}{T} \leq \sqrt{\frac{\epsilon}{LT}}.$$

Using Proposition 7, we have

$$\begin{aligned}
&\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)} \cdots \mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} \\
&\qquad \leq \|\pi_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi(a)}\|_{TV} + \frac{L(k+1)^2}{2T} \\
&\qquad = \|\pi_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi(a)}\|_{TV} + \frac{LT}{2}\left(\frac{k+1}{T}\right)^2 \\
&\qquad \leq \|\pi_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi(a)}\|_{TV} + \frac{LT}{2}\left(\sqrt{\frac{\epsilon}{LT}}\right)^2 \\
&\qquad = \|\pi_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi(a)}\|_{TV} + \frac{\epsilon}{2}.
\end{aligned}$$

Next, from Corollary 3, as long as $\epsilon < 1/\sqrt{n}$ and $\sqrt{\frac{\epsilon}{LT}} - \frac{1}{T} \leq \frac{\epsilon(1-\sqrt{n}\epsilon)}{4Ln^{3/2}t_{mix}(\epsilon/2)}$, we have

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(\mathbf{a}+\frac{\mathbf{1}}{\mathbf{T}}\right)}\cdots\mathbf{P}_{\phi\left(\mathbf{a}+\frac{\mathbf{k}}{\mathbf{T}}\right)} - \pi_{\phi\left(\mathbf{a}+\frac{\mathbf{k}}{\mathbf{T}}\right)}\|_{TV} \leq \epsilon$$

for

$$a \leq a + \frac{k}{T} \leq a + \sqrt{\frac{\epsilon}{LT}} - \frac{1}{T}.$$

It should be clear that as $\epsilon \to 0$,

$$\sqrt{\frac{\epsilon}{LT}} - \frac{1}{T} \leq \frac{\epsilon(1-\sqrt{n}\epsilon)}{4Ln^{3/2}t_{mix}(\epsilon/2)}$$

when

$$T \geq \frac{16L^5t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{8L^3t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{L}{\epsilon}.$$

This tells us that as $\epsilon \to 0$ and

$$T \geq \frac{16L^5t_{mix}^4(\epsilon/2)}{\epsilon^3} + \frac{8L^3t_{mix}^2(\epsilon/2)}{\epsilon^2} + \frac{L}{\epsilon}$$

we have

$$\|\pi_{\phi(a)}\mathbf{P}_{\phi\left(a+\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi\left(a+\frac{k}{T}\right)} - \pi_{\phi\left(a+\frac{k}{T}\right)}\|_{TV} < \epsilon$$

for $1 \leq k \leq k_{max}$.

3.4.6 Proof of Proposition 9

Let $x, y \in [a, c]$ with $x < y$.

First suppose that $x, y \in [a, b]$.

We see then that $\left|\phi(x) - \phi(y)\right| = \left|\phi_1(x) - \phi_1(y)\right|$.

This would imply that

$$\frac{1}{\max\{L_1, L_2\}}\big|x-y\big| \le \frac{1}{L_1}\big|x-y\big| \le \big|\phi(x)-\phi(y)\big| \le L_1\big|x-y\big| \le \max\{L_1, L_2\}\big|x-y\big|.$$

Now suppose that $x, y \in [b, c]$.

We see then that $\big|\phi(x) - \phi(y)\big| = \big|\phi_2(x) - \phi_2(y)\big|$.

This would imply that

$$\frac{1}{\max\{L_1, L_2\}}\big|x-y\big| \le \frac{1}{L_2}\big|x-y\big| \le \big|\phi(x)-\phi(y)\big| \le L_2\big|x-y\big| \le \max\{L_1, L_2\}\big|x-y\big|.$$

Finally suppose that $x \in [a, b]$ and $y \in [b, c]$.

We see then that

$$\begin{aligned}\big|\phi(x) - \phi(y)\big| &= \big|\phi_1(x) - \phi_2(y)\big| \\ &= \big|\phi_1(x) - \phi_1(b) + \phi_2(b) - \phi_2(y)\big|.\end{aligned}$$

Notice that if both ϕ_1 and ϕ_2 are strictly increasing, then

$$\begin{aligned}\big|\phi_1(x) - \phi_1(b) + \phi_2(b) - \phi_2(y)\big| &= \phi_1(b) - \phi_1(x) + \phi_2(y) - \phi_2(b) \\ &= \big|\phi_1(x) - \phi_1(b)\big| + \big|\phi_2(b) - \phi_2(y)\big|.\end{aligned}$$

Similarly, if both ϕ_1 and ϕ_2 are strictly decreasing, then

$$\begin{aligned}\big|\phi_1(x) - \phi_1(b) + \phi_2(b) - \phi_2(y)\big| &= \phi_1(x) - \phi_1(b) + \phi_2(b) - \phi_2(y) \\ &= \big|\phi_1(x) - \phi_1(b)\big| + \big|\phi_2(b) - \phi_2(y)\big|.\end{aligned}$$

In either case we see that $\left|\phi(x) - \phi(y)\right| = \left|\phi_1(x) - \phi_1(b)\right| + \left|\phi_2(b) - \phi_2(y)\right|$.

We see then that both

$$\frac{1}{\max\{L_1, L_2\}} \left(\left|x - b\right| + \left|b - y\right|\right) \leq \frac{1}{L_1}\left|x - b\right| + \frac{1}{L_2}\left|b - y\right| \leq \left|\phi(x) - \phi(y)\right|$$

and

$$\left|\phi(x) - \phi(y)\right| \leq L_1\left|x - b\right| + L_2\left|b - y\right| \leq \max\{L_1, L_2\} \left(\left|x - b\right| + \left|b - y\right|\right).$$

Because $x < b$ and $b < y$, we see that

$$\begin{aligned} \left|x - b\right| + \left|b - y\right| &= b - x + y - b \\ &= y - x \\ &= \left|x - y\right|. \end{aligned}$$

This would imply that

$$\frac{1}{\max\{L_1, L_2\}}\left|x-y\right| \leq \frac{1}{L_2}\left|x-y\right| \leq \left|\phi(x)-\phi(y)\right| \leq L_2\left|x-y\right| \leq \max\{L_1, L_2\}\left|x-y\right|.$$

We see that regardless of where x and y are in the interval $[a, c]$, ϕ has the right inequality to be a bi-Lipschitz continuous function with Lipschitz constant $\max\{L_1, L_2\}$ and because depending on the maximum value, there are $x, y \in [a, c]$ to where this inequality is tight, ϕ will be bi-Lipschitz continuous with Lipschitz constant $\max\{L_1, L_2\}$.

3.4.7 Proof of Theorem 12

For a partition $\mathcal{P}_M$ and positive integer T large enough, there exists positive integers $k_{1,T}, \cdots, k_{2M-1,T}$ so that $k_{2M-1,T} = T$ and

$$\frac{k_{i,T}}{T} \leq a_i \leq \frac{k_{i,T}+1}{T}$$

for $1 \leq i \leq 2M-2$.

For this partition $\mathcal{P}_M$ and value of T, we see then that for $1 \leq l \leq k_{1,T}$,

$$\pi_{\mathbf{0}} \mathbf{P}_{\Phi(\frac{1}{T})} \cdots \mathbf{P}_{\Phi(\frac{l}{T})} - \pi_{\Phi(\frac{l}{T})} = \pi_{\mathbf{0}} \mathbf{P}_{\phi_1(\frac{1}{T})} \cdots \mathbf{P}_{\phi_1(\frac{l}{T})} - \pi_{\phi_1(\frac{l}{T})}$$

and for $1 \leq i \leq 2M-2$ and $k_{i,T}+1 \leq l \leq k_{i+1,T}$,

$$\begin{aligned}
&\pi_{\mathbf{0}} \mathbf{P}_{\Phi(\frac{1}{T})} \cdots \mathbf{P}_{\Phi(\frac{l}{T})} - \pi_{\Phi(\frac{l}{T})} \\
&\quad = \pi_{\mathbf{0}} \mathbf{P}_{\phi_1(\frac{1}{T})} \cdots \mathbf{P}_{\phi_1\left(\frac{k_{1,T}}{T}\right)} - \pi_{\phi_1\left(\frac{k_{1,T}}{T}\right)} \\
&\quad + \sum_{j=1}^{i-1} \left(\pi_{\phi_{j+1}\left(\frac{k_{j,T}+1}{T}\right)} \mathbf{P}_{\phi_{j+1}\left(\frac{k_{j,T}+2}{T}\right)} \cdots \mathbf{P}_{\phi_{j+1}\left(\frac{k_{j+1,T}}{T}\right)} - \pi_{\phi_{j+1}\left(\frac{k_{j+1,T}}{T}\right)} \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \prod_{j_1=k_{j_1,T}+1}^{l} \mathbf{P}_{\Phi(\frac{j_1}{T})} \\
&\quad + \pi_{\phi_{i+1}\left(\frac{k_{i,T}+1}{T}\right)} \mathbf{P}_{\phi_{i+1}\left(\frac{k_{i,T}+2}{T}\right)} \cdots \mathbf{P}_{\phi_{i+1}(\frac{l}{T})} - \pi_{\phi_{i+1}(\frac{l}{T})} \\
&\quad + \sum_{j=1}^{i} \left(\pi_{\Phi\left(\frac{k_{j,T}}{T}\right)} - \pi_{\Phi\left(\frac{k_{j,T}+1}{T}\right)} \right) \prod_{j_1=k_{j,T}+1}^{l} \mathbf{P}_{\Phi(\frac{j_1}{T})}.
\end{aligned}$$

Taking the total variation norm to either side and using the triangle inequality and the fact that $\|(\mu-\nu)\mathbf{P}\|_{TV} \leq \|\mu-\nu\|_{TV}$ for any probability distributions μ and ν and stochastic matrix $\mathbf{P}$, we see that for $1 \leq l \leq k_{1,T}$,

$$\|\pi_{\mathbf{0}} \mathbf{P}_{\Phi(\frac{1}{T})} \cdots \mathbf{P}_{\Phi(\frac{l}{T})} - \pi_{\Phi(\frac{l}{T})}\|_{TV} = \|\pi_{\mathbf{0}} \mathbf{P}_{\phi_1(\frac{1}{T})} \cdots \mathbf{P}_{\phi_1(\frac{l}{T})} - \pi_{\phi_1(\frac{l}{T})}\|_{TV}$$

and for $1 \le i \le 2M-2$ and $k_{i,T}+1 \le l \le k_{i+1,T}$,

$$\begin{aligned}
&\|\pi_{\mathbf{0}}\mathbf{P}_{\Phi\left(\frac{1}{T}\right)}\cdots\mathbf{P}_{\Phi\left(\frac{l}{T}\right)} - \pi_{\Phi\left(\frac{l}{T}\right)}\|_{TV} \\
&\quad = \|\pi_{\mathbf{0}}\mathbf{P}_{\phi_1\left(\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi_1\left(\frac{k_{1,T}}{T}\right)} - \pi_{\phi_1\left(\frac{k_{1,T}}{T}\right)}\|_{TV} \\
&\quad + \sum_{j=1}^{i-1}\|\pi_{\phi_{j+1}\left(\frac{k_{j,T}+1}{T}\right)}\mathbf{P}_{\phi_{j+1}\left(\frac{k_{j,T}+2}{T}\right)}\cdots\mathbf{P}_{\phi_{j+1}\left(\frac{k_{j+1,T}}{T}\right)} - \pi_{\phi_{j+1}\left(\frac{k_{j+1,T}}{T}\right)}\|_{TV} \\
&\quad + \|\pi_{\phi_{i+1}\left(\frac{k_{i,T}+1}{T}\right)}\mathbf{P}_{\phi_{i+1}\left(\frac{k_{i,T}+2}{T}\right)}\cdots\mathbf{P}_{\phi_{i+1}\left(\frac{l}{T}\right)} - \pi_{\phi_{i+1}\left(\frac{l}{T}\right)}\|_{TV} \\
&\quad + \sum_{j=1}^{i}\|\pi_{\phi_j\left(\frac{k_{j,T}}{T}\right)} - \pi_{\phi_j(a_j)}\|_{TV} + \|\pi_{\phi_{j+1}(a_j)} - \pi_{\phi_{j+1}\left(\frac{k_{j,T}+1}{T}\right)}\|_{TV}.
\end{aligned}$$

Because ϕ_1 is strictly increasing on $[0, k_{1,T}/T]$ and ϕ_i is either strictly increasing or strictly decreasing on $[(k_{i,T}+1)/T, k_{i+1,T}/T]$ for $1 \le i \le 2M-2$ we can use Corollary 5 to say that as $\epsilon \to 0$, if

$$\begin{aligned}
T \ge\ & \frac{16L^5(4M-2)^3 t_{mix}^4(\epsilon/(8M-4))}{\epsilon^3} \\
&+ \frac{8L^3(4M-2)^2 t_{mix}^2(\epsilon/(8M-4))}{\epsilon^2} + \frac{L(4M-2)}{\epsilon}
\end{aligned}$$

then we have for $1 \le l \le k_{1,T}$,

$$\|\pi_0\mathbf{P}_{\phi_1\left(\frac{1}{T}\right)}\cdots\mathbf{P}_{\phi_1\left(\frac{l}{T}\right)} - \pi_{\phi_1\left(\frac{l}{T}\right)}\|_{TV} < \frac{\epsilon}{4M-2}$$

and for $1 \le i \le 2M-2$ and $k_{i,T}+1 \le l \le k_{i+1,T}$,

$$\|\pi_{\phi_{i+1}\left(\frac{k_{i,T}+1}{T}\right)}\mathbf{P}_{\phi_{i+1}\left(\frac{k_{i,T}+2}{T}\right)}\cdots\mathbf{P}_{\phi_{i+1}\left(\frac{l}{T}\right)} - \pi_{\phi_{i+1}\left(\frac{l}{T}\right)}\|_{TV} < \frac{\epsilon}{4M-2}.$$

This implies that regardless of the subinterval containing k/T, we have that

$$\begin{aligned}&\|\pi_{\mathbf{0}}\mathbf{P}_{\Phi\left(\frac{1}{T}\right)}\cdots\mathbf{P}_{\Phi\left(\frac{k}{T}\right)}-\pi_{\Phi\left(\frac{k}{T}\right)}\|_{TV}\\&\quad\leq\sum_{j=1}^{2M-1}\|\pi_{\phi_j\left(\frac{k_{j,T}}{T}\right)}-\pi_{\phi_j(a_j)}\|_{TV}+\|\pi_{\phi_{j+1}(a_j)}-\pi_{\phi_{j+1}\left(\frac{k_{j,T}+1}{T}\right)}\|_{TV}\\&\quad+\epsilon/2.\end{aligned}$$

We finally use Corollary 3 to say that on the inteval I_i, if

$$\delta^*=\frac{\epsilon((8M-4)-\sqrt{n}\epsilon)}{4L_in^{3/2}t_{mix}(\epsilon/(8M-4))}$$

we have for all $\left|s^*-s\right|\leq\delta^*$, $\|\pi_{\phi_i(s^*)}-\pi_{\phi_i(s)}\|_{TV}\leq\epsilon/(8M-4)$.

This would imply that as $\epsilon\to 0$, if

$$\begin{aligned}T\geq&\ \frac{16L^5(4M-2)^3t_{mix}^4(\epsilon/(8M-4))}{\epsilon^3}\\&+\frac{8L^3(4M-2)^2t_{mix}^2(\epsilon/(8M-4))}{\epsilon^2}+\frac{L(4M-2)}{\epsilon}\end{aligned}$$

then

$$\|\pi_{\mathbf{0}}\mathbf{P}_{\Phi\left(\frac{1}{T}\right)}\cdots\mathbf{P}_{\Phi\left(\frac{k}{T}\right)}-\pi_{\Phi\left(\frac{k}{T}\right)}\|_{TV}\leq\epsilon.$$

This implies that

$$t_{sad}(\mathbf{P_0},\mathbf{P_1},\epsilon)=O\left(\frac{t_{mix}^4(\epsilon/2)}{\epsilon^3}\right).$$

Chapter 4

The Stable Adiabatic Time Versus The Spectral Gap For Discrete Markov Chains

Before we derived the results in Chapter 3 we had a different goal in mind. Our goal was to bound the stable adiabatic time with a spectral gap measurement, rather than a mixing time measurement. The beginning of this chapter is dedicated to a detailed explanation of the motivation of this goal. Section 4.1 will discuss one form of the Quantum Adiabatic Theorem, which is related our research. Our goal was to asymptotically bound the stable adiabatic time as a function of ϵ by an inverse power of the smallest spectral gap over a linear adiabatic evolution multiplied by an inverse power of ϵ. The latter sections provide bounds for the stable adiabatic time in different scenarios, however, we see that these results will not be as general and concise as the results in Chapter 3. In these sections we see some partial evidence that the stable adiabatic time is asymptotically bounded by a constant multiple of an inverse cube of the smallest spectral gap multiplied by the multiplicative inverse of ϵ.

4.1 The Quantum Adiabatic Theorem

Adiabatic transitions are probably best known in the context of quantum mechanics. There is a well-known Quantum Adiabatic Theorem documented in many sources. In this section we will outline one version of the Quantum Adiabatic Theorem as it was given in [3]. The result of this theorem motivated us to pursue an analogue for Markov processes.

For $s \in [0,1]$ let $\mathbf{H}(s)$ be a Hamiltonian, also called an energy function, dependent on the parameter s. Although the authors of [3] do not require the Hamiltonian to have a finite number of states, we will only focus on finite state Hamiltonian operators. Hamiltonian operators over n states are $l^2(\mathbb{C}^n)$-norm preserving, i.e the energy functions are Hermitian. We call $\mathbf{H}(0)$ and $\mathbf{H}(1)$ the initial and final Hamiltonians respectively. We will use the notation $\|\mathbf{H}\|$ to denote $\max_{s\in[0,1]} \|\mathbf{H}(s)\|$ where we will denote $\|\cdot\|$ as the usual operator norm.

The quantum adiabatic results often concern one eigenstate of the energy function, the ground state, because the proofs follow from the results on one eigenstate. Let $\mathbf{\Phi}(s)$ be the ground state of $\mathbf{H}(s)$ with eigenvalue $\gamma(s)$. Hamiltonians are often used like the generators of continuous-time Markov processes (we will see more about this in Chapters 5, 6 and 8), so for a given $T > 0$ when we say that we apply the adiabatic evolution given by $\mathbf{H}$ and $\mathbf{\Phi}$ for time T we mean that we initialize a system in the state $\mathbf{\Phi}(0)$ and then apply the continuously varying Hamiltonian $\mathbf{H}(t/T)$ for time $t \in [0,T]$.

Given $\epsilon > 0$ the Quantum Adiabatic Theorem informally says that if we assume that the change in the Hamiltonian happens slowly enough, by selecting a large enough value of T, then when we apply the adiabatic evolution given by $\mathbf{H}$ and $\mathbf{\Phi}$ for time T we will be in an ϵ-ball of $\mathbf{\Phi}(1)$. The time T is called the quantum adiabatic time. The Quantum Adiabatic Theorem addresses how large this quantum adiabatic time must be to guarantee the above result. We cite the theorem from [3].

Theorem 13 *For $s \in [0,1]$ let $\mathbf{H}(s)$ be a time dependent Hamiltonian, let $\mathbf{\Phi}(s)$ be the ground state and let $\gamma(s)$ be the eigenvalue associated with the ground state. Assume that for any $s \in [0,1]$ all other eigenvalues of $\mathbf{H}(s)$ are either smaller than $\gamma(s) - \Delta$ or larger than $\gamma(s) + \Delta$ (i.e. there is a spectral gap of Δ around $\gamma(s)$). Consider the adiabatic evolution given by $\mathbf{H}$ and $\mathbf{\Phi}$ for time T. Then, the following condition is enough to guarantee that the final state is at distance at most ϵ from $\mathbf{\Phi}(1)$:*

$$T \geq \frac{10^5}{\epsilon^2} \max \left\{ \frac{\|\mathbf{H}^{'}\|^3}{\Delta^4}, \frac{\|\mathbf{H}^{'}\|\|\mathbf{H}^{''}\|}{\Delta^3} \right\} . \tag{4.1}$$

In our work we try make an analogue to the quantum process described above, only now it is for time inhomogeneous Markov processes. The definition of the adiabatic time will be the appropriate analogue. Here the ground state in quantum mechanics corresponds to the stationary state of a Markov process. The Hamiltonian operator described above, in the context of quantum mechanics, most accurately corresponds to the generator matrix of a time-inhomogeneous, continuous-time Markov process.

Notice that the quantum adiabatic time is bounded by the inverse square of ϵ multiplied by an inverse power of the smallest spectral gap over the adiabatic evolution given by $\mathbf{H}$ and $\mathbf{\Phi}$ for time T, however, there are some operator norm measurements that we are not necessarily able to write in terms of ϵ or the smallest spectral gap. The bound on the quantum adiabatic time led us to believe that we could find a similar bound on the stable adiabatic time for a $l^1(\mathbb{R}^n)$-norm preserving Markov processes. The following sections outline our attempts to find this bound.

4.2 Two-State Markov Chains

This section establishes a relationship between the stable adiabatic time and the smallest spectral gap over the entire transition for two-state, discrete-time Markov chains under a linear adiabatic evolution, see Definition 12 in Section 2.1.

To begin, I adopt a new notation for our probability transition matrices that agrees with the notation for birth-death processes. We will return to birth-death processes in Section 4.6. We define constants $p_1^{(0)}, p_1^{(1)}, q_2^{(0)}, q_2^{(1)} \in [0,1]$. The general initial and final two-state probability transition matrices are written as follows:

$$\mathbf{P_0} = \begin{bmatrix} 1-p_1^{(0)} & p_1^{(0)} \\ q_2^{(0)} & 1-q_2^{(0)} \end{bmatrix} \quad \text{and} \quad \mathbf{P_1} = \begin{bmatrix} 1-p_1^{(1)} & p_1^{(1)} \\ q_2^{(1)} & 1-q_2^{(1)} \end{bmatrix}$$

We can succinctly write a linear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ if we define $p_1^{(t)} = (1-t)p_1^{(0)} + tp_1^{(1)}$ and $q_2^{(t)} = (1-t)q_2^{(0)} + tq_2^{(1)}$. We see then that

$$\mathbf{P_t} = (1-t)\mathbf{P_0} + t\mathbf{P_1} = \begin{bmatrix} 1-p_1^{(t)} & p_1^{(t)} \\ q_2^{(t)} & 1-q_2^{(t)} \end{bmatrix}$$

I first want to explore the spectral structure of $\mathbf{P_t}$. The following Proposition finds the eigenvalues of $\mathbf{P_t}$ and the proof of this proposition is given in Section 4.3.

Proposition 10 *The eigenvalues of* $\mathbf{P_t}$ *are* 1 *and* $1-\left(p_1^{(t)}+q_2^{(t)}\right)$.

If $\mathbf{P_t}$ is irreducible and aperiodic, then $\left|1-\left(p_1^{(t)}+q_2^{(t)}\right)\right| \neq 1$. In the case that $\mathbf{P_t}$ is irreducible and aperiodic we use Definition 10 to see that the spectral gap of $\mathbf{P_t}$ is $\Delta = 1-\left|1-\left(p_1^{(t)}+q_2^{(t)}\right)\right|$. In this chapter, however, we reserve the symbol Δ for the smallest spectral gap of a linear adiabatic evolution.

Definition 23 *Let* $\{\mathbf{P_t}\}_{t\in[0,1]}$ *be a linear evolution between two stochastic matrices* $\mathbf{P_0}$ *and* $\mathbf{P_1}$ *that are both irreducible and aperiodic. Letting* $\lambda_1(t)=1$ *be the largest eigenvalue in modulus of* $\mathbf{P_t}$ *and* $\lambda_2(t)$ *be the second largest eigenvalue in modulus of* $\mathbf{P_t}$ *we define the* <u>*smallest spectral gap*</u> *of the adiabatic transition to be*

$$\Delta = \inf_{0\le t\le 1}\{1-\left|\lambda_2(t)\right|\}. \tag{4.2}$$

Clearly, in this section $\Delta = \inf_{0\le t\le 1}\{1-\left|1-\left(p_1^{(t)}+q_2^{(t)}\right)\right|\}$. We also denote $\Delta_t = p_1^{(t)}+q_2^{(t)}$ to ease notation throughout this section.

The following two propositions describe the eigenvectors of $\mathbf{P_t}$. The proofs of these propositions can be found in Section 4.3.

Proposition 11 *For* $t\in[0,1]$ *the stationary distribution of* $\mathbf{P_t}$ *is*

$$\pi_t = \begin{bmatrix} \frac{q_2^{(t)}}{\Delta_t} & \frac{p_1^{(t)}}{\Delta_t} \end{bmatrix}. \tag{4.3}$$

Proposition 12 *For $t \in [0,1]$*

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \mathbf{P_t} = (1-\Delta_t) \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{4.4}$$

.

Using Proposition 11 we calculate the difference between fractionally consecutive stationary distributions. The work behind this calculation is given in Section 4.3.

Proposition 13 *Given $T \in \mathbb{N}$ we have that*

$$\pi_{\frac{j-1}{T}} - \pi_{\frac{j}{T}} = \frac{p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}}{T \Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \tag{4.5}$$

for $j = 1, \cdots, T$.

Proposition 13 tells us that the difference of fractionally consecutive stationary distributions is an eigenvector of $\mathbf{P_t}$ for all $t \in [0,1]$. This idea helped us measure $\|\pi_0 \mathbf{P_{\frac{1}{T}}} \cdots \mathbf{P_{\frac{k}{T}}} - \pi_{\frac{k}{T}}\|_{TV}$ in the following lemma through some telescoping algebra. The proof is given in Section 4.3.

Lemma 1 *One can show that*

$$\|\pi_0 \mathbf{P_{\frac{1}{T}}} \cdots \mathbf{P_{\frac{k}{T}}} - \pi_{\frac{k}{T}}\|_{TV} = \frac{\left| p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)} \right|}{T} \Big| \sum_{j=1}^{k} \left(\frac{\prod_{m=j}^{k} \left(1 - \Delta_{\frac{m}{T}}\right)}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \right) \Big|. \tag{4.6}$$

With this measurement we can find a bound of the stable adiabatic time for the two-state case in terms of the smallest spectral gap over the entire linear adiabatic evolution. We find this bound in the following theorem. The proof can be found in Section 4.3.

Theorem 14 *Given $\epsilon > 0$ we have for a linear adiabatic evolution between two-state Markov chains $\mathbf{P_0}$ and $\mathbf{P_1}$,*

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{1}{\epsilon\Delta^3}. \tag{4.7}$$

In this case we have a nice bound of the stable adiabatic time, but this is a severely limited result. In Sections 4.4 and 4.6, we attempt to find a similar result for a general n-state Markov chain.

4.3 Proofs

4.3.1 Proof of Proposition 10

We can see that the characteristic equation of $\mathbf{P_t}$ will be

$$(1 - p_1^{(t)} - \lambda) \cdot (1 - q_2^{(t)} - \lambda) - p_1^{(t)} \cdot q_2^{(t)} = 0.$$

We can write this as

$$\lambda^2 - (2 - \Delta_t)\lambda + (1 - \Delta_t) = 0.$$

This factors to

$$(\lambda - 1) \cdot (\lambda - (1 - \Delta_t)) = 0.$$

Indeed the roots are 1 and $1 - \Delta_t$.

4.3.2 Proof of Proposition 11

First notice that $\Delta_t > 0$ and both $p_1^{(t)} \geq 0$ and $q_2^{(t)} \geq 0$. This implies that the entries of our suggested π_t are nonnegative.

Next notice that

$$\begin{aligned}\pi_t(1)+\pi_t(2) &= \frac{q_2^{(t)}}{\Delta_t}+\frac{p_1^{(t)}}{\Delta_t}\\ &= \frac{p_1^{(t)}+q_2^{(t)}}{\Delta_t}\\ &= \frac{\Delta_t}{\Delta_t}\\ &= 1.\end{aligned}$$

This suggests that π_t is a probability distribution. Finally we show that

$$\begin{aligned}\pi_t\mathbf{P_t} &= \begin{bmatrix} \frac{q_2^{(t)}}{\Delta_t} & \frac{p_1^{(t)}}{\Delta_t}\end{bmatrix}\cdot\begin{bmatrix} 1-p_1^{(t)} & p_1^{(t)}\\ q_2^{(t)} & 1-q_2^{(t)}\end{bmatrix}\\ &= \begin{bmatrix} \frac{q_2^{(t)}}{\Delta_t}-\frac{p_1^{(t)}q_2^{(t)}}{\Delta_t}+\frac{p_1^{(t)}q_2^{(t)}}{\Delta_t} & \frac{p_1^{(t)}q_2^{(t)}}{\Delta_t}+\frac{p_1^{(t)}}{\Delta_t}-\frac{p_1^{(t)}q_2^{(t)}}{\Delta_t}\end{bmatrix}\\ &= \begin{bmatrix} \frac{q_2^{(t)}}{\Delta_t} & \frac{p_1^{(t)}}{\Delta_t}\end{bmatrix}\\ &= \pi_t.\end{aligned}$$

4.3.3 Proof of Proposition 12

$$\begin{aligned}\begin{bmatrix} 1 & -1\end{bmatrix}\mathbf{P_t} &= \begin{bmatrix} 1 & -1\end{bmatrix}\begin{bmatrix} 1-p_1^{(t)} & p_1^{(t)}\\ q_2^{(t)} & 1-q_2^{(t)}\end{bmatrix}\\ &= \begin{bmatrix} 1-p_1^{(t)}-q_2^{(t)} & p_1^{(t)}-1+q_2^{(t)}\end{bmatrix}\\ &= (1-\Delta_t)\begin{bmatrix} 1 & -1\end{bmatrix}.\end{aligned}$$

4.3.4 Proof of Proposition 13

Remember from Proposition 11 that

$$\pi_t = \begin{bmatrix} \frac{p_1^{(t)}}{\Delta_t} & \frac{q_2^{(t)}}{\Delta_t} \end{bmatrix}.$$

This will imply that

$$\begin{aligned}
&\pi_{\frac{j-1}{T}} - \pi_{\frac{j}{T}} \\
&= \begin{bmatrix} \frac{q_2^{(\frac{j-1}{T})}}{\Delta_{\frac{j-1}{T}}} - \frac{q_2^{(\frac{j}{T})}}{\Delta_{\frac{j}{T}}} & \frac{p_1^{(\frac{j-1}{T})}}{\Delta_{\frac{j-1}{T}}} - \frac{p_1^{(\frac{j}{T})}}{\Delta_{\frac{j}{T}}} \end{bmatrix} \\
&= \frac{1}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} q_2^{(\frac{j-1}{T})} \cdot \Delta_{\frac{j}{T}} - q_2^{(\frac{j}{T})} \cdot \Delta_{\frac{j-1}{T}} & p_1^{(\frac{j-1}{T})} \cdot \Delta_{\frac{j}{T}} - p_1^{(\frac{j}{T})} \cdot \Delta_{\frac{j-1}{T}} \end{bmatrix} \\
&= \frac{1}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} q_2^{(\frac{j-1}{T})} p_1^{(\frac{j}{T})} - q_2^{(\frac{j}{T})} p_1^{(\frac{j-1}{T})} & p_1^{(\frac{j-1}{T})} q_2^{(\frac{j}{T})} - p_1^{(\frac{j}{T})} q_2^{(\frac{j-1}{T})} \end{bmatrix} \\
&= \frac{q_2^{(\frac{j-1}{T})} p_1^{(\frac{j}{T})} - q_2^{(\frac{j}{T})} p_1^{(\frac{j-1}{T})}}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \\
&= \frac{\frac{j}{T} q_2^{(0)}(p_1^{(1)} - p_1^{(0)}) + \frac{j-1}{T} p_1^{(0)}(q_2^{(1)} - q_2^{(0)}) - q_2^{(\frac{j}{T})} p_1^{(\frac{j-1}{T})}}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \\
&= \frac{\frac{j}{T} q_2^{(0)} p_1^{(1)} + \frac{j-1}{T} p_1^{(0)} q_2^{(1)} - \frac{j}{T} p_1^{(0)}(q_2^{(1)} - q_2^{(0)}) - \frac{j-1}{T} q_2^{(0)}(p_1^{(1)} - p_1^{(0)})}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \\
&= \frac{\frac{j}{T} q_2^{(0)} p_1^{(1)} + \frac{j-1}{T} p_1^{(0)} q_2^{(1)} - \frac{j}{T} p_1^{(0)} q_2^{(1)} - \frac{j-1}{T} q_2^{(0)} p_1^{(1)}}{\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \\
&= \frac{p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}}{T\Delta_{\frac{j-1}{T}}\Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix}.
\end{aligned}$$

4.3.5 Proof of Lemma 1

Notice that

$$\begin{aligned}
\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}} &= (\pi_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}) \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} \\
&+ (\pi_{\frac{k-2}{T}} - \pi_{\frac{k-1}{T}}) \mathbf{P}_{\frac{\mathbf{k-1}}{\mathbf{T}}} \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} \\
&+ (\pi_{\frac{k-3}{T}} - \pi_{\frac{k-2}{T}}) \mathbf{P}_{\frac{\mathbf{k-2}}{\mathbf{T}}} \mathbf{P}_{\frac{\mathbf{k-1}}{\mathbf{T}}} \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} \\
&+ \cdots \\
&+ (\pi_0 - \pi_{\frac{1}{T}}) \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \mathbf{P}_{\frac{\mathbf{2}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}.
\end{aligned}$$

Indeed

$$\begin{aligned}
\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}} &= \sum_{j=1}^{k} \left(\pi_{\frac{j-1}{T}} - \pi_{\frac{j}{T}} \right) \mathbf{P}_{\frac{\mathbf{j}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} \\
&= \sum_{j=1}^{k} \left(\pi_{\frac{j-1}{T}} - \pi_{\frac{j}{T}} \right) \prod_{m=j}^{k} \mathbf{P}_{\frac{\mathbf{m}}{\mathbf{T}}}.
\end{aligned}$$

We can see from Proposition 13 that this becomes

$$\begin{aligned}
&\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}} \\
&\quad = \frac{p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}}{T} \sum_{j=1}^{k} \frac{1}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \begin{bmatrix} 1 & -1 \end{bmatrix} \prod_{m=j}^{k} \mathbf{P}_{\frac{\mathbf{m}}{\mathbf{T}}}.
\end{aligned}$$

From Proposition 12, we can use that the fact that $\begin{bmatrix} 1 & -1 \end{bmatrix}$ is an eigenvector for all $\mathbf{P_t}$ to show

$$\begin{aligned}
&\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}} \\
&\quad = \frac{p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}}{T} \sum_{j=1}^{k} \left(\frac{\prod_{m=j}^{k} \left(1 - \Delta_{\frac{m}{T}}\right)}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \right) \begin{bmatrix} 1 & -1 \end{bmatrix}.
\end{aligned}$$

Now taking the the norm, we find

$$\begin{aligned}&\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} \\ &\qquad = \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T} \Big| \sum_{j=1}^{k} \left(\frac{\prod_{m=j}^{k} \left(1 - \Delta_{\frac{m}{T}}\right)}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \right) \Big|.\end{aligned}$$

4.3.6 Proof of Theorem 14

From Lemma 1 we have that

$$\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} = \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T} \Big| \sum_{j=1}^{k} \left(\frac{\prod_{m=j}^{k} \left(1 - \Delta_{\frac{m}{T}}\right)}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \right) \Big|.$$

We see then that

$$\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} \leq \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T} \sum_{j=1}^{k} \left(\frac{\prod_{m=j}^{k} |1 - \Delta_{\frac{m}{T}}|}{\Delta_{\frac{j-1}{T}} \Delta_{\frac{j}{T}}} \right).$$

Because $\Delta \leq 1 - \left|1 - \Delta_t\right| \leq \Delta_t$ for all $t \in [0,1]$ we see that

$$\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} \leq \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T\Delta^2} \sum_{j=1}^{k} \left(\prod_{m=j}^{k} \sup_{0 \leq t \leq 1} \{\left|1 - \Delta_t\right|\} \right).$$

It is clear then that

$$\begin{aligned}\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} &\leq \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T\Delta^2} \sum_{j=1}^{k} \left(\sup_{0 \leq t \leq 1} \{\left|1 - \Delta_t\right|\} \right)^{k-j+1} \\ &\leq \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T\Delta^2} \sum_{j=0}^{k} \left(\sup_{0 \leq t \leq 1} \{\left|1 - \Delta_t\right|\} \right)^{j}.\end{aligned}$$

Noticing the geometric series we can write this as

$$\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} \leq \frac{|p_1^{(1)} q_2^{(0)} - p_1^{(0)} q_2^{(1)}|}{T\Delta^2} \left(\frac{1 - (\sup_{0 \leq t \leq 1}\{|1 - \Delta_t|\})^{k+1}}{1 - \sup_{0 \leq t \leq 1}\{|1 - \Delta_t|\}} \right).$$

Because $\inf_{0\le t\le 1}\{1-\left|1-\Delta_t\right|\} = 1-\sup_{0\le t\le 1}\{\left|1-\Delta_t\right|\}$ it is easy to see that

$$\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} \le \frac{1}{T\Delta^3}.$$

Because this is true for $1 \le k \le T$, then

$$\max\{\|\pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{TV} : 1 \le k \le T\} \le \frac{1}{T\Delta^3}.$$

Now for $\epsilon > 0$ setting

$$\frac{1}{T\Delta^3} \le \epsilon$$

and solving for T we see that

$$T \ge \frac{1}{\epsilon\Delta^3}.$$

This implies that

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \le \frac{1}{\epsilon\Delta^3}.$$

4.4 Reversible Markov Chains

This section attempts to improve upon the results from Section 4.2. We consider n-state, discrete-time Markov chains that are called reversible. Recall that Definition 11 in Section 1.3 defines a reversible Markov chain. We again attempt to bound the stable adiabatic time by the smallest spectral gap of a linear adiabatic evolution, but now it is between two reversible Markov chains. We will derive a result that is similar to the results from Section 3.1. To achieve our result we have to explore what it means to be reversible and to develop this requires the introduction of the following inner products.

Definition 24 *For vectors* $\mathbf{x}, \mathbf{y}, \pi_t \in \mathbb{R}^n$ *we define the inner product*

$$< \mathbf{x}, \mathbf{y} >_{\pi_t} = \sum_{i=1}^{n} \mathbf{x}(i) \cdot \mathbf{y}(i) \cdot \pi_t(i). \tag{4.8}$$

Definition 25 *For vectors* $\mathbf{x}, \mathbf{y}, \pi_t \in \mathbb{R}^n$ *we define the inner product*

$$< \mathbf{x}, \mathbf{y} >_{\frac{1}{\pi_t}} = \sum_{i=1}^{n} \mathbf{x}(i) \cdot \mathbf{y}(i) \cdot \frac{1}{\pi_t(i)}. \tag{4.9}$$

Before we talk about reversibility we relate the norm created by the inner product in Definition 25 and the total variation norm. This is a small detour, but this relationship will be important in the development of our result. This proposition has been shown in [6], but we include its proof in Section 4.5 since it is so short.

Proposition 14 *For any probability distribution* ν *and any time* $t \in [0, 1]$,

$$\|\nu - \pi_t\|_{TV} \leq \frac{1}{2} \|\nu - \pi_t\|_{\frac{1}{\pi_t}}. \tag{4.10}$$

The inner products from Definitions 24 and 25 are commonly used in the context of stochastic processes and we want to use Definition 24 to redefine reversible Markov chains in a second way. We highlight the following proposition from [6] and we include its proof in Section 4.5.

Proposition 15 $\mathbf{P_t}$ *is self-adjoint with with respect to* $< \cdot, \cdot >_{\pi_t}$ *if and only if* $\mathbf{P_t}$ *is reversible with with respect to* π_t.

The results in this section will apply to transitions, $\mathbf{P_t}$, that are self-adjoint with respect to $< \cdot, \cdot >_{\pi_t}$ for all $t \in [0, 1]$.

The inner products defined above are important because we can find bases of $\mathbb{R}^n$ consisting of left eigenvectors and right eigenvectors of $\mathbf{P}_t$ that are orthonormal with respect to $< \cdot, \cdot >_{\pi_t}$ and $< \cdot, \cdot >_{\frac{1}{\pi_t}}$ respectively. To find these bases we must first define a matrix and perform the spectral theorem of linear algebra on this matrix. We assume that $\pi_t(i) \neq 0$ for all $t \in [0,1]$. The following proposition can be found in [6] and we include its proof in Section 4.5.

Proposition 16 *Defining* $\mathbf{P_t^*} = \mathbf{D_t}^{\frac{1}{2}}\mathbf{P_t}\mathbf{D_t}^{-\frac{1}{2}}$ *where* $\mathbf{D_t} = diag\{\pi_t(1), \cdots, \pi_t(n)\}$ *is a* $n \times n$ *diagonal matrix, we see that* $\mathbf{P_t^*}$ *is a symmetric matrix.*

Because $\mathbf{P_t^*}$ is symmetric, we see that it is self-adjoint with respect to the usual Euclidean inner product. A nice property of self-adjoint matrices with real entries is that all the eigenvalues of the matrix are real. For a given $t \in [0,1]$ let $\lambda_1(t), \cdots, \lambda_n(t)$ be the not-necessarily-distinct, real eigenvalues for $\mathbf{P_t^*}$ ordered such that $|\lambda_1(t)| \geq \cdots \geq |\lambda_n(t)|$. Because $\mathbf{P_t^*}$ is self-adjoint with respect to the Euclidean inner product, we know that there exists an orthonormal basis with respect to the Euclidean inner product for $\mathbb{R}^n$ consisting of left eigenvectors of $\mathbf{P_t^*}$. Let $\mathbf{w}_{1,t}, \cdots, \mathbf{w}_{n,t}$ be this orthonormal basis with respect to the Euclidean inner product such that $\mathbf{w}_{i,t}$ corresponds to the eigenvalue $\lambda_i(t)$. Because $\mathbf{P_t^*}$ is symmetric, we see that $\mathbf{w}_{1,t}, \cdots, \mathbf{w}_{n,t}$ is also an orthonormal basis with respect to the Euclidean inner product whose transposes consist of right eigenvectors. Here $\mathbf{w}_{i,t}^T$ corresponds to $\lambda_i(t)$.

For $1 \leq i \leq n$ define $\mathbf{u}_{i,t}$ such that $\mathbf{w}_{i,t} = \mathbf{u}_{i,t}\mathbf{D_t}^{-\frac{1}{2}}$ and define $\mathbf{v}_{i,t}$ such that $\mathbf{w}_{i,t} = \mathbf{v}_{i,t}\mathbf{D_t}^{\frac{1}{2}}$. The following two propositions tell us that we have found a collection of left eigenvectors of $\mathbf{P_t}$ and a collection of right eigenvectors of $\mathbf{P_t}$. These were also shown in [6] and we will show the proofs of these propositions in Section 4.5.

Proposition 17 *For* $1 \leq i \leq n$, $\lambda_i(t)$ *is an eigenvalue for* $\mathbf{P_t}$ *and* $\mathbf{u}_{i,t}$ *is a left eigenvector of* $\mathbf{P_t}$ *that corresponds to* $\lambda_i(t)$.

Proposition 18 *For $1 \leq i \leq n$, $\mathbf{v}_{i,t}^T$ is a right eigenvector of $\mathbf{P_t}$ that corresponds to $\lambda_i(t)$.*

Now we relate these above collections of eigenvectors to the two inner products that we defined at the beginning of this section. We will see that the collection of left eigenvectors of $\mathbf{P_t}$ are orthonormal with respect to $< \cdot, \cdot >_{\frac{1}{\pi_t}}$ and using the dimension of $\mathbb{R}^n$ we conclude that this collection of eigenvectors forms a basis of $\mathbb{R}^n$. Similarly we see that the collection of right eigenvectors forms an orthonomral basis of $\mathbb{R}^n$ with respect to $< \cdot, \cdot >_{\pi_t}$. These bases were expressed in [6]. We summarize this information in the following propositions, and we prove that these eigenvectors form an orthonormal basis in Section 4.5.

Proposition 19 $\mathbf{u}_{1,t}, \cdots, \mathbf{u}_{n,t}$ *is an orthonormal basis for $\mathbb{R}^n$ with respect to $< \cdot, \cdot >_{\frac{1}{\pi_t}}$.*

Proposition 20 $\mathbf{v}_{1,t}, \cdots, \mathbf{v}_{n,t}$ *is an orthonormal basis for $\mathbb{R}^n$ with respect to $< \cdot, \cdot >_{\pi_t}$.*

Proposition 19 is used in the proof of the following Proposition. The inequality in the Proposition was coined the Perron-Frobenius inequality in [6]. We originally tried to use this inequality along with some telescoping algebra to find a bound of the stable adiabatic time, but the norm created by $< \cdot, \cdot >_{\frac{1}{\pi_t}}$ is called a 'local' norm and we were unable to find a reasonable bound for a general reversible matrix. The proof of the following proposition was included in Section 4.5.

Proposition 21 *For self-adjoint matrices with respect to $< \cdot, \cdot >_{\pi_t}$, we have that for any distribution ν^T,*

$$\|\nu \mathbf{P_t} - \pi_t\|_{\frac{1}{\pi_t}} \leq |\lambda_2(t)| \|\nu - \pi_t\|_{\frac{1}{\pi_t}}. \tag{4.11}$$

We can also use Proposition 19 in the development of a process called eigenvalue perturbation. This process can be referenced in [25]. We use eigenvalue perturbation to find an upper bound of the $\|\cdot\|_{\frac{1}{\pi_t}}$-norm of the difference of fractionally consecutive stationary distributions in our linear adiabatic evolution with respect to the smallest spectral gap over the entire evolution. We summarize this in the following proposition. The proof of this proposition can be found in Section 4.5.

Proposition 22 *Given $1 \leq i \leq T$ we see that*

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{1}{T\Delta}\|\pi_{\frac{i-1}{T}}\left(\mathbf{P_1} - \mathbf{P_0}\right)\|_{\frac{1}{\pi_{\frac{i}{T}}}} \tag{4.12}$$

We can now apply the mathematics discussed in Propositions 21 and 22 to find an upper bound for $\|\pi_0 \mathbf{P}_{\frac{1}{T}} \cdots \mathbf{P}_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV}$ with respect to the smallest spectral gap of the linear adiabatic evolution between $\mathbf{P}_0$ and $\mathbf{P}_1$ and some other terms. The following lemma describes this bound and we include the proof of this lemma in Section 4.5.

Lemma 2 *Define $\nu_{\frac{k}{T}} = \pi_0 \mathbf{P}_{\frac{\mathbf{1}}{\mathbf{T}}} \cdots \mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}}$ for $1 \leq k \leq T$ and $\nu_0 = \pi_0$. Also let Δ be the smallest spectral gap over the linear adiabatic evolution between $\mathbf{P}_0$ and $\mathbf{P}_1$.*

Letting

$$c(\frac{k}{T}) = \min_{0 \leq i \leq k}\left\{\min_{1 \leq m \leq n}\left\{\pi_{\frac{i}{T}}(m)\right\}\right\}$$

we see that under the assumption that $\pi_t(m) > 0$ *for all* $t \in [0,1]$ *and* $1 \le m \le n$,

$$\begin{aligned}
&\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \\
&\le \frac{1}{2T\Delta} \max_{0 \le i \le k-1} \left\{ \|\pi_{\frac{i}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right\} \\
&\quad \sum_{j=0}^{k-1} (1-\Delta)^{j+1} \left(1 + \frac{1}{T\Delta} \frac{1}{\sqrt{c(\frac{k}{T})}} \max_{0 \le i \le k-1} \left\{ \|\pi_{\frac{i}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right\} \right)^{j+1}
\end{aligned} \tag{4.13}$$

At this point our goal was to find the nature of the stable adiabatic time strictly of a function of Δ but there were some complications getting this bound. As we have mentioned earlier, $\|\cdot\|_{\frac{1}{\pi_t}}$ is a 'local' norm, so we cannot determine the nature of

$$\max_{0 \le i \le T-1} \left\{ \|\pi_{\frac{i}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right\}$$

in terms of Δ as $\Delta \to 0$. Due to this, we wrote the most descriptive result in this setting. The following theorem describes a bound of the stable adiabatic time and we include the proof of this theorem in Section 4.5

Theorem 15 *Let* $\epsilon > 0$ *and let* Δ *be the smallest spectral gap over the linear adiabatic evolution between* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *where both matrices are reversible Markov chains with respect to* π_0 *and* π_1 *respectively. Letting*

$$c^* = \min_{t \in [0,1]} \left\{ \min_{1 \le m \le n} \{\pi_t(m)\} \right\}$$

we see under the assumption that $\pi_t(m) > 0$ *for* $t \in [0,1]$ *and* $1 \le m \le n$,

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \le \frac{\frac{1}{\epsilon} + 1}{\Delta^2 \sqrt{c^*}} \max_{T^* \ge 2} \left\{ \max_{0 \le i \le T^*-1} \left\{ \|\pi_{\frac{i}{T^*}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T^*}}}} \right\} \right\} \tag{4.14}$$

Because we were unable to find a general result for reversible Markov chains, we decided to look at a subclass of these Markov chains called Birth-Death Markov chains. Instead of applying what we know about Birth-Death Markov chains to the previous theorem, we instead introduce an entirely new development of the bound. This is done to highlight the many attempts we made toward our final result and to give the reader some food for thought. We return to Birth-Death Markov chains in Section 4.6.

4.5 Proofs

4.5.1 Proof of Proposition 14

We start by looking at the right hand side of the equation squared.

$$\begin{aligned}
\|\nu - \pi_t\|^2_{\frac{1}{\pi_t}} &= \sum_{i=1}^{n} (\nu(i) - \pi_t(i))^2 \frac{1}{\pi_t(i)} \\
&= \sum_{i=1}^{n} \left(\frac{\nu(i)}{\pi_t(i)} - 1 \right)^2 \pi_t(i).
\end{aligned}$$

Using the Cauchy-Schwarz inequality, we see that

$$\left(\sum_{i=1}^{n} |\frac{\nu(i)}{\pi_t(i)} - 1| \sqrt{\pi_t(i)} \right)^2 \leq \sum_{i=1}^{n} \left(|\frac{\nu(i)}{\pi_t(i)} - 1| \sqrt{\pi_t(i)} \right)^2.$$

Because $\pi_t(i) \leq 1$ for $1 \leq i \leq n$, we see that

$$\begin{aligned}
\left(\sum_{i=1}^{n} |\frac{\nu(i)}{\pi_t(i)} - 1| \sqrt{\pi_t(i)} \right)^2 \pi_t(i) &\leq \sum_{i=1}^{n} \left(|\frac{\nu(i)}{\pi_t(i)} - 1| \sqrt{\pi_t(i)} \right)^2 \\
&= \|\nu - \pi_t\|^2_{\frac{1}{\pi_t}}.
\end{aligned}$$

However

$$(2\|\nu-\pi_t\|_{TV})^2 = \left(\sum_{i=1}^{n} |\nu(i)-\pi_t(i)|\right)^2$$
$$= \left(\sum_{i=1}^{n} |\frac{\nu(i)}{\pi_t(i)} - 1|\sqrt{\pi_t(i)}\right)^2 \pi_t(i).$$

We therefore see that

$$\|\nu-\pi_t\|_{TV} \le \frac{1}{2}\|\nu-\pi_t\|_{\frac{1}{\pi_t}}.$$

4.5.2 Proof of Proposition 15

Letting $\mu, \nu \in \mathbb{R}^n$ be column vectors we have that if $\mathbf{P_t}$ is self-adjoint with respect to $< \cdot, \cdot >_{\pi_t}$, then $< \mu, \mathbf{P_t}\nu >_{\pi_t} = < \mathbf{P_t}\mu, \nu >_{\pi_t}$.

Letting $\mu = \mathbf{e}_i$ and $\nu = \mathbf{e}_j$ be standard basis vectors, we see that $< \mu, \mathbf{P_t}\nu >_{\pi_t} = \pi_t(i)\mathbf{P_t}(i,j)$ and $< \mathbf{P_t}\mu, \nu >_{\pi_t} = \pi_t(j)\mathbf{P_t}(j,i)$.

It is clear that $\pi_t(i)\mathbf{P_t}(i,j) = \pi_t(j)\mathbf{P_t}(j,i)$ for $1 \le i,j \le n$, therefore $\mathbf{P_t}$ is reversible with respect to π_t.

If $\mathbf{P_t}$ is reversible with respect to π_t, then $< \mathbf{e}_i, \mathbf{P_t}\mathbf{e}_j >_{\pi_t} = < \mathbf{P_t}\mathbf{e}_i, \mathbf{e}_j >_{\pi_t}$ for $1 \le i,j \le n$.

For column vectors $\mu, \nu \in \mathbb{R}^n$ there exist constants $c_1, \cdots, c_n$ and $d_1, \cdots, d_n \in \mathbb{R}$

such that $\mu = \sum_{i=1}^{n} c_i \mathbf{e}_i$ and $\nu = \sum_{j=1}^{n} d_j \mathbf{e}_j$ we have that

$$\begin{aligned}
< \mu, \mathbf{P_t} \nu >_{\pi_t} &=< \sum_{i=1}^{n} c_i \mathbf{e}_i, \mathbf{P_t} \sum_{j=1}^{n} d_j \mathbf{e}_j >_{\pi_t} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i d_j < \mathbf{e}_i, \mathbf{P_t} \mathbf{e}_j >_{\pi_t} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i d_j < \mathbf{P_t} \mathbf{e}_i, \mathbf{e}_j >_{\pi_t} \\
&=< \mathbf{P_t} \sum_{i=1}^{n} c_i \mathbf{e}_i, \sum_{j=1}^{n} d_j \mathbf{e}_j >_{\pi_t} \\
&=< \mathbf{P_t} \mu, \nu >_{\pi_t} .
\end{aligned}$$

It is clear that if $\mathbf{P_t}$ is reversible with respect to π_t, then $\mathbf{P_t}$ is self-adjoint with respect to $< \cdot, \cdot >_{\pi_t}$.

4.5.3 Proof of Proposition 16

We see that

$$(\mathbf{P}_\mathbf{t}^*(i,j))_{n \times n} = \left(\frac{\sqrt{\pi_t(i)}}{\sqrt{\pi_t(j)}} \mathbf{P_t}(i,j) \right)_{n \times n}$$

Notice that for reversible Markov chains,

$$\begin{aligned}
\frac{\sqrt{\pi_t(i)}}{\sqrt{\pi_t(j)}} \mathbf{P_t}(i,j) &= \frac{1}{\sqrt{\pi_t(i)}\sqrt{\pi_t(j)}} \pi_t(i) \mathbf{P_t}(i,j) \\
&= \frac{1}{\sqrt{\pi_t(i)}\sqrt{\pi_t(j)}} \pi_t(j) \mathbf{P_t}(j,i) \\
&= \frac{\sqrt{\pi_t(j)}}{\sqrt{\pi_t(i)}} \mathbf{P_t}(j,i).
\end{aligned}$$

This tell us that $\mathbf{P}_\mathbf{t}^*(i,j) = \mathbf{P}_\mathbf{t}^*(j,i)$.

Therefore $\mathbf{P}^*_\mathbf{t}$ is symmetric.

4.5.4 Proof of Proposition 17

For $1 \leq i \leq n$, we see that

$$\begin{aligned}
\mathbf{u}_{i,t}\mathbf{P}_\mathbf{t} &= \mathbf{u}_{i,t}\left(\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{P}^*_\mathbf{t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}}\right) \\
&= \left(\mathbf{u}_{i,t}\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\right)\mathbf{P}^*_\mathbf{t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}} \\
&= \mathbf{w}_{i,t}\mathbf{P}^*_\mathbf{t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}} \\
&= \lambda_i(t)\mathbf{w}_{i,t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}} \\
&= \lambda_i(t)\left(\mathbf{u}_{i,t}\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\right)\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}} \\
&= \lambda_i(t)\mathbf{u}_{i,t}\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}} \\
&= \lambda_i(t)\mathbf{u}_{i,t}.
\end{aligned}$$

4.5.5 Proof of Proposition 18

For $1 \leq i \leq n$, we see that

$$\begin{aligned}
\mathbf{P}_\mathbf{t}\mathbf{v}^T_{i,t} &= \left(\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{P}^*_\mathbf{t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}}\right)\mathbf{v}^T_{i,t} \\
&= \mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{P}^*_\mathbf{t}\left(\mathbf{v}_{i,t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}}\right)^T \\
&= \mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{P}^*_\mathbf{t}\mathbf{w}^T_{i,t} \\
&= \lambda_i(t)\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{w}^T_{i,t} \\
&= \lambda_i(t)\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\left(\mathbf{v}_{i,t}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}}\right)^T \\
&= \lambda_i(t)\mathbf{D}_\mathbf{t}{}^{-\frac{1}{2}}\mathbf{D}_\mathbf{t}{}^{\frac{1}{2}}\mathbf{v}^T_{i,t} \\
&= \lambda_i(t)\mathbf{v}^T_{i,t}.
\end{aligned}$$

4.5.6 Proof of Proposition 19

Notice that because $\mathbf{w}_{1,t}, \cdot, \mathbf{w}_{n,t}$ is orthonormal with respect to the Euclidean inner product, denoted $< \cdot, \cdot >_2$, we see that

$$\begin{aligned}
< \mathbf{u}_{i,t}, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} &= \mathbf{u}_{i,t}\mathbf{D_t}^{-1}\mathbf{u}_{j,t}^T \\
&= \mathbf{u}_{i,t}\mathbf{D_t}^{-\frac{1}{2}}\left(\mathbf{D_t}^{-\frac{1}{2}}\right)^T\mathbf{u}_{j,t}^T \\
&= \left(\mathbf{u}_{i,t}\mathbf{D_t}^{-\frac{1}{2}}\right)\cdot\left(\mathbf{u}_{j,t}\mathbf{D_t}^{-\frac{1}{2}}\right)^T \\
&= \mathbf{w}_{i,t}\cdot\mathbf{w}_{j,t}^T \\
&=< \mathbf{w}_{i,t}, \mathbf{w}_{j,t} >_2 \\
&= \delta_{i,j}.
\end{aligned}$$

4.5.7 Proof of Proposition 20

Notice again that

$$\begin{aligned}
< \mathbf{v}_{i,t}, \mathbf{v}_{j,t} >_{\pi_t} &= \mathbf{v}_{i,t}\mathbf{D_t}\mathbf{v}_{j,t}^T \\
&= \mathbf{v}_{i,t}\mathbf{D_t}^{\frac{1}{2}}\left(\mathbf{D_t}^{\frac{1}{2}}\right)^T\mathbf{v}_{j,t}^T \\
&= \left(\mathbf{v}_{i,t}\mathbf{D_t}^{\frac{1}{2}}\right)\cdot\left(\mathbf{v}_{j,t}\mathbf{D_t}^{\frac{1}{2}}\right)^T \\
&= \mathbf{w}_{i,t}\cdot\mathbf{w}_{j,t}^T \\
&=< \mathbf{w}_{i,t}, \mathbf{w}_{j,t} >_2 \\
&= \delta_{i,j}.
\end{aligned}$$

4.5.8 Proof of Proposition 21

Notice that $\|\nu\mathbf{P_t} - \pi_t\|_{\frac{1}{\pi_t}} = \|(\nu - \pi_t)\mathbf{P_t}\|_{\frac{1}{\pi_t}}$.

Because $\mathbf{u}_{1,t}, \cdots, \mathbf{u}_{n,t}$ is an orthonormal basis, as shown in proposition 19, we can say that

$$\nu - \pi_t = \sum_{j=1}^{n} < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}.$$

Remember that $\mathbf{u}_{1,t} = \pi_t$.

This means that

$$\begin{aligned}
< \nu - \pi_t, \mathbf{u}_{1,t} >_{\frac{1}{\pi_t}} &= \sum_{i=1}^{n} (\nu(i) - \pi_t(i))\pi_t(i)\frac{1}{\pi_t(i)} \\
&= \sum_{i=1}^{n} \nu(i) - \pi_t(i) \\
&= \sum_{i=1}^{n} \nu(i) - \sum_{i=1}^{n} \pi_t(i) \\
&= 1 - 1 \\
&= 0.
\end{aligned}$$

Therefore

$$\nu - \pi_t = \sum_{j=2}^{n} < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}.$$

We see, because the $\mathbf{u}_{i,t}$ are orthogonal, that

$$\begin{aligned}
\|\nu - \pi_t\|^2_{\frac{1}{\pi_t}} &= \| \sum_{j=2}^{n} < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&= \sum_{j=2}^{n} \| < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}}.
\end{aligned}$$

We also see that

$$(\nu - \pi_t)\mathbf{P_t} = \sum_{j=2}^{n} \lambda_i(t) < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}.$$

This, along with the fact that the $\mathbf{u}_{i,t}$ are orthogonal, implies that

$$\begin{aligned}
\|(\nu - \pi_t)\mathbf{P_t}\|^2_{\frac{1}{\pi_t}} &= \|\sum_{j=2}^{n} \lambda_i(t) < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&= \sum_{j=2}^{n} \|\lambda_i(t) < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&= \sum_{j=2}^{n} |\lambda_i(t)|^2 \| < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&\leq \sum_{j=2}^{n} |\lambda_2(t)|^2 \| < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&= |\lambda_2(t)|^2 \sum_{j=2}^{n} \| < \nu - \pi_t, \mathbf{u}_{j,t} >_{\frac{1}{\pi_t}} \mathbf{u}_{j,t}\|^2_{\frac{1}{\pi_t}} \\
&= |\lambda_2(t)|^2 \|\nu - \pi_t\|^2_{\frac{1}{\pi_t}}.
\end{aligned}$$

We conclude that $\|\nu\mathbf{P_t} - \pi_t\|_{\frac{1}{\pi_t}} \leq |\lambda_2(t)| \|\nu - \pi_t\|_{\frac{1}{\pi_t}}$.

4.5.9 Proof of Proposition 22

Suppose that I want to perturb $\mathbf{P_t}$ by a small amount δ. By this I mean that $\mathbf{P_{t+\delta}} = \mathbf{P_t} + \delta\,(\mathbf{P_1} - \mathbf{P_0})$.

Under this definition we expect the corresponding eigenvalues and eigenvectors for the equations involving $\mathbf{P_{t+\delta}}$ to resemble

$$\lambda_i(t+\delta) = \lambda_i(t) + \lambda_i[\delta]$$

$$\mathbf{u}_{i,t+\delta} = \mathbf{u}_{i,t} + \mathbf{u}_i[\delta]$$

where $\lambda_i[\delta]$ and $\mathbf{u}_i[\delta]$ are the right quantities to make the equations work.

We now want to solve the equation

$$\mathbf{u}_{i,t+\delta}\mathbf{P}_{\mathbf{t}+\delta} = \lambda_i(t+\delta)\mathbf{u}_{i,t+\delta}$$

by expanding all the terms.

We see that

$$\left(\mathbf{u}_{i,t} + \mathbf{u}_i[\delta]\right)\left(\mathbf{P}_\mathbf{t} + \delta\left(\mathbf{P}_\mathbf{1} - \mathbf{P}_\mathbf{0}\right)\right) = \left(\lambda_i(t) + \lambda_i[\delta]\right)\left(\mathbf{u}_{i,t} + \mathbf{u}_i[\delta]\right).$$

This will expand to

$$\begin{aligned}\mathbf{u}_{i,t}\mathbf{P}_\mathbf{t} + \delta\mathbf{u}_{i,t}\left(\mathbf{P}_\mathbf{1} - \mathbf{P}_\mathbf{0}\right) + \mathbf{u}_i[\delta]\mathbf{P}_\mathbf{t} + \delta\mathbf{u}_i[\delta]\left(\mathbf{P}_\mathbf{1} - \mathbf{P}_\mathbf{0}\right)& \\ = \lambda_i(t)\mathbf{u}_{i,t} + \lambda_i(t)\mathbf{u}_i[\delta] + \lambda_i[\delta]\mathbf{u}_{i,t} + \lambda_i[\delta]\mathbf{u}_i[\delta].&\end{aligned}$$

We can first cancel the terms from the equation $\mathbf{u}_{i,t}\mathbf{P}_\mathbf{t} = \lambda_i(t)\mathbf{u}_{i,t}$, then we can collect the terms in the equation to separate higher order terms of δ objects.

$$\begin{aligned}\left(\delta\mathbf{u}_{i,t}\left(\mathbf{P}_\mathbf{1} - \mathbf{P}_\mathbf{0}\right) + \mathbf{u}_i[\delta]\mathbf{P}_\mathbf{t}\right) + \delta\mathbf{u}_i[\delta]\left(\mathbf{P}_\mathbf{1} - \mathbf{P}_\mathbf{0}\right)& \\ = \left(\lambda_i(t)\mathbf{u}_i[\delta] + \lambda_i[\delta]\mathbf{u}_{i,t}\right) + \lambda_i[\delta]\mathbf{u}_i[\delta].&\end{aligned}$$

Letting $\mathbf{h}_i[t,\delta] = \delta\mathbf{u}_i[\delta]\left(\mathbf{P_1} - \mathbf{P_0}\right) - \lambda_i[\delta]\mathbf{u}_i[\delta]$ we can write

$$\delta\mathbf{u}_{i,t}\left(\mathbf{P_1} - \mathbf{P_0}\right) + \mathbf{u}_i[\delta]\mathbf{P_t} + \mathbf{h}_i[t,\delta] = \lambda_i(t)\mathbf{u}_i[\delta] + \lambda_i[\delta]\mathbf{u}_{i,t}.$$

It was mentioned earlier that the left eigenvalues of $\mathbf{P_t}$ are orthonormal with respect to $\|\cdot\|_{\frac{1}{\pi_t}}$ and therefore they create a basis for $\mathbb{R}^n$.

This means that we can write our vectors

$$\mathbf{u}_i[\delta] = \sum_{j=1}^{n} a_{ij}\mathbf{u}_{j,t}$$

where the $a_{ij} \in \mathbb{R}$.

Plugging this sum into our previous equation, we have that

$$\delta\mathbf{u}_{i,t}\left(\mathbf{P_1} - \mathbf{P_0}\right) + \sum_{j=1}^{n} a_{ij}\mathbf{u}_{j,t}\mathbf{P_t} + \mathbf{h}_i[t,\delta] = \lambda_i(t)\sum_{j=1}^{n} a_{ij}\mathbf{u}_{j,t} + \lambda_i[\delta]\mathbf{u}_{i,t}.$$

Now we would multiply both sides of the equation on the right by $\mathbf{D_t^{-1}}\mathbf{u}_{k,t}^T$ where $k \neq i$, we see that

$$\delta \mathbf{u}_{i,t} \left(\mathbf{P_1} - \mathbf{P_0}\right) \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T + \sum_{j=1}^{n} a_{ij} \lambda_j(t) \mathbf{u}_{j,t} \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T + \mathbf{h}_i[t,\delta] \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T$$
$$= \lambda_i(t) \sum_{j=1}^{n} a_{ij} \mathbf{u}_{j,t} \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T + \lambda_i[\delta] \mathbf{u}_{i,t} \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T.$$

Using the orthonormality of the inner product this simplifies to

$$\delta \mathbf{u}_{i,t} \left(\mathbf{P_1} - \mathbf{P_0}\right) \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T + a_{ik} \lambda_k(t) + \mathbf{h}_i[t,\delta] \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T = \lambda_i(t) a_{ik}.$$

Solving for the a_{ik} terms, we see that

$$a_{ik} = \frac{\delta \mathbf{u}_{i,t} \left(\mathbf{P_1} - \mathbf{P_0}\right) \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T}{\lambda_i(t) - \lambda_k(t)} + \frac{\mathbf{h}_i[t,\delta] \mathbf{D_t^{-1}} \mathbf{u}_{k,t}^T}{\lambda_i(t) - \lambda_k(t)}.$$

Also remember that $a_{ij} = \left(\mathbf{u}_{i,t+\delta} - \mathbf{u}_{i,t}\right) \mathbf{D_t^{-1}} \mathbf{u}_{j,t}^T$ by definition.

This would imply that

$$\mathbf{u}_{i,t+\delta} - \mathbf{u}_{i,t} = \left(\mathbf{u}_{i,t+\delta} - \mathbf{u}_{i,t}\right) \mathbf{D_t^{-1}} \mathbf{u}_{i,t}^T \mathbf{u}_{i,t} + \sum_{j=1:j\neq i}^{n} \frac{\delta \mathbf{u}_{i,t} \left(\mathbf{P_1} - \mathbf{P_0}\right) \mathbf{D_t^{-1}} \mathbf{u}_{j,t}^T}{\lambda_i(t) - \lambda_j(t)} u_{j,t}$$
$$+ \sum_{j=1:j\neq i}^{n} \frac{\mathbf{h}_i[t,\delta] \mathbf{D_t^{-1}} \mathbf{u}_{j,t}^T}{\lambda_i(t) - \lambda_j(t)} u_{j,t}.$$

In our case, we are particularly interested in $\mathbf{u}_{1,t+\delta} - \mathbf{u}_{1,t} = \pi_{t+\delta} - \pi_t$. This is ideal because $(\mathbf{u}_{1,t+\delta} - \mathbf{u}_{1,t})\,\mathbf{D_t^{-1}}\mathbf{u}_{1,t}^T = 0$.

Using the notation that $\mathbf{u}_{1,t} = \pi_t$ and noting that $\lambda_1(t) = 1$ for $t \in [0,1]$, the previous equation becomes

$$\pi_{t+\delta} - \pi_t = \delta \sum_{j=2}^{n} \frac{\pi_t\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D_t^{-1}}\mathbf{u}_{j,t}^T}{1-\lambda_j(t)}\mathbf{u}_{j,t} + \sum_{j=2}^{n} \frac{\mathbf{h}_1[t,\delta]\mathbf{D_t^{-1}}\mathbf{u}_{j,t}^T}{1-\lambda_j(t)}\mathbf{u}_{j,t}.$$

Naturally we can combine these two sums, express $\mathbf{h}_1[t,\delta] = (\pi_{t+\delta} - \pi_t)\,(\mathbf{P_1} - \mathbf{P_0})$ and evaluate this at $t = \frac{i}{T}$ and $\delta = -\frac{1}{T}$ to find that

$$\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}} = -\frac{1}{T}\sum_{j=2}^{n} \frac{\pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T}{1-\lambda_j(\frac{i}{T})}\mathbf{u}_{j,\frac{i}{T}}.$$

Taking the squared norm of either side we see that orthogonality gives us

$$\begin{aligned}
\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} &= \| -\frac{1}{T}\sum_{j=2}^{n} \frac{\pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T}{1-\lambda_j(\frac{i}{T})}\mathbf{u}_{j,\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&= \left(\frac{1}{T}\sum_{j=2}^{n} \frac{\pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T}{1-\lambda_j(\frac{i}{T})}\right)^2 \cdot \|\mathbf{u}_{j,\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&\leq \frac{1}{T^2}\sum_{j=2}^{n} \left(\frac{\pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T}{1-\lambda_j(\frac{i}{T})}\right)^2 \|\mathbf{u}_{j,\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&\leq \frac{1}{T^2\Delta^2}\sum_{j=2}^{n} \left(\pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T\right)^2 \|\mathbf{u}_{j,\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&= \frac{1}{T^2\Delta^2}\|\sum_{j=2}^{n} \pi_{\frac{i-1}{T}}\,(\mathbf{P_1} - \mathbf{P_0})\,\mathbf{D}_{\frac{\mathbf{i}}{\mathbf{T}}}^{-1}\mathbf{u}_{j,\frac{i}{T}}^T\mathbf{u}_{j,\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&= \frac{1}{T^2\Delta^2}\|\pi_{\frac{i-1}{T}}(\mathbf{P_1} - \mathbf{P_0})\|^2_{\frac{1}{\pi_{\frac{i}{T}}}}.
\end{aligned}$$

Now taking the square root of either side, we find that

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \le \frac{1}{T\Delta}\|\pi_{\frac{i-1}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i}{T}}}}.$$

4.5.10 Proof of Lemma 2

We see that $\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}} = \nu_{\frac{k-1}{T}}\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}$.

This implies that $\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} = \|\nu_{\frac{k-1}{T}}\mathbf{P}_{\frac{\mathbf{k}}{\mathbf{T}}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}$.

Using Proposition 21, we see that

$$\begin{aligned}
&\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad\le \left|\lambda_2(\frac{k}{T})\right| \|\nu_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad= \left|\lambda_2(\frac{k}{T})\right| \|\nu_{\frac{k-1}{T}} - \pi_{\frac{k-1}{T}} + \pi_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad\le \left|\lambda_2(\frac{k}{T})\right| \|\nu_{\frac{k-1}{T}} - \pi_{\frac{k-1}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} + \left|\lambda_2(\frac{k}{T})\right| \|\pi_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad= \left|\lambda_2(\frac{k}{T})\right| \sqrt{\sum_{m=1}^{n} (\nu_{\frac{k-1}{T}}(m) - \pi_{\frac{k-1}{T}}(m))^2 \frac{1}{\pi_{\frac{k-1}{T}}(m)} \frac{\pi_{\frac{k-1}{T}}(m)}{\pi_{\frac{k}{T}}(m)}} \\
&\qquad\qquad\qquad\qquad + \left|\lambda_2(\frac{k}{T})\right| \|\pi_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad\le \left|\lambda_2(\frac{k}{T})\right| \max_{1\le m\le n} \left\{ \sqrt{\frac{\pi_{\frac{k-1}{T}}(m)}{\pi_{\frac{k}{T}}(m)}} \right\} \|\nu_{\frac{k-1}{T}} - \pi_{\frac{k-1}{T}}\|_{\frac{1}{\pi_{\frac{k-1}{T}}}} \\
&\qquad\qquad\qquad\qquad + \left|\lambda_2(\frac{k}{T})\right| \|\pi_{\frac{k-1}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}.
\end{aligned}$$

Repeating this process for every $\|\nu_{\frac{i}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}}$ with $1 \leq i \leq k-1$ and collecting terms we see that

$$\begin{aligned}
&\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \\
&\quad \leq \sum_{j=0}^{k-1} \left[\prod_{i=k-j}^{k} \left|\lambda_2(\frac{i}{T})\right| \right] \|\pi_{\frac{k-j-1}{T}} - \pi_{\frac{k-j}{T}}\|_{\frac{1}{\pi_{\frac{k-j}{T}}}} \\
&\quad \left[\prod_{i=k-j}^{k-1} \left(1 + \max_{1\leq m\leq n} \left\{ \frac{\sqrt{\pi_{\frac{i}{T}}(m)} - \sqrt{\pi_{\frac{i+1}{T}}(m)}}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\} \right) \right]
\end{aligned}$$

We note that

$$\begin{aligned}
&\max_{1\leq m\leq n} \left\{ \frac{\sqrt{\pi_{\frac{i}{T}}(m)} - \sqrt{\pi_{\frac{i+1}{T}}(m)}}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \\
&= \max_{1\leq m\leq n} \left\{ \frac{\pi_{\frac{i}{T}}(m) - \pi_{\frac{i+1}{T}}(m)}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)} + \sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \\
&\leq \left[\max_{1\leq m\leq n} \left\{ \frac{\pi_{\frac{i}{T}}(m) - \pi_{\frac{i+1}{T}}(m)}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\} \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)} + \sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\} \right]^2 \\
&= \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)} + \sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \max_{1\leq m\leq n} \left\{ \frac{\pi_{\frac{i}{T}}(m) - \pi_{\frac{i+1}{T}}(m)}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \\
&\leq \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)} + \sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \max_{1\leq m\leq n} \left\{ \left[\frac{\pi_{\frac{i}{T}}(m) - \pi_{\frac{i+1}{T}}(m)}{\sqrt{\pi_{\frac{i+1}{T}}(m)}} \right]^2 \right\} \\
&\leq \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)} + \sqrt{\pi_{\frac{i+1}{T}}(m)}} \right\}^2 \sum_{j=1}^{n} \left[\frac{\pi_{\frac{i}{T}}(j) - \pi_{\frac{i+1}{T}}(j)}{\sqrt{\pi_{\frac{i+1}{T}}(j)}} \right]^2 \\
&\leq \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}} \right\}^2 \sum_{j=1}^{n} \left[\frac{\pi_{\frac{i}{T}}(j) - \pi_{\frac{i+1}{T}}(j)}{\sqrt{\pi_{\frac{i+1}{T}}(j)}} \right]^2
\end{aligned}$$

Noticing that

$$\sum_{m=1}^{n}\left[\frac{\pi_{\frac{i}{T}}(m)-\pi_{\frac{i+1}{T}}(m)}{\sqrt{\pi_{\frac{i+1}{T}}(m)}}\right]^2=\|\pi_{\frac{i}{T}}-\pi_{\frac{i+1}{T}}\|^2_{\frac{1}{\pi_{\frac{i+1}{T}}}}$$

we can conclude by taking the square root of either side

$$\begin{aligned}&\max_{1\le m\le n}\left\{\frac{\sqrt{\pi_{\frac{i}{T}}(m)}-\sqrt{\pi_{\frac{i+1}{T}}(m)}}{\sqrt{\pi_{\frac{i+1}{T}}(m)}}\right\}\\&\qquad\le\max_{1\le m\le n}\left\{\frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}}\right\}\|\pi_{\frac{i}{T}}-\pi_{\frac{i+1}{T}}\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}.\end{aligned}$$

Plugging this into our previous equation, we have that

$$\|\nu_{\frac{k}{T}}-\pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \tag{4.15}$$

$$\begin{aligned}&\le\sum_{j=0}^{k-1}\left[\prod_{i=k-j}^{k}\left|\lambda_2(\frac{i}{T})\right|\right]\|\pi_{\frac{k-j-1}{T}}-\pi_{\frac{k-j}{T}}\|_{\frac{1}{\pi_{\frac{k-j}{T}}}}\\&\left[\prod_{i=k-j}^{k-1}\left(1+\max_{1\le m\le n}\left\{\frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}}\right\}\|\pi_{\frac{i}{T}}-\pi_{\frac{i+1}{T}}\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right)\right]\end{aligned} \tag{4.16}$$

Because $\left|\lambda_2(\frac{j}{T})\right|\le(1-\Delta)$ for $0\le j\le k-1$ we see that

$$\begin{aligned}&\|\nu_{\frac{k}{T}}-\pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}\\&\le\sum_{j=0}^{k-1}(1-\Delta)^{j+1}\|\pi_{\frac{k-j-1}{T}}-\pi_{\frac{k-j}{T}}\|_{\frac{1}{\pi_{\frac{k-j}{T}}}}\\&\left[\prod_{i=k-j}^{k-1}\left(1+\max_{1\le m\le n}\left\{\frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}}\right\}\|\pi_{\frac{i}{T}}-\pi_{\frac{i+1}{T}}\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right)\right]\end{aligned}$$

Now using the result from Proposition 22 we see that

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}$$
$$\leq \frac{1}{T\Delta}\sum_{j=0}^{k-1}(1-\Delta)^{j+1}\|\pi_{\frac{k-j-1}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{k-j}{T}}}}$$
$$\left[\prod_{i=k-j}^{k-1}\left(1+\frac{1}{T\Delta}\max_{1\leq m\leq n}\left\{\frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}}\right\}\|\pi_{\frac{i}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right)\right]$$

Letting

$$c(\frac{k}{T}) = \min_{0\leq i\leq k}\left\{\min_{1\leq m\leq n}\left\{\pi_{\frac{i}{T}}(m)\right\}\right\}$$

we compare $\|\pi_{\frac{i}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}$ to

$$\max_{0\leq i\leq k-1}\left\{\|\pi_{\frac{i}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right\}$$

for each $0 \leq i \leq k-1$ to see that

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}$$
$$\leq \frac{1}{T\Delta}\max_{0\leq i\leq k-1}\left\{\|\pi_{\frac{i}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right\}$$
$$\sum_{j=0}^{k-1}(1-\Delta)^{j+1}\left(1+\frac{1}{T\Delta}\frac{1}{\sqrt{c(\frac{k}{T})}}\max_{0\leq i\leq k-1}\left\{\|\pi_{\frac{i}{T}}(\mathbf{P_1}-\mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right\}\right)^{j+1}$$

Using the fact that $\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \leq \frac{1}{2}\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}}$ we come to our result.

4.5.11 Proof of Theorem 15

From Equation 4.13 in Lemma 2 we see that

$$
\begin{aligned}
&\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \\
&\leq \frac{1}{2T\Delta} \max_{0\leq i\leq k-1} \left\{ \|\pi_{\frac{i}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right\} \\
&\sum_{j=0}^{k-1} (1-\Delta)^{j+1} \left(1 + \frac{1}{T\Delta} \frac{1}{\sqrt{c(\frac{k}{T})}} \max_{0\leq i\leq k-1} \left\{ \|\pi_{\frac{i}{T}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right\} \right)^{j+1}
\end{aligned}
$$

where

$$
c(\frac{k}{T}) = \min_{0\leq i\leq k} \left\{ \min_{1\leq m\leq n} \left\{ \pi_{\frac{i}{T}}(m) \right\} \right\}.
$$

Let

$$
T = \frac{M}{\Delta^2\sqrt{c^*}} \max_{T^*\geq 2} \left\{ \max_{1\leq i\leq T^*-1} \left\{ \|\pi_{\frac{i}{T^*}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T^*}}}} \right\} \right\}
$$

where $M > 1$ is a positive constant.

Because $\sqrt{c^*} \leq 2$ and because of the geometric nature of the series, we see that

$$
\begin{aligned}
\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \leq & \frac{\Delta\sqrt{c^*}}{2M} \sum_{j=0}^{k-1} (1-\Delta)^{j+1} \left(1 + \frac{\Delta}{M}\right)^{j+1} \\
\leq & \frac{\Delta}{M} \sum_{j=0}^{k-1} \left(1 - \left(\frac{M-1}{M}\Delta + \frac{1}{M}\Delta^2 \right) \right)^j \\
\leq & \frac{\Delta}{M} \frac{1 - \left(1 - \left(\frac{M-1}{M}\Delta + \frac{1}{M}\Delta^2\right)\right)^k}{\frac{M-1}{M}\Delta + \frac{1}{M}\Delta^2} \\
\leq & \frac{\Delta}{M} \frac{1}{\frac{M-1}{M}\Delta + \frac{1}{M}\Delta^2} \\
\leq & \frac{1}{(M-1)+\Delta}.
\end{aligned}
$$

Given $\epsilon > 0$ if we let $M = \frac{1}{\epsilon} + 1$, then we see that

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \leq \epsilon.$$

Now we see that if

$$T = \frac{\frac{1}{\epsilon}+1}{\Delta^2\sqrt{c^*}} \max_{T^* \geq 2} \left\{ \max_{0 \leq i \leq T^*-1} \left\{ \|\pi_{\frac{i}{T^*}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T^*}}}} \right\} \right\}$$

then for all $1 \leq k \leq T$

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \leq \epsilon.$$

This would imply that

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{\frac{1}{\epsilon}+1}{\Delta^2\sqrt{c^*}} \max_{T^* \geq 2} \left\{ \max_{0 \leq i \leq T^*-1} \left\{ \|\pi_{\frac{i}{T^*}}(\mathbf{P_1} - \mathbf{P_0})\|_{\frac{1}{\pi_{\frac{i+1}{T^*}}}} \right\} \right\}.$$

4.6 Birth-Death Markov Chains

In this section we seek to improve upon our two-state results and reversible Markov chain results from the Sections 4.2 and 4.4. We consider n-state, discrete-time Markov chains that are called birth-death. This is one of the simplest types of n-state Markov chains because it restricts communication between states that are not adjacent. Birth-death Markov chains are a specific kind of reversible Markov chain, described in Section 4.4, and the two-state Markov chains, described in Section 4.2, are a two-state birth-death chain. We again attempted to compare the stable adiabatic time of these types of Markov chains to the smallest spectral gap of the linear adiabatic evolution of their probability transition matrices. We used a variety of techniques that gave us results that were similar to the results in Chapter 3, but they did highlight some of the problems that arose when comparing the stable adiabatic time to the smallest spectral gap.

Birth-death process have many applications in a variety of areas such as queueing theory. We devote Chapter 7 for an application of a very specific birth-death process outlined in [27]. One can reference [14] for a full description of a birth-death chain, but in this section we will outline some important properties of these chains.

We now want to define the probability transition matrices for two n-state birth-death processes. We will again call them $\mathbf{P_0}$ and $\mathbf{P_1}$.

Let $\mathbf{P_0}(1,j) = \mathbf{P_1}(1,j) = 0$ for $j > 2$. For $2 \leq i \leq n-1$ let $\mathbf{P_0}(i,j) = \mathbf{P_1}(i,j) = 0$ for $j < i-1$ and $j > i+1$. Let $\mathbf{P_0}(n,j) = \mathbf{P_1}(n,j) = 0$ for $j < n-1$. For $a \in \{0,1\}$ let $p_1^{(a)}, \cdots p_{n-1}^{(a)}, r_1^{(a)}, \cdots, r_n^{(a)}, q_2^{(a)}, \cdots, q_n^{(a)} \in [0,1]$ such that for $2 \leq i \leq n-1$, $q_i^{(a)} + r_i^{(a)} + p_i^{(a)} = 1$, $r_1^{(a)} + p_1^{(a)} = 1$ and $q_n^{(a)} + r_n^{(a)} = 1$. Now define for $2 \leq i \leq n-1$, $\mathbf{P_a}(i,i-1) = q_i^{(a)}$, $\mathbf{P_a}(i,i) = r_i^{(a)}$, $\mathbf{P_a}(i,i+1) = p_i^{(a)}$, $\mathbf{P_a}(1,1) = r_1^{(a)}$, $\mathbf{P_a}(1,2) = p_1^{(a)}$, and $\mathbf{P_a}(n,n-1) = q_n^{(a)}$, $\mathbf{P_a}(n,n) = r_n^{(a)}$.

For these matrices to be irreducible and aperiodic, we must have for $a \in \{0,1\}$ that $p_i^{(a)} \neq 0$ for $1 \leq i \leq n-1$ and $q_i^{(a)} \neq 0$ for $2 \leq i \leq n$.

The linear adiabatic evolution between $\mathbf{P_0}$ and $\mathbf{P_1}$ will define a class of discrete-time, birth-death processes with the following probability transition matrix for a given $t \in [0,1]$:

$$\mathbf{P_t} = \begin{bmatrix} r_1^{(t)} & p_1^{(t)} & 0 & \cdots & \cdots & 0 & 0 & 0 \\ q_2^{(t)} & r_2^{(t)} & p_2^{(t)} & \ddots & \cdots & 0 & 0 & 0 \\ 0 & q_3^{(t)} & r_3^{(t)} & \ddots & \ddots & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \ddots & r_{n-2}^{(t)} & p_{n-2}^{(t)} & 0 \\ 0 & 0 & 0 & \cdots & \ddots & q_{n-1}^{(t)} & r_{n-1}^{(t)} & p_{n-1}^{(t)} \\ 0 & 0 & 0 & \cdots & \cdots & 0 & q_n^{(t)} & r_n^{(t)} \end{bmatrix}$$

where $q_i^{(t)} = (1-t)q_i^{(0)} + tq_i^{(1)}$ for $2 \leq i \leq n$, $r_i^{(t)} = (1-t)r_i^{(t)} + tr_i^{(t)}$ for $1 \leq i \leq n$ and $p_i^{(t)} = (1-t)p_i^{(0)} + tp_i^{(1)}$ for $1 \leq i \leq n-1$.

In this section we are further assuming that there exists a $\delta > 0$ such that for $a \in \{0, 1\}$, $p_i^{(a)} \geq \delta$ for $1 \leq i \leq n-1$ and $q_i^{(a)} \geq \delta$ for $2 \leq i \leq n$.

We first want to understand the nature of the stationary distribution. The following proposition gives the stationary distribution the proof of this proposition, which can be found in [14], is transcribed in Section 4.7 to emphasize how one finds the stationary distribution.

Proposition 23 *For the class of irreducible, aperiodic, discrete-time, birth-death processes based on a linear adiabatic evolution between* $\mathbf{P_0}$ *and* $\mathbf{P_1}$*, we have that the stationary distribution of* $\mathbf{P_t}$ *is*

$$\pi_t(i) = \frac{\left(\prod_{j=i+1}^{n} q_j^{(t)}\right)\left(\prod_{j=1}^{i-1} p_j^{(t)}\right)}{\sum_{k=1}^{n}\left(\prod_{j=k+1}^{n} q_j^{(t)}\right)\left(\prod_{j=1}^{k-1} p_j^{(t)}\right)} \tag{4.17}$$

for $1 \leq i \leq n$*.*

Now that we have considered the probability transition matrix and stationary distribution of these birth-death chains, we return our attention to comparing the stable adiabatic time to the smallest spectral gap of the linear adiabatic evolution between $\mathbf{P}_0$ and $\mathbf{P}_1$. To do this, we first study the second largest eigenvalue of $\mathbf{P}_t$ as a function of t, but not necessarily second largest in modulus. The following proposition gives a new representation of this eigenvalue that is commonly called the Rayleigh quotient, but the formula applies to reversible Markov chains. This proposition was cited from [6], however, we include its proof in Section 4.7.

Proposition 24 *Given $t \in [0,1]$, if we let $\lambda_2(t)$ be the second largest eigenvalue of $\mathbf{P}_t$ where $\mathbf{P}_t$ is a reversible matrix with respect to π_t, then*

$$\lambda_2(t) = 1 - \inf\left\{ \frac{<(\mathbf{I} - \mathbf{P_t})x, x>_{\pi_t}}{<x,x>_{\pi_t} - (<x,\mathbf{1}>_{\pi_t})^2} \big| x \neq \mathbf{0} \text{ and } <x,\mathbf{1}>_{\pi_t} = 0 \right\}. \tag{4.18}$$

We would like to use the Rayleigh quotient to find a bound for the stable adiabatic time, but to do this we must first simplify the quotient by using an expression called the Dirichlet form. The following proposition describes the Dirichlet form and its proof, which can be found in [6], is given in Section 4.7.

Proposition 25 *For a reversible matrix $\mathbf{P}_t$ with respect to π_t, we have that*

$$<(\mathbf{I} - \mathbf{P_t})x, x>_{\pi_t} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \pi_t(i)\mathbf{P_t}(i,j)\left(x(j) - x(i)\right)^2. \tag{4.19}$$

Combining Propositions 24 and 25, we find an alternative expression for the second largest eigenvalue of $\mathbf{P_t}$, which we summarize in the following corollary.

Corollary 6 *Given $t \in [0,1]$, if we let $\lambda_2(t)$ be the second largest eigenvalue of*

$\mathbf{P}_t$ *where* $\mathbf{P}_t$ *is a reversible matrix with respect to* π_t*, then*

$$\lambda_2(t) = 1-$$
$$\inf\left\{\frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\pi_t(i)\mathbf{P_t}(i,j)\left(x(j)-x(i)\right)^2}{\sum_{i=1}^{n}\pi_t(i)x(i)^2}\Big| x \neq \mathbf{0} \textit{ and } < x, \mathbf{1} >_{\pi_t} = 0\right\}. \tag{4.20}$$

We are also concerned with bounds of the second largest eigenvalue. One such bound is the lower conductance bound on the second largest eigenvalue. We sum up this result from [6] in the following proposition, but we first define some quantities necessary for the statement of the proposition.

Recall that the state space $E = \{1, \cdots, n\}$ and let $B \subseteq E$ with $B \neq \emptyset$. We define the following quantities:

$$\pi_t(B) = \sum_{i \in B} \pi_t(i) \tag{4.21}$$

$$F_t(B) = \sum_{i \in B, j \in E \setminus B} \pi_t(i)\mathbf{P_t}(i,j) \tag{4.22}$$

$$\psi_t(B, \mathbf{P_t}) = \frac{F_t(B)}{\pi_t(B)} \tag{4.23}$$

$$\phi_t(\mathbf{P_t}) = \inf_{B \subseteq E}\{\psi_t(B, \mathbf{P_t}) | \pi_t(B) \leq 1/2\}. \tag{4.24}$$

For a proof of the proposition, I refer you to [6].

Proposition 26 *Assume that* $\mathbf{P}_t$ *is a reversible matrix with respect to* π_t *and let* $\lambda_2(t)$ *be the second largest real eigenvalue of* $\mathbf{P}_t$*. We then have the following inequality:*

$$1 - 2 \cdot \phi_t(\mathbf{P_t}) \leq \lambda_2(t). \tag{4.25}$$

Recalling the proof of Lemma 2, we see that before using the inequality from

Proposition 22, we derived Equation 4.15:

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{\frac{1}{\pi_{\frac{k}{T}}}} \leq \sum_{j=0}^{k-1} \left[\prod_{i=k-j}^{k} \left|\lambda_2(\frac{i}{T})\right| \right] \|\pi_{\frac{k-j-1}{T}} - \pi_{\frac{k-j}{T}}\|_{\frac{1}{\pi_{\frac{k-j}{T}}}}$$
$$\left[\prod_{i=k-j}^{k-1} \left(1 + \max_{1\leq m\leq n} \left\{ \frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}} \right\} \|\pi_{\frac{i}{T}} - \pi_{\frac{i+1}{T}}\|_{\frac{1}{\pi_{\frac{i+1}{T}}}} \right) \right]$$

where $\nu_{\frac{k}{T}} = \pi_0 \mathbf{P}_{\frac{1}{T}} \cdots \mathbf{P}_{\frac{k}{T}}$ and $\lambda_2(t)$ is the second largest eigenvalue in modulus.

Our goal is to find a bound for $\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV}$. Under the assumption that $\mathbf{P}_t$ is a birth-death probability transition matrix and $\lambda_2(t)$ is the second largest eigenvalue and the second largest eigenvalue in modulus for $t \in [0,1]$, we are able to utilize Corollary 6 and Proposition 26 to find a bound of $\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}}$ for $1 \leq i \leq n$. We summarize this in the following proposition and include the proof of this proposition in Section 4.7.

Proposition 27 *Consider a linear adiabatic evolution between two birth-death matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ with $\lambda_2(t)$ the second largest eigenvalue for all $t \in [0,1]$. If we assume there exists a $\delta > 0$ such that for $a \in \{0,1\}$, $p_i^{(a)} \geq \delta$ for $1 \leq i \leq n-1$ and $q_i^{(a)} \geq \delta$ for $2 \leq i \leq n$, then*

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{\sqrt{8}}{\Delta^2 T} \frac{n^2}{\sqrt{\delta^{3(n-1)}}}. \tag{4.26}$$

We now apply this apply this to Equation 4.15 along with some algebra to derive the following corollary. The proof of this corollary is given in Section 4.7.

Corollary 7 *Consider a linear adiabatic evolution between two birth-death matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ with $\lambda_2(t)$ the second largest eigenvalue for all $t \in [0,1]$ and Δ is the smallest spectral gap over the entire evolution. If we assume there exists a $\delta > 0$ such that for $a \in \{0,1\}$, $p_i^{(a)} \geq \delta$ for $1 \leq i \leq n-1$ and $q_i^{(a)} \geq \delta$ for*

$2 \leq i \leq n$, *then*

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \leq \frac{\sqrt{2}}{\Delta^2 T} \frac{n^2}{\sqrt{\delta^{3(n-1)}}} \sum_{j=0}^{k-1} (1-\Delta)^j \left(1 + \frac{\sqrt{8}}{\Delta^2 T} \frac{n^{5/2}}{\delta^{2(n-1)}}\right)^j. \quad (4.27)$$

We use this corollary to find a bound for the stable adiabatic time for a linear adiabatic evolution between birth-death matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ when $\lambda_2(t)$ is the second largest eigenvalue for all $t \in [0,1]$ and for $a \in \{0,1\}$, $p_i^{(a)} \geq \delta$ for $1 \leq i \leq n-1$ and $q_i^{(a)} \geq \delta$ for $2 \leq i \leq n$. The proof of this theorem is in Section 4.7.

Theorem 16 *Consider a linear adiabatic evolution between two birth-death matrices* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *with* $\lambda_2(t)$ *the second largest eigenvalue for all* $t \in [0,1]$ *and* Δ *is the smallest spectral gap over the entire evolution. If we assume there exists a* $\delta > 0$ *such that for* $a \in \{0,1\}$, $p_i^{(a)} \geq \delta$ *for* $1 \leq i \leq n-1$ *and* $q_i^{(a)} \geq \delta$ *for* $2 \leq i \leq n$, *then*

$$t_{sad}(\mathbf{P_0}, \mathbf{P_1}, \epsilon) \leq \frac{M^*}{\epsilon \Delta^3} \quad (4.28)$$

where M^* *is a constant that depends on* n *and* δ.

It is important to note that M^* is not a general constant, so this result doesn't tell us much in general. It appears that for the right conditions on these types of Markov chains, the asymptotic bound of the stable adiabatic time would be the multiplicative inverse of ϵ multiplied by the inverse cube of smallest spectral gap. You can see that the results in Chapter 3 are much clearer and concise. However, the results in this section combined with the results in Theorems 2 and 3, may lead one to believe that the asymptotic bound in Chapter 3 can be improved to the inverse cube of the largest mixing time.

4.7 Proofs

4.7.1 Proof of Proposition 23

Remember that $\pi_t = \pi_t \mathbf{P_t}$ for $t \in [0, 1]$.

For this specific case, this implies that $\pi_t(1) = r_1^{(t)}\pi_t(1) + q_2^{(t)}\pi_t(2)$.
It also implies that for $2 \leq i \leq n-1$ that $\pi_t(i) = p_{i-1}^{(t)}\pi_t(i-1) + r_i^{(t)}\pi_t(i) + q_{i+1}^{(t)}\pi_t(i+1)$.
Finally it implies that $\pi_t(n) = p_{n-1}^{(t)}\pi_t(n-1) + r_n^{(t)}\pi_t(n)$.

We see now that

$$\pi_t(2) = \frac{(1 - r_1^{(t)})}{q_2^{(t)}}\pi_t(1) = \frac{p_1^{(t)}}{q_2^{(t)}}\pi_t(1).$$

Because this process is reversible with respect to $< \cdot, \cdot >_{\pi_t}$, we see that $p_{i-1}^{(t)}\pi_t(i-1) = q_i^{(t)}\pi_t(i)$ for $2 \leq i \leq n-1$.

We therefore have $q_{i+1}^{(t)}\pi_t(i+1) = (1 - q_i^{(t)} - r_i^{(t)})\pi_t(i) = p_i^{(t)}\pi_t(i)$ for $2 \leq i \leq n-1$.

If we write this as $\pi_t(i+1) = \frac{p_i^{(t)}}{q_{i+1}^{(t)}}\pi_t(i)$ for $2 \leq i \leq n-1$, then you can use the recursion to see that

$$\pi_t(i+1) = \frac{\prod_{j=1}^{i} p_j^{(t)}}{\prod_{j=2}^{i+1} q_j^{(t)}}\pi_t(1)$$

for $2 \leq i \leq n-1$.

Because π_t is a probability distribution, we see that

$$\sum_{k=1}^{n} \frac{\prod_{j=1}^{k-1} p_j^{(t)}}{\prod_{j=2}^{k} q_j^{(t)}}\pi_t(1) = 1$$

which in turn implies that

$$\pi_t(1) = \frac{1}{\sum_{k=1}^{n} \frac{\prod_{j=1}^{k-1} p_j^{(t)}}{\prod_{j=2}^{k} q_j^{(t)}}}.$$

We now see that

$$\pi_t(i) = \frac{\frac{\prod_{j=1}^{i-1} p_j^{(t)}}{\prod_{j=2}^{i} q_j^{(t)}}}{\sum_{k=1}^{n} \frac{\prod_{j=1}^{k-1} p_j^{(t)}}{\prod_{j=2}^{k} q_j^{(t)}}}.$$

After some algebra to simplify the numerator and denominator, i.e. using the convention for the product notation that if $j > n$, then for a sequence of number $\{a_k\}_{k=1}^{n}$, $\prod_{k=j}^{n} a_k = 1$, we see that

$$\pi_t(i) = \frac{\left(\prod_{j=i+1}^{n} q_j^{(t)}\right)\left(\prod_{j=1}^{i-1} p_j^{(t)}\right)}{\sum_{k=1}^{n}\left(\prod_{j=k+1}^{n} q_j^{(t)}\right)\left(\prod_{j=1}^{k-1} p_j^{(t)}\right)}$$

for $1 \leq i \leq n$.

4.7.2 Proof of Proposition 24

Assuming that the column vector $\mathbf{x} \neq \mathbf{0}$ and $< \mathbf{x}, \mathbf{1} >_{\pi_t} = 0$, we see that for $t \in [0, 1]$

$$\begin{aligned}
\frac{< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t}}{< \mathbf{x}, \mathbf{x} >_{\pi_t} - (< \mathbf{x}, \mathbf{1} >_{\pi_t})^2} &= \frac{< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t}}{< \mathbf{x}, \mathbf{x} >_{\pi_t}} \\
&= \frac{< \mathbf{x}, \mathbf{x} >_{\pi_t} - < \mathbf{P_t}\mathbf{x}, \mathbf{x} >_{\pi_t}}{< \mathbf{x}, \mathbf{x} >_{\pi_t}} \\
&= 1 - \frac{< \mathbf{P_t}\mathbf{x}, \mathbf{x} >_{\pi_t}}{< \mathbf{x}, \mathbf{x} >_{\pi_t}}.
\end{aligned}$$

We see then that

$$1 - \inf\left\{\frac{<(\mathbf{I}-\mathbf{P_t})\mathbf{x},\mathbf{x}>_{\pi_t}}{<\mathbf{x},\mathbf{x}>_{\pi_t} - (<\mathbf{x},\mathbf{1}>_{\pi_t})^2}\big|\mathbf{x}\neq\mathbf{0} \text{ and } <\mathbf{x},\mathbf{1}>_{\pi_t}=0\right\}$$
$$= \sup\left\{\frac{<\mathbf{P_t}\mathbf{x},\mathbf{x}>_{\pi_t}}{<\mathbf{x},\mathbf{x}>_{\pi_t}}\big|\mathbf{x}\neq\mathbf{0} \text{ and } <\mathbf{x},\mathbf{1}>_{\pi_t}=0\right\}.$$

We know there exists scalars $c_2, \cdots, c_n$ such that $\mathbf{x} = c_2\mathbf{v}_{2,t} + \cdots + c_n\mathbf{v}_{n,t}$ where $\mathbf{v}_{2,t}, \cdots, \mathbf{v}_{n,t}$ are defined as a subset of the orthonormal basis introduced in Proposition 18. We see that this set is orthonormal with respect to $<\cdot,\cdot>_{\pi_t}$ as shown in Proposition 20 and for reversible matrices all eigenvalues are real. This implies that

$$\begin{aligned}
<\mathbf{P_t}\mathbf{x},\mathbf{x}>_{\pi_t} &= \lambda_2(t)c_2^2 + \cdots \lambda_n(t)c_n^2 \\
&\leq \lambda_2(t)c_2^2 + \cdots \lambda_2(t)c_n^2 \\
&\leq \lambda_2(t) <\mathbf{x},\mathbf{x}>_{\pi_t} .
\end{aligned}$$

This implies that

$$1 - \inf\left\{\frac{<(\mathbf{I}-\mathbf{P_t})\mathbf{x},\mathbf{x}>_{\pi_t}}{<\mathbf{x},\mathbf{x}>_{\pi_t} - (<\mathbf{x},\mathbf{1}>_{\pi_t})^2}\big|\mathbf{x}\neq\mathbf{0} \text{ and } <\mathbf{x},\mathbf{1}>_{\pi_t}=0\right\} \leq \lambda_2(t).$$

Noticing that when $\mathbf{x} = \mathbf{v}_{2,t}$,

$$\frac{<\mathbf{P_t}\mathbf{x},\mathbf{x}>_{\pi_t}}{<\mathbf{x},\mathbf{x}>_{\pi_t}} = \lambda_2(t)$$

we have that the supremum is reached.

We can conclude then that

$$\lambda_2(t) = 1 - \inf\left\{ \frac{< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t}}{< \mathbf{x}, \mathbf{x} >_{\pi_t} - (< \mathbf{x}, \mathbf{1} >_{\pi_t})^2} \middle| \mathbf{x} \neq \mathbf{0} \text{ and } < \mathbf{x}, \mathbf{1} >_{\pi_t} = 0 \right\}.$$

4.7.3 Proof of Proposition 25

To begin we look expand $< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t}$ where $\mathbf{x}$ is a column vector.

$$\begin{aligned}
< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t} &= \sum_{i=1}^{n} \left(\mathbf{x}(i) - \sum_{j=1}^{n} \mathbf{P_t}(i,j)\mathbf{x}(j) \right) \mathbf{x}(i)\pi_t(i) \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \mathbf{P_t}(i,j)\mathbf{x}(i) - \sum_{j=1}^{n} \mathbf{P_t}(i,j)\mathbf{x}(j) \right) \mathbf{x}(i)\pi_t(i) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t(i)\mathbf{P_t}(i,j)\mathbf{x}(i)\left(\mathbf{x}(i) - \mathbf{x}(j)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t(i)\mathbf{P_t}(i,j)\left(\mathbf{x}(i)^2 - \mathbf{x}(i)\mathbf{x}(j)\right).
\end{aligned}$$

By changing the index of the above equation, replacing i with j, we have that

$$< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t} = \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t(j)\mathbf{P_t}(j,i)\left(\mathbf{x}(j)^2 - \mathbf{x}(i)\mathbf{x}(j)\right).$$

Because this chain is reversible with respect to π_t for all $t \in [0,1]$, we have that $\pi_t(i)\mathbf{P_t}(i,j) = \pi_t(j)\mathbf{P_t}(j,i)$. Using this in the above equation, we have that

$$< (\mathbf{I} - \mathbf{P_t})\mathbf{x}, \mathbf{x} >_{\pi_t} = \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_t(i)\mathbf{P_t}(i,j)\left(\mathbf{x}(j)^2 - \mathbf{x}(i)\mathbf{x}(j)\right).$$

If we now add the right hand side of the above equation to

$\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_t(i)\mathbf{P_t}(i,j)\left(\mathbf{x}(i)^2-\mathbf{x}(i)\mathbf{x}(j)\right)$, we have that

$$2 < (\mathbf{I}-\mathbf{P_t})\mathbf{x},\mathbf{x} >_{\pi_t} = \sum_{i=1}^{n}\sum_{j=1}^{n}\pi_t(i)\mathbf{P_t}(i,j)\left(\mathbf{x}(j)^2-2\mathbf{x}(i)\mathbf{x}(j)+\mathbf{x}(i)^2\right).$$

Noticing that when $i=j$, the entry of the double-sum is 0, and reversibility guarantees that the (i,j)-entry of the double-sum is equivalent to the (j,i)-entry of the double-sum, if we divide both sides of the equation by 2 we have that

$$< (\mathbf{I}-\mathbf{P_t})\mathbf{x},\mathbf{x} >_{\pi_t} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\pi_t(i)\mathbf{P_t}(i,j)\left(\mathbf{x}(j)-\mathbf{x}(i)\right)^2.$$

4.7.4 Proof of Proposition 27

We see from Corollary 6 that for any $i,T\in\mathbb{N}$ with $i\le T$,

$$\Delta \le 1-\lambda_2(\frac{i}{T}) = \inf$$
$$\left\{\frac{\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\pi_{\frac{i}{T}}(j)\mathbf{P}_{\frac{\mathbf{i}}{\mathbf{T}}}(j,k)\left(\mathbf{x}(k)-\mathbf{x}(j)\right)^2}{\sum_{j=1}^{n}\pi_{\frac{i}{T}}(j)\mathbf{x}(j)^2}\Big|\mathbf{x}\neq\mathbf{0}\text{ and } <\mathbf{x},\mathbf{1}>_{\pi_{\frac{i}{T}}}=0\right\}.$$

Selecting

$$\mathbf{x}(j) = \frac{\pi_{\frac{i-1}{T}}(j)-\pi_{\frac{i}{T}}(j)}{\pi_{\frac{i}{T}}(j)}$$

and using the fact that for a birth-death chain

$$\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\pi_{\frac{i}{T}}(j)\mathbf{P}_{\frac{\mathbf{i}}{\mathbf{T}}}(j,k)\left(\mathbf{x}(j+1)-\mathbf{x}(j)\right)^2 = \sum_{j=1}^{n-1}\pi_{\frac{i}{T}}(j)p_j^{(\frac{i}{T})}\left(\mathbf{x}(k)-\mathbf{x}(j)\right)^2$$

we have that

$$\Delta \leq \frac{\sum_{j=1}^{n-1} \pi_{\frac{i}{T}}(j) p_j^{(\frac{i}{T})} \left(\frac{\pi_{\frac{i-1}{T}}(j+1) - \pi_{\frac{i}{T}}(j+1)}{\pi_{\frac{i}{T}}(j+1)} - \frac{\pi_{\frac{i-1}{T}}(j) - \pi_{\frac{i}{T}}(j)}{\pi_{\frac{i}{T}}(j)} \right)^2}{\sum_{j=1}^{n} \frac{\left(\pi_{\frac{i-1}{T}}(j) - \pi_{\frac{i}{T}}(j)\right)^2}{\pi_{\frac{i}{T}}(j)}}.$$

Noticing that the denominator of this equation is $\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|^2_{\frac{1}{\pi_{\frac{i}{T}}}}$, we can perform some algebra to show that

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{1}{\sqrt{\Delta}} \sqrt{\sum_{j=1}^{n-1} \frac{p_j^{(\frac{i}{T})} \left(\pi_{\frac{i-1}{T}}(j+1)\pi_{\frac{i}{T}}(j) - \pi_{\frac{i}{T}}(j+1)\pi_{\frac{i-1}{T}}(j)\right)^2}{\left(\pi_{\frac{i}{T}}(j+1)\right)^2 \pi_{\frac{i}{T}}(j)}}.$$

Because birth-death chains are reversible, we see that $\pi_t(j)p_j^{(t)} = \pi_t(j+1)q_{j+1}^{(t)}$. Using this to replace some of the terms in the above equation, we derive

$$\begin{aligned}
&\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&\quad \leq \frac{1}{\sqrt{\Delta}} \sqrt{\sum_{j=1}^{n-1} \frac{p_j^{(\frac{i}{T})} \left(\frac{p_j^{(\frac{i-1}{T})}}{q_{j+1}^{(\frac{i-1}{T})}} \pi_{\frac{i-1}{T}}(j)\pi_{\frac{i}{T}}(j) - \frac{p_j^{(\frac{i}{T})}}{q_{j+1}^{(\frac{i}{T})}} \pi_{\frac{i}{T}}(j)\pi_{\frac{i-1}{T}}(j) \right)^2}{\left(\frac{p_j^{(\frac{i}{T})}}{q_{j+1}^{(\frac{i}{T})}} \pi_{\frac{i}{T}}(j) \right)^2 \pi_{\frac{i}{T}}(j)}} \\
&\quad \leq \frac{1}{\sqrt{\Delta}} \sqrt{\sum_{j=1}^{n-1} \frac{\left(\pi_{\frac{i-1}{T}}(j)\right)^2 \left(p_j^{(\frac{i-1}{T})} q_{j+1}^{(\frac{i}{T})} - p_j^{(\frac{i}{T})} q_{j+1}^{(\frac{i-1}{T})}\right)^2}{\pi_{\frac{i}{T}}(j) p_j^{(\frac{i}{T})} \left(q_{j+1}^{(\frac{i-1}{T})}\right)^2}} \\
&\quad \leq \frac{1}{\sqrt{\Delta}} \sqrt{\sum_{j=1}^{n-1} \frac{\left(\pi_{\frac{i-1}{T}}(j)\right)^2 \left(\pi_{\frac{i-1}{T}}(j+1)\right)^2 \left(p_j^{(\frac{i-1}{T})} q_{j+1}^{(\frac{i}{T})} - p_j^{(\frac{i}{T})} q_{j+1}^{(\frac{i-1}{T})}\right)^2}{\pi_{\frac{i}{T}}(j) p_j^{(\frac{i}{T})} \left(\pi_{\frac{i-1}{T}}(j+1) q_{j+1}^{(\frac{i-1}{T})}\right)^2}}.
\end{aligned}$$

By using the formulae $p_j^{(\frac{i-1}{T})} = p_j^{(\frac{i}{T})} - \frac{1}{T}\left(p_j^{(1)} - p_j^{(0)}\right)$ and $q_{j+1}^{(\frac{i-1}{T})} = q_{j+1}^{(\frac{i}{T})} - \frac{1}{T}\left(q_{j+1}^{(1)} - q_{j+1}^{(0)}\right)$, performing some basic algebra and using the reversibility argument we can expand terms in the previous inequality to get the following:

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{1}{\sqrt{\Delta}T}\sqrt{\sum_{j=1}^{n-1} \frac{\left(\pi_{\frac{i-1}{T}}(j)\right)^2 \left(\pi_{\frac{i-1}{T}}(j+1)\right)^2 \left(p_j^{(0)}q_{j+1}^{(1)} - p_j^{(1)}q_{j+1}^{(0)}\right)^2}{\left(\pi_{\frac{i}{T}}(j)p_j^{(\frac{i}{T})}\right)^3}}.$$

If we now return our attention to our conductance bound in Proposition 26.

We divide the set E into two distinct sets. Define $B_j = \{1, \cdots, j\}$ and $B_j^C = \{j+1, \cdots, n\}$.

Using the notation introduced in Section 4.6, we see that for a birth-death chain

$$\pi_t(B_j) = \sum_{l=1}^{j} \pi_t(l)$$

$$F_t(B_j) = \pi_t(j)p_j^{(t)}$$

$$\psi_t(B_j, \mathbf{P_t}) = \frac{\pi_t(j)p_j^{(t)}}{\sum_{l=1}^{j} \pi_t(l)}.$$

We similarly see for a birth-death chain that

$$\pi_t(B_j^C) = \sum_{l=j+1}^{n} \pi_t(l)$$

$$F_t(B_j^C) = \pi_t(j+1)q_{j+1}^{(t)} = \pi_t(j)p_j^{(t)}$$

$$\psi_t(B_j^C, \mathbf{P_t}) = \frac{\pi_t(j)p_j^{(t)}}{\sum_{l=j+1}^{n} \pi_t(l)}.$$

Because either $\pi_t(B_j) \leq 1/2$ or $\pi_t(B_j^C) \leq 1/2$, or both simultaneously, we can see that

$$\phi_t(\mathbf{P_t}) \leq \frac{\pi_t(j)p_j^{(t)}}{\min\left\{\pi_t(B_j), \pi_t(B_j^C)\right\}}$$

for all $t \in [0,1]$ and $1 \leq j \leq n$.

Using the conductance bound from Proposition 26 we that for all $t \in [0,1]$ and $1 \leq j \leq n$

$$\frac{\min\left\{\pi_t(B_j), \pi_t(B_j^C)\right\}\Delta}{2} \leq \pi_t(j)p_j^{(t)}.$$

We now bound $1/\left(\pi_{\frac{i}{T}}p_j^{(\frac{i}{T})}\right)$ in our bound of $\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}}$. We see that

$$\begin{aligned}
&\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \\
&\quad\leq \frac{1}{\sqrt{\Delta}T}\sqrt{\sum_{j=1}^{n-1}\frac{8\left(\pi_{\frac{i-1}{T}}(j)\right)^2\left(\pi_{\frac{i-1}{T}}(j+1)\right)^2\left(p_j^{(0)}q_{j+1}^{(1)} - p_j^{(1)}q_{j+1}^{(0)}\right)^2}{\left(\min\left\{\pi_{\frac{i}{T}}(B_j), \pi_{\frac{i}{T}}(B_j^C)\right\}\Delta\right)^3}} \\
&\quad\leq \frac{\sqrt{8}}{\Delta^2 T}\sqrt{\sum_{j=1}^{n-1}\frac{\left(\pi_{\frac{i-1}{T}}(j)\right)^2\left(\pi_{\frac{i-1}{T}}(j+1)\right)^2\left(p_j^{(0)}q_{j+1}^{(1)} - p_j^{(1)}q_{j+1}^{(0)}\right)^2}{\left(\min\left\{\pi_{\frac{i}{T}}(B_j), \pi_{\frac{i}{T}}(B_j^C)\right\}\right)^3}}.
\end{aligned}$$

We know that $\pi_{\frac{i-1}{T}}(j) \leq 1$ for $1 \leq j \leq n$ and $|p_j^{(0)}q_{j+1}^{(1)} - p_j^{(1)}q_{j+1}^{(0)}| \leq 1$ for $1 \leq j \leq n-1$, so

$$\left(\pi_{\frac{i-1}{T}}(j)\right)^2\left(\pi_{\frac{i-1}{T}}(j+1)\right)^2\left(p_j^{(0)}q_{j+1}^{(1)} - p_j^{(1)}q_{j+1}^{(0)}\right)^2 \leq 1.$$

for all $1 \leq j \leq n-1$.

This now shows that

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{\sqrt{8}}{\Delta^2 T}\sqrt{\sum_{j=1}^{n-1}\frac{1}{\left(\min\left\{\pi_{\frac{i}{T}}(B_j), \pi_{\frac{i}{T}}(B_j^C)\right\}\right)^3}}.$$

Notice that $\pi_{\frac{i}{T}}(j) \leq \pi_{\frac{i}{T}}(B_j)$ and $\pi_{\frac{i}{T}}(j+1) \leq \pi_{\frac{i}{T}}(B_j^C)$, which implies that $\min\left\{\pi_{\frac{i}{T}}(j), \pi_{\frac{i}{T}}(j+1)\right\} \leq \min\left\{\pi_{\frac{i}{T}}(B_j), \pi_{\frac{i}{T}}(B_j^C)\right\}$. This implies that

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \leq \frac{\sqrt{8}}{\Delta^2 T}\sqrt{\sum_{j=1}^{n-1} \frac{1}{\left(\min\left\{\pi_{\frac{i}{T}}(j), \pi_{\frac{i}{T}}(j+1)\right\}\right)^3}}.$$

Because $\left(\prod_{l=k+1}^{n} q_l^{(t)}\right)\left(\prod_{l=1}^{k-1} p_l^{(t)}\right) \leq 1$ for $1 \leq k \leq n$, $t \in [0,1]$, we see that

$$\sum_{k=1}^{n}\left(\prod_{l=k+1}^{n} q_l^{(t)}\right)\left(\prod_{l=1}^{k-1} p_l^{(t)}\right) \leq n.$$

This tells us that

$$\frac{\delta^{n-1}}{n} \leq \frac{\left(\prod_{l=j+1}^{n} q_l^{(\frac{i}{T})}\right)\left(\prod_{l=1}^{j-1} p_l^{(\frac{i}{T})}\right)}{n} \leq \pi_{\frac{i}{T}}(j)$$

for all $1 \leq j \leq n$.

We now argue that

$$\begin{aligned}
\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} &\leq \frac{\sqrt{8}}{\Delta^2 T}\sqrt{\sum_{j=1}^{n-1} \frac{n^3}{\delta^{3(n-1)}}} \\
&\leq \frac{\sqrt{8}}{\Delta^2 T}\frac{n^2}{\sqrt{\delta^{3(n-1)}}}.
\end{aligned}$$

4.7.5 Proof of Corollary 7

If we consider Equation 4.15 from Lemma 2 and apply Proposition 14 we derive the following inequality:

$$\begin{aligned}\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \le \frac{1}{2}\sum_{j=0}^{k-1}\left[\prod_{i=k-j}^{k}\left|\lambda_2(\frac{i}{T})\right|\right]\|\pi_{\frac{k-j-1}{T}} - \pi_{\frac{k-j}{T}}\|_{\frac{1}{\pi_{\frac{k-j}{T}}}} \\ \left[\prod_{i=k-j}^{k-1}\left(1+\max_{1\le m\le n}\left\{\frac{1}{\sqrt{\pi_{\frac{i}{T}}(m)}}\right\}\|\pi_{\frac{i}{T}} - \pi_{\frac{i+1}{T}}\|_{\frac{1}{\pi_{\frac{i+1}{T}}}}\right)\right]\end{aligned}$$

where $\nu_{\frac{k}{T}} = \pi_0\mathbf{P}_{\frac{1}{T}}\cdots\mathbf{P}_{\frac{k}{T}}$ for a linear adiabatic evolution between $\mathbf{P}_0$ and $\mathbf{P}_1$ and $\lambda_2(t)$ is the second largest eigenvalue in modulus, however, we assumed that the second largest eigenvalue is the second largest eigenvalue in modulus

We also showed in Proposition 27

$$\|\pi_{\frac{i-1}{T}} - \pi_{\frac{i}{T}}\|_{\frac{1}{\pi_{\frac{i}{T}}}} \le \frac{\sqrt{8}}{\Delta^2 T}\frac{n^2}{\sqrt{\delta^{3(n-1)}}}.$$

Because $\left|\lambda_2(t)\right| \le 1-\Delta$ and $1/\sqrt{\pi_t(m)} \le \sqrt{n/\delta^{n-1}}$ for all $t \in [0,1]$ and $1 \le m \le n$, we see that

$$\begin{aligned}\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} &\le \frac{1}{2}\sum_{j=0}^{k-1}(1-\Delta)^{j+1}\,\frac{\sqrt{8}}{\Delta^2 T}\frac{n^2}{\sqrt{\delta^{3(n-1)}}}\left(1+\frac{\sqrt{8}}{\Delta^2 T}\frac{n^{5/2}}{\delta^{2(n-1)}}\right)^j \\ &\le \frac{\sqrt{2}}{\Delta^2 T}\frac{n^2}{\sqrt{\delta^{3(n-1)}}}\sum_{j=0}^{k-1}(1-\Delta)^j\left(1+\frac{\sqrt{8}}{\Delta^2 T}\frac{n^{5/2}}{\delta^{2(n-1)}}\right)^j.\end{aligned}$$

4.7.6 Proof of Theorem 16

We showed in Corollary 7 that

$$\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} \le \frac{\sqrt{2}}{\Delta^2 T}\frac{n^2}{\sqrt{\delta^{3(n-1)}}}\sum_{j=0}^{k-1}(1-\Delta)^j\left(1+\frac{\sqrt{8}}{\Delta^2 T}\frac{n^{5/2}}{\delta^{2(n-1)}}\right)^j.$$

We now let

$$T = \frac{\sqrt{8}n^{5/2}M}{\Delta^3\delta^{2(n-1)}}$$

where $M > 1$ is constant.

We can see that

$$\begin{aligned}\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} &\leq \frac{\Delta\sqrt{\delta^{(n-1)}}}{2\sqrt{n}M}\sum_{j=0}^{k-1}\left[(1-\Delta)\left(1+\frac{\Delta}{M}\right)\right]^j \\ &\leq \frac{\Delta\sqrt{\delta^{(n-1)}}}{2\sqrt{n}M}\sum_{j=0}^{k-1}\left[1-\frac{(M-1)\Delta}{M}-\frac{\Delta^2}{M}\right]^j.\end{aligned}$$

Noticing that this is a geometric series, we have that

$$\begin{aligned}\|\nu_{\frac{k}{T}} - \pi_{\frac{k}{T}}\|_{TV} &\leq \frac{\Delta\sqrt{\delta^{(n-1)}}}{2v\sqrt{n}M}\left(\frac{1-\left(1-\frac{(M-1)\Delta}{M}-\frac{\Delta^2}{M}\right)^k}{1-\left(1-\frac{(M-1)\Delta}{M}-\frac{\Delta^2}{M}\right)}\right) \\ &\leq \frac{\Delta\sqrt{\delta^{(n-1)}}}{2\sqrt{n}M}\left(\frac{1}{\frac{(M-1)\Delta}{M}+\frac{\Delta^2}{M}}\right) \\ &\leq \frac{\sqrt{\delta^{(n-1)}}}{2\sqrt{n}\left((M-1)+\Delta\right)}.\end{aligned}$$

We can see that for any $\epsilon > 0$ we can select an integer N not depending on ϵ such that $M = N/\epsilon > 1$ such that

$$\frac{\sqrt{\delta^{n-1}}}{2\sqrt{n}\left((M-1)+\Delta\right)} \leq \epsilon.$$

Because this is true for all $1 \leq k \leq T$, we see that

$$t_{sad}\left(\mathbf{P}_0, \mathbf{P}_1, \epsilon\right) \leq \frac{\sqrt{8}n^{5/2}N}{\epsilon\Delta^3\delta^{2(n-1)}}.$$

Letting $M^* = \left(\sqrt{8}n^{5/2}N\right)/\delta^{2(n-1)}$, we see that

$$t_{sad}\left(\mathbf{P}_0, \mathbf{P}_1, \epsilon\right) \leq \frac{M^*}{\epsilon\Delta^3}.$$

Chapter 5

Background on Continuous Markov Processes

We now turn our attention to continuous-time Markov chains. Our research group has written papers on two applications of the adiabatic time to continuous-time Markov chains. We will return to these applications in Chapters 7 and 8. Before we discuss these applications, we first must develop the necessary tools to discuss continuous-time Markov chains. Section 5.1 will introduce all the proper definitions of continuous-time Markov chains and will describe the sufficient conditions to have a unique stationary distribution for these Markov chains. Section 5.3 will define the mixing time for continuous-time Markov chains.

5.1 Continuous Markov Chains

This section considers the creation and development of continuous-time Markov chains. We will adopt notation from earlier chapters using discrete-time. For example, recall that discrete-time Markov chains have associated with them a sequence of stochastic matrices. In Section 2.5 we adopted the notation $\mathcal{P}_n$ to describe the space of $n \times n$ stochastic matrices. We now want to define continuous-time Markov chains over a finite state space $E = \{1, \cdots, n\}$.

Definition 26 *A continuous-time Markov chain is a random process in the space of functions* $\mathbf{X} = \{X : \mathbb{R}^+ \to E\}$*, where each function has a probability associated with it uniquely determined (up to initial distribution). This probability is governed by a multivariate function of stochastic matrices* $\mathbf{P} : \left\{(a,b) \in \mathbb{R}^+ \times \mathbb{R}^+ \middle| a \leq b\right\} \to \mathcal{P}_n$ *which give the conditional probability for times* $s, t \in \mathbb{R}^+$ *with* $s \leq t$*:* $\mathbb{P}(X(t) = j | X(s) = i) = p(s,t)_{ij}$ *where the* $p(s,t)_{ij}$ *is the* ij *entry of* $\mathbf{P}(s,t)$*.*

The matrices $\mathbf{P}(s,t)$ are called the probability transition matrices from time s to time t. As we did in discrete-time, we can classify Markov chains as time-homogeneous or time-inhomogeneous. We will first focus on time-homogeneous Markov chains so that we can define the stationary distribution of a continuous-time Markov chain, similar to the process outlined for discrete-time Markov chains.

Definition 27 *A continuous-time Markov chain is said to be time-homogeneous if* $p(s,t)_{ij} = p(0, t-s)_{ij}$ *for all* $i, j \in E$ *and* $(s,t) \in \left\{(a,b) \in \mathbb{R}^+ \times \mathbb{R}^+ \middle| a \leq b\right\}$*. We would simply write* $p(s,t)_{ij} = p(t-s)_{ij}$ *and* $\mathbf{P}(s,t) = \mathbf{P}(t-s)$*.*

A continuous-time Markov chain is said to be time-inhomogeneous if there exists $(s,t) \in \left\{(a,b) \in \mathbb{R}^+ \times \mathbb{R}^+ \middle| a \leq b\right\}$ *such that* $p(s,t)_{ij} \neq p(0, t-s)_{ij}$ *for some*

$i, j \in E$.

To better understand continuous-time, time-homogeneous Markov chains, we reference the following proposition, which is cited from [12]. Because the proof of this proposition is straightforward, I will include it in Section 5.2.

Proposition 28 *Consider the probability transition matrix function $\mathbf{P} : \mathbb{R}^+ \to E$ for a continuous-time, time-homogeneous Markov chain over a finite state space. Let $\{\mathbf{P}(t) | t \geq 0\}$ be the codomain of this function. Then*

(1) $\mathbf{P}(0) = \mathbb{I}$

(2) $\mathbf{P}(t)$ *is stochastic for all $t \geq 0$ and for $s, t \geq 0$*

(3) $\mathbf{P}(s + t) = \mathbf{P}(s)\mathbf{P}(t)$.

The codomain, $\{\mathbf{P}(t) | t \geq 0\}$, having the properties mentioned above is called a <u>stochastic semigroup</u>. The following definition outlines one property that we insist on the stochastic semigroups having throughout our study of continuous-time, time-homogeneous Markov chains.

Definition 28 *The semigroup $\{\mathbf{P}(t) | t \geq 0\}$ is called <u>standard</u> if $\mathbf{P}(t) \to \mathbb{I}$ as $t \to 0$.*

We also demand that our continuous-time, time-homogeneous Markov chains have the following property.

Definition 29 *The semigroup $\{\mathbf{P}(t) : t \geq 0\}$ for a continuous-time, time-homogeneous Markov chain is called <u>irreducible</u> if for all $(i, j) \in E \times E$ we have $p(s)_{ij} > 0$ for some $s \geq 0$.*

The goal in requiring continuous-time, time-homogeneous Markov chains to have standard, irreducible semigroups is to guarantee the existence of a unique stationary distribution, however, the definition of stationary distribution must also change in the continuous setting.

Definition 30 *For a continuous-time, time-homogeneous Markov chain with stochastic semigroup* $\{\mathbf{P}(t) : t \geq 0\}$*, any left-handed eigenvector associated with the eigenvalue* 1 *is called a stationary distribution, denoted* π*, only if* π_j *is a real number and* $\pi_j \geq 0$ *for* $1 \leq j \leq n$*. In particular,* $\pi\mathbf{P}(t) = \pi$ *for all* $t \geq 0$*.*

The following theorem explains the conditions necessary to make existence of a stationary distribution imply uniqueness of the stationary distribution. A sketch of the proof was given in [12].

Theorem 17 *For a continuous-time, time-homogeneous Markov chain with standard, irreducible semigroup* $\{\mathbf{P}(t) : t \geq 0\}$*,*

(a) *if there exists a stationary distribution* π *then it is unique and*

$$p(t)_{ij} \to \pi_j \text{ as } t \to 0 \text{ for all } i, j \in E$$

(b) *if there is no stationary distribution, then* $p(t)_{ij} \to 0$ *as* $t \to 0$ *for all* $i, j \in E$*.*

We know that the semigroup $\{\mathbf{P}(t) : t \geq 0\}$ being standard implies that $p(t)_{ij}$ is continuous for all $t \geq 0$ and $i, j \in E$. If we suppose that the Markov chain is in state i at time t, then for a small amount of time h, we either have that the Markov chain remains in state i with probability $p(h)_{ii} + o(h)$ or moves to state $j \neq i$ with probability $p(h)_{ij} + o(h)$. For h small we can assume that the Markov chain doesn't travel to multiple states within the time interval $(t, t+h)$. It has been shown that $p(h)_{ij}$ is approximately linear in h when h is small. This would imply that $\mathbf{P}(t)$ is differentiable for all $t \geq 0$ and $\mathbf{P}'(0)$ is a constant matrix. The following definition gives a name to this matrix.

Definition 31 *The* <u>*generator matrix,*</u> $\mathbf{Q}$ *of a continuous-time, time homogeneous Markov chain with probability transition matrix* $\mathbf{P}(t)$ *for* $t \geq 0$ *is a constant matrix that satisfies the Kolmogorov Backward equations*

$$\mathbf{P'}(t) = \mathbf{QP}(t).$$

Because E is a finite, we see that solving this matrix differential equation gives us that $\mathbf{P}(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!}\mathbf{Q}^j$, and because $\mathbf{P}(t)$ is stochastic for all $t \geq 0$ we see that $\mathbf{Q1} = \mathbf{0}$.

For a continuous time Markov chain $\{X(t)\}_{t\geq 0}$ on a finite state space E with a bounded generator matrix $\mathbf{Q} = (q_{ij})_{i,j\in E}$ and $\lambda = \max_{i\in E} \sum_{j\neq i} q_{ij}$, the upper bound on the departure rates of all states, a process called <u>uniformization</u> [12] gives transition probabilities to be

$$\mathbf{P}(t) = \sum_{j=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} \mathbf{P}^j = e^{\mathbf{Q}t}, \tag{5.1}$$

where the matrix $\mathbf{P} = \mathbb{I} + \frac{1}{\lambda}\mathbf{Q}$. The matrix $\mathbf{P}(t)$ denotes the transition probabilities at time t.

Notice that $\mathbf{P} = \mathbb{I} + \frac{1}{\lambda}\mathbf{Q}$ is a stochastic matrix. If π is a stationary distribution of $\mathbf{P}$ as a discrete probability transition matrix, then π is the stationary distribution of $\mathbf{P}(t)$ as a continuous probability transition matrix. Theorem 17 would guarantee that this stationary distribution is unique. The existence of a unique stationary distribution will lead to and analogue concept of the mixing time introduced in Section 1.3. We will outline the mixing time in Section 5.3.

5.2 Proofs

5.2.1 Proof of Proposition 28

First I show that $\mathbf{P}(0) = \mathbb{I}$. This is straightforward.

Note that $p(0)_{ij} = \mathbb{P}\left(X(0) = j | X(0) = i\right)$. If $i \neq j$, then this probability is by definition 0, and if $i = j$ this probability is by definition 1.

This implies that $\mathbf{P}(0) = \mathbb{I}$.

Next I show that $\mathbf{P}(t)$ is stochastic.

$$
\begin{aligned}
\left(\mathbf{P}(t)\mathbf{1}\right)_i &= \sum_{j=1}^{n} p(t)_{ij} \\
&= \mathbb{P}\left(\cup_{j=1}^{n} \{X(t) = j\} | X(0) = i\right) \\
&= \sum_{j=1}^{n} \mathbb{P}\left(X(t) = j | X(0) = i\right) \\
&= 1.
\end{aligned}
$$

This implies that $\mathbf{P}(t)\mathbf{1} = \mathbf{1}$.

Finally I show that $\mathbf{P}(s+t) = \mathbf{P}(s)\mathbf{P}(t)$.

$$
\begin{aligned}
p(s+t)_{ij} &= \mathbb{P}\left(X(s+t)=j|X(0)=i\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X(s+t)=j|X(s)=k, X(0)=i\right)\mathbb{P}\left(X(s)=k|X(0)=i\right)
\end{aligned}
$$

Using the Markov property of Markov chains and the time homogeneity of the chain, we see that

$$
\begin{aligned}
p(s+t)_{ij} &= \sum_{k=1}^{n} \mathbb{P}\left(X(s+t)=j|X(s)=k\right)\mathbb{P}\left(X(s)=k|X(0)=i\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(X(t)=j|X(0)=k\right)\mathbb{P}\left(X(s)=k|X(0)=i\right) \\
&= \sum_{k=1}^{n} p(t)_{kj}p(s)_{ik}.
\end{aligned}
$$

We see now that $\mathbf{P}(s+t) = \mathbf{P}(s)\mathbf{P}(t)$.

5.3 Mixing Time for Continuous Markov Chains

If you recall Definition 9 from Section 1.3, the mixing time was defined to answer how the structure of the irreducible and aperiodic matrix affects how quickly the Markov chain converges to its stationary distribution. In Section 5.1 we discussed analogue sufficient conditions for a continuous-time, time-homogeneous Markov chain to converge to its unique stationary distribution. In this section we are going to define the mixing time to describe how quickly the continuous-time Markov chains described in Section 5.1 converges to their stationary distributions in the total variation norm.

Definition 32 *Let $\{\mathbf{P}(t) | t \geq 0\}$ be a standard, irreducible semigroup for a continuous-time, time-homogeneous Markov chain with stationary distribution π. Given an $\epsilon > 0$, the time $t_{mix}(\mathbf{P}, \epsilon)$ is called the mixing time if it is the infimum over $t \geq 0$ of*

$$\max_{\nu} \|\nu \mathbf{P}(t) - \pi\|_{TV} \leq \epsilon \tag{5.2}$$

where the maximum is take over all probability distributions ν.

In Chapter 6 we will define an analogue of the adiabatic time for continuous-time Markov chains and we will asymptotically bound the adiabatic time as $\epsilon \to 0$ by an inverse power of ϵ multiplied by a power of the mixing time. We could analogously try to bound the adiabatic time by the inverse spectral gap as in Theorems 2 and 3, however, as we have discussed in Chapter 4, adiabatic times and mixing times make a more natural comparison.

Chapter 6

The Adiabatic Time Versus The Mixing Time For Continuous Markov Chains

We now turn our attention to defining the adiabatic time for continuous-time Markov chains and we will seek to asymptotically bound the adiabatic time as $\epsilon \to 0$ as an inverse power of ϵ multiplied by a power of the mixing time. Section 6.1 will do this for a linear evolution between two bounded generators and Section 6.3 will do this for a general evolution between two bounded generators. We will provide examples in both sections to claim that these asymptotic bounds are optimal.

6.1 Linear Evolution

We start by defining a linear evolution between the generator matrices of two continuous-time, time-homogeneous Markov chains. This will help us define the continuous-time, time-inhomogeneous Markov chains necessary for the study of the adiabatic time. We introduced the following definition in [16].

Definition 33 *Let* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *be two bounded generators for continuous-time, time-homogeneous Markov chains with standard, irreducible semigroups* $\{\mathbf{P}[\mathbf{0}](s)|s \geq 0\}$ *and* $\{\mathbf{P}[\mathbf{1}](s)|s \geq 0\}$ *associated with* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *respectively. We call* $\mathbf{Q}[0]$ *the initial generator matrix and* $\mathbf{Q}[1]$ *the final generator matrix. We define a class of generator matrices based on a* <u>*linear evolution*</u> *between* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *to be* $\{\mathbf{Q}[t]\}_{t\in[0,1]}$ *so that*

$$\mathbf{Q}[t] = (1-t)\mathbf{Q}[0] + t\mathbf{Q}[1] \tag{6.1}$$

for each $t \in [0,1]$.

For $t \in [0,1]$, if we let $\{\mathbf{P}[\mathbf{t}](s)|s \geq 0\}$ be the standard, irreducible semigroup associated with the generator $\mathbf{Q}[t]$, then we define π_t to be the stationary distribution $\mathbf{P}[\mathbf{t}](s)$ for $s \geq 0$.

We use this linear evolution between $\mathbf{Q}[0]$ and $\mathbf{Q}[1]$ to define a special class of time-inhomogeneous Markov chains. Given $T > 0$ and $t_1, t_2 \geq 0$ such that $0 \leq t_1 \leq t_2 \leq T$, let $\mathbf{P}_T(t_1, t_2)$ denote a matrix of transition probabilities of a continuous-time, time-inhomogeneous Markov chain generated by $\mathbf{Q}[\frac{t}{T}]$ over the time interval $[t_1, t_2]$. If we select $t_1 = 0$ and $t_2 = T$, we would have the following differential equation on the interval $[0, T]$ describing our time-inhomogeneous Markov chain

$$\frac{d\mathbf{P}_T(0,t)}{dt} = \mathbf{Q}\left[\frac{t}{T}\right]\mathbf{P}_T(0,t). \tag{6.2}$$

With this time-dependent generator we define the adiabatic time to be the first time that the continuous-time Markov chain generated by $\mathbf{Q}[\frac{t}{T}]$ reaches a a resulting distribution that is 'close enough' to the stationary distribution of the continuous-time, time-homogeneous Markov chain governed by the generator $\mathbf{Q}[1]$.

Definition 34 *Given a linear adiabatic evolution between the bounded generators $\mathbf{Q}[0]$ and $\mathbf{Q}[1]$, we let $\mathbf{P}_T(0,T)$ denote a matrix of transition probabilities of a continuous-time, time-inhomogeneous Markov chain generated by a linear adiabatic evolution $\mathbf{Q}[\frac{t}{T}]$ over the time interval $[0,T]$. Given $\epsilon > 0$, a time $t_{ad}\left(\mathbf{P[0]}, \mathbf{P[1]}, \epsilon\right)$ is called the <u>adiabatic time</u> if it is the infimum of $T > 0$ such that*

$$\max_{\nu} \|\nu \mathbf{P}_T(0,T) - \pi_1\|_{TV} \leq \epsilon \tag{6.3}$$

where the maximum is taken over all probability distributions ν.

We now find a bound of the adiabatic time with respect to the mixing time of the time-homogeneous Markov chain governed by $\mathbf{Q}[1]$. We will see that this result agrees with the asymptotic bound found in Corollary 1 as $\epsilon \to 0$. We now state the following result from [16]. We will include a proof of this theorem in Section 6.2.

Theorem 18 *Take λ such that*

$$\lambda \geq \max_{i \in E} \sum_{j : j \neq i} \mathbf{Q}[0](i,j)$$

and

$$\lambda \geq \max_{i \in E} \sum_{j : j \neq i} \mathbf{Q}[1](i,j).$$

Then the adiabatic time

$$t_{ad}\left(\mathbf{P}[\mathbf{0}], \mathbf{P}[\mathbf{1}], \epsilon\right) \leq \frac{\lambda t_{mix}^{2}(\mathbf{P}[\mathbf{1}], \epsilon/2)}{\epsilon} + \theta \tag{6.4}$$

where $\theta = t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2) + \epsilon/(4\lambda)$.

The following example provides generators $\mathbf{Q}[0]$ and $\mathbf{Q}[1]$ that show the result from Theorem 18 are optimal.

Example 3 *(The lower bound.)*

$$\mathbf{Q}[0] = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & -1 \end{pmatrix}$$

and

$$\mathbf{Q}[1] = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

In Section 6.3 we will expand the types of time-inhomogeneous Markov chains we consider by defining a more general type of adiabatic evolution and we will find a similar result.

6.2 Proofs

6.2.1 Proof of Theorem 18

Observe that $\lambda \geq \max_{i \in E} \sum_{j:j \neq i} \mathbf{Q}\left[\frac{t}{T}\right](i,j)$ for $0 \leq t \leq T$.

Take

$$T = K\left(1 - \frac{1}{2K}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2) \text{ and}$$
$$N = (K-1)\left(1 - \frac{1}{2K}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2),$$

and let $\mathbf{P}[\mathbf{1}] = e^{t\mathbf{Q}[1]}$ denote the probability transtion matrix associated with the generator $\mathbf{Q}[1]$.

Now, we let $\mathbf{P}_0 = \mathbb{I} + \frac{1}{\lambda}\mathbf{Q}[0]$ and $\mathbf{P}_1 = \mathbb{I} + \frac{1}{\lambda}\mathbf{Q}[1]$. Note that $\mathbf{P}_0$ and $\mathbf{P}_1$ are the probability transition matrices for time-homogeneous, discrete-time Markov chains and

$$\nu \mathbf{P}_T(0,T) = \nu_{\mathbf{N}} \mathbf{P}_T(N,T) = \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} \frac{(\lambda(T-N))^j}{j!} e^{-\lambda(T-N)} \mathbf{I}_j \right),$$

where $\nu_{\mathbf{N}} = \nu \mathbf{P}_T(0,N)$ and

$$\mathbf{I}_j = \frac{j!}{(T-N)^j} \int \cdots \int_{N<s_1<s_2<\cdots<s_n<T} \left[\left(1 - \frac{s_1}{T}\right)\mathbf{P}_0 + \frac{s_1}{T}\mathbf{P}_1\right] \cdots \left[\left(1 - \frac{s_j}{T}\right)\mathbf{P}_0 + \frac{s_j}{T}\mathbf{P}_1\right] ds_1 \cdots ds_j$$

i.e. the order statistics of j arrivals within the $[N,T]$ time interval. We used the fact that, when condition on the number of arrivals, the arrival times of a Poisson process are distributed as an order statistic of uniform random variables.

Hence

$$
\begin{aligned}
&\nu \mathbf{P}_T\,(0,T) \\
&= \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} (\lambda(T-N)))^j \frac{e^{-\lambda(T-N)}}{(T-N)^j T^j j!} \mathbf{P}_1^j \int_N^T \cdots \int_N^T s_1 \cdots s_j ds_1 \cdots ds_j \right) \\
&\quad\quad + \mathcal{E} \\
&= e^{-\lambda\left(1-\frac{1}{2k}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} \frac{\lambda^j t_{mix}^j(\mathbf{P}[\mathbf{1}],\epsilon/2)}{j!} \mathbf{P}_1^j \right) + \mathcal{E} \\
&= e^{-\frac{\lambda t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)}{2K-1}} \nu_{\mathbf{N}} \mathbf{P}[\mathbf{1}]\,(t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)) + \mathcal{E}
\end{aligned}
$$

where $\mathcal{E}$ is the rest of the terms. Thus, the total variation distance,

$$
\max_{\nu} \|\nu \mathbf{P}_T\,(0,T) - \pi_1\|_{TV} \leq e^{-\frac{\lambda t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)}{2K-1}} \epsilon/2 + S_N.
$$

Taking $K \geq \frac{\lambda t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)}{\epsilon} + 1/2$, we bound the error term

$$
\begin{aligned}
S_N &= \|\mathcal{E} - \pi_1\|_{TV} \\
&\leq 1 - e^{-\lambda\left(1-\frac{1}{2K}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \sum_{j=0}^{\infty} \frac{\lambda^j t_{mix}^j(\mathbf{P}[\mathbf{1}],\epsilon/2)}{j!} \\
&= 1 - e^{-\frac{\lambda t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)}{2K-1}} \\
&\leq \epsilon/2
\end{aligned}
$$

as $\epsilon < -2\log\left(1 - \frac{\epsilon}{2}\right)$. Therefore

$$
t_{ad}\,(\mathbf{P}[\mathbf{0}],\mathbf{P}[\mathbf{1}],\epsilon) \leq \left(\frac{\lambda t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)}{\epsilon} + 1/2 \right) (t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2) + \epsilon/(2\lambda))\,.
$$

6.3 General Evolution

We now extend our results to more general adiabatic evolutions between the generator matrices of two continuous-time, time-homogeneous Markov chains. We introduced the following definition in [4].

Definition 35 *Let* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *be two bounded generators for continuous-time, time-homogeneous Markov chains with standard, irreducible semigroups* $\{\mathbf{P}[\mathbf{0}](s)|s \geq 0\}$ *and* $\{\mathbf{P}[\mathbf{1}](s)|s \geq 0\}$ *associated with* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *respectively. We call* $\mathbf{Q}[0]$ *the initial generator matrix and* $\mathbf{Q}[1]$ *the final generator matrix. We define a class of generator matrices based on a* <u>*general evolution*</u> *between* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *to be* $\{\mathbf{Q}[t]\}_{t\in[0,1]}$ *so that*

$$\mathbf{Q}[t](i,j) = (1 - \phi_{i,j}(t))\mathbf{Q}[0](i,j) + \phi_{i,j}(t)\mathbf{Q}[1](i,j) \tag{6.5}$$

where $\phi_{i,j} : [0,1] \to [0,1]$ *are continuous functions such that* $\phi_{i,j}(0) = 0$ *and* $\phi_{i,j}(1) = 1$ *for all* $1 \leq i,j \leq n$ *and* $\sum_j \phi_{i,j}(t)\left(\mathbf{Q}[1](i,j) - \mathbf{Q}[0](i,j)\right) = 0$ *for all* $t \in [0,1]$ *and each* $1 \leq i \leq n$.

We use the general evolution between $\mathbf{Q}[0]$ and $\mathbf{Q}[1]$ to define a special class of time-inhomogeneous Markov chains. Given $T > 0$, let $\mathbf{P}_T(0,T)$ denote a matrix of transition probabilities of a continuous-time, time-inhomogeneous Markov chain generated by $\mathbf{Q}[\frac{t}{T}]$ over the time interval $[0,T]$. We again have the following differential equation on the interval $[0,T]$ describing our time-inhomogeneous Markov chain

$$\frac{d\mathbf{P}_T(0,t)}{dt} = \mathbf{Q}\left[\frac{t}{T}\right]\mathbf{P}_T(0,t).$$

We now make the analogous definition of the adiabatic time of the time-inhomogeneous, continuous time Markov chain.

Definition 36 *Given a general adiabatic evolution between the bounded generators* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$*, we let* $\mathbf{P}_T(0,T)$ *denote a matrix of transition probabilities of a continuous-time, time-inhomogeneous Markov chain generated by a general adiabatic evolution* $\mathbf{Q}[\frac{t}{T}]$ *over the time interval* $[0,T]$*. Given* $\epsilon > 0$*, a time* $t_{ad}(\mathbf{P}[\mathbf{0}], \mathbf{P}[\mathbf{1}], \epsilon)$ *is called the* <u>*adiabatic time*</u> *if it is the infimum of* $T > 0$ *such that*

$$\max_{\nu} \|\nu \mathbf{P}_T(0,T) - \pi_1\|_{TV} \leq \epsilon$$

where the maximum is taken over all probability distributions ν*.*

As we did before, we find a bound of the adiabatic time with respect to the mixing time of the time-homogeneous Markov chain governed by $\mathbf{Q}[1]$. We will see that this result agrees with the asymptotic bound found in Theorem 7 as $\epsilon \to 0$. We state the following result from [4]. We will include a proof of this theorem in Section 6.4.

Theorem 19 *Suppose* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *are bounded generators for continuous-time, time-homogeneous Markov chains with standard, irreducible semigroups* $\{\mathbf{P}[\mathbf{0}](s) | s \geq 0\}$ *and* $\{\mathbf{P}[\mathbf{1}](s) | s \geq 0\}$ *associated with* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$ *respectively. Consider a general adiabatic evolution between* $\mathbf{Q}[0]$ *and* $\mathbf{Q}[1]$*. Let* $\epsilon > 0$*. Let* ϕ *be the piecewise minimum function of the* $\phi_{i,j}$ *functions, if* m *is the first positive integer such that* $\phi^{(m)}(1) \neq 0$*. If we take* λ *such that*

$$\lambda \geq \max_{i \in E} \sum_{j : j \neq i} \mathbf{Q}[0](i,j) \text{ and } \lambda \geq \max_{i \in E} \sum_{j : j \neq i} \mathbf{Q}[1](i,j)$$

then

$$t_{ad}(\mathbf{P}[\mathbf{0}], \mathbf{P}[\mathbf{1}], \epsilon) = \mathcal{O}\left(\left[\frac{\lambda}{\epsilon} \right]^{\frac{1}{m}} t_{mix}^{\frac{m+1}{m}} (\mathbf{P}[1], \epsilon/2) \right) \tag{6.6}$$

To show that this bound is optimal, you can reference the generators introduced in Example 3. We can also check that the result in Theorem 19 is scale invariant as we did in [4]. For a positive value M, we scale the initial and final generators to $\frac{1}{M}\mathbf{Q}[0]$ and $\frac{1}{M}\mathbf{Q}[1]$ respectively. Then the adiabatic evolution is slowed down M times, and the new adiabatic time should be of order $M\left[\frac{\lambda}{\epsilon}\right]^{\frac{1}{m}} t_{mix}^{\frac{m+1}{m}}(\mathbf{P}[1], \epsilon/2)$ with the mixing time and λ taken before the scaling. On the other hand, the new mixing time will be $Mt_{mix}(\mathbf{P}[1], \epsilon/2)$ and the new λ is $\frac{\lambda}{M}$ as the rates are M times lower. Plugging the new parameters into the theorem, we obtain

$$\left[\frac{\lambda}{M\epsilon}\right]^{\frac{1}{m}} \left(Mt_{mix}(\mathbf{P}[1], \epsilon/2)\right)^{\frac{m+1}{m}} = M\left[\frac{\lambda}{\epsilon}\right]^{\frac{1}{m}} t_{mix}^{\frac{m+1}{m}}(\mathbf{P}[1], \epsilon/2)$$

confirming that the theorem is invariant under time scaling.

6.4 Proofs

6.4.1 Proof of Theorem 19

Define $\hat{\mathbf{Q}}$ to be a Markov generator with off-diagonal entries

$$\hat{\mathbf{Q}}(i,j) = \frac{1-\phi_{i,j}(t)}{1-\phi(t)}\mathbf{Q}[0](i,j) + \frac{\phi_{i,j}(t)-\phi(t)}{1-\phi(t)}\mathbf{Q}[1](i,j).$$

Then writing

$$\mathbf{Q}[t](i,j) = (1-\phi_{i,j}(t))\mathbf{Q}[0](i,j) + (\phi_{i,j}(t)-\phi(t))\mathbf{Q}[1](i,j) + \phi(t)\mathbf{Q}[1](i,j)$$

would imply that

$$\mathbf{Q}[t] = (1-\phi(t))\hat{\mathbf{Q}} + \phi(t)\mathbf{Q}[1].$$

Observe that

$$\lambda \geq \max_{i \in E} \sum_{j:j \neq i} \hat{\mathbf{Q}}(i,j) \text{ and } \lambda \geq \max_{i \in E} \sum_{j:j \neq i} \mathbf{Q}\left[\frac{t}{T}\right](i,j)$$

as

$$\lambda \geq \max_{i \in E} \sum_{j:j \neq i} \mathbf{Q}[0](i,j) \text{ and } \lambda \geq \max_{i \in E} \sum_{j:j \neq i} \mathbf{Q}[1](i,j).$$

Let $\mathbf{P}[\mathbf{1}](t) = e^{t\mathbf{Q}[1]}$ denote the continuous-time, probability transition matrix associated with the generator $\mathbf{Q}[1]$ and let $\mathbf{P}_0 = \mathbb{I} + \frac{1}{\lambda}\hat{\mathbf{Q}}$ and $\mathbf{P}_1 = \mathbb{I} + \frac{1}{\lambda}\mathbf{Q}[1]$.

The probability transition matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ are for time-homogeneous, discrete-time, Markov chains. Conditioning on the number of arrivals within the $[N,T]$ time interval

$$\nu\mathbf{P}_T(0,T) = \nu_{\mathbf{N}}\mathbf{P}_T(N,T) = \nu_{\mathbf{N}}\left(\sum_{j=0}^{\infty} \frac{(\lambda(T-N))^j}{j!} e^{-\lambda(T-N)}\mathbf{I}_j\right)$$

where $\nu_{\mathbf{N}} = \nu\mathbf{P}_T(0,N)$ and

$$\mathbf{I}_j = \frac{j!}{(T-N)^j}\int \cdots \int_{N<s_1<\cdots<s_j<T} \left[(1-\phi(\frac{s_1}{T}))\mathbf{P}_0 + \phi(\frac{s_1}{T})\mathbf{P}_1\right] \cdots \left[(1-\phi(\frac{s_j}{T}))\mathbf{P}_0 + \phi(\frac{s_j}{T})\mathbf{P}_1\right] ds_1 \cdots ds_j$$

i.e. the order statistics of j arrivals within the $[N,T]$ time interval.

Therefore, combining the terms with $\mathbf{P}[\mathbf{1}]$, we obtain

$$\begin{aligned}
&\nu \mathbf{P}_T(0,T) \\
&= \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} \frac{\lambda^j \mathbf{P}[\mathbf{1}]^j}{j!} e^{-\lambda(T-N)} \int_N^T \cdots \int_N^T \phi(\frac{s_1}{T}) \cdots \phi(\frac{s_j}{T}) ds_1 \cdots ds_j \right) + \mathcal{E} \\
&= e^{-\lambda(T-N)} \nu_N \left(\sum_{j=0}^{\infty} \frac{\lambda^j T^j}{j!} \mathbf{P}[\mathbf{1}]^j \left(\int_{\frac{N}{T}}^1 \phi(x) dx \right)^j \right) + \mathcal{E}
\end{aligned}$$

where $\mathcal{E}$ is the rest of the terms.

Take $K > 0$ and define

$$T = \left(\int_{\frac{K-1}{K}}^1 \phi(x) dx \right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2)$$

and

$$N = \frac{K-1}{K} \left(\int_{\frac{K-1}{K}}^1 \phi(x) dx \right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2).$$

Recall the approximation of the minimum function ϕ around $x = 1$

$$\phi(x) = 1 + \frac{\phi^{(m)}(1)(x-1)^m}{m!} + \mathcal{O}\left(\left| x - 1 \right|^{m+1} \right)$$

and therefore

$$\int_{\frac{K-1}{K}}^1 \phi(x) dx = \frac{1}{K} \left(1 + \frac{\gamma(K)}{K^m} \right)$$

where $\gamma(K) = (-1)^m \frac{\phi^{(m)}(1)}{(m+1)!} + \mathcal{O}\left(K^{-1}\right)$. Thus we can write

$$\begin{aligned}
&\nu \mathbf{P}_T(0,T) \\
&= e^{-\lambda(T-N)} \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} \frac{\lambda^j (T-N)^j}{j!} \mathbf{P}[\mathbf{1}]^j \left[1 + \gamma(K) \left(\frac{T-N}{T} \right)^m \right]^j \right) + \mathcal{E}.
\end{aligned}$$

We see, using a standard uniformization argument, that

$$\begin{aligned}
&\nu P_T(0,T) \\
&= e^{-\lambda\left(1+\frac{\gamma(K)}{K^m}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \nu_{\mathbf{N}} \left(\sum_{j=0}^{\infty} \frac{\lambda^j t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)^j}{j!} \mathbf{P}[\mathbf{1}]^j \right) + \mathcal{E} \\
&= e^{-\lambda\left(\frac{\gamma(K)}{K^m+\gamma(K)}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \nu_{\mathbf{N}} \exp\{\mathbf{Q}[1] t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)\} + \mathcal{E}.
\end{aligned}$$

Now, since $(-1)^m \phi^{(m)}(1) \leq 0$, we have that, for any probability distribution ν,

$$\begin{aligned}
&\|\nu \mathbf{P}_T(0,T) - \pi_1\|_{TV} \\
&= e^{-\lambda\left(\frac{\gamma(K)}{K^m+\gamma(K)}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \|\nu \exp\{\mathbf{Q}[1] t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)\} - \pi_1\|_{TV} + S_N
\end{aligned}$$

where, by the triangle inequality

$$0 \leq S_N \leq 1 - e^{-\lambda\left(\frac{\gamma(K)}{K^m+\gamma(K)}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \left(\sum_{j=0}^{\infty} \frac{\lambda^j (t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2))^j}{j!} \right)$$

and, by definition of $t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)$,

$$e^{-\lambda\left(\frac{\gamma(K)}{K^m+\gamma(K)}\right)^{-1} t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} \|\nu \exp\{\mathbf{Q}[1] t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)\} - \pi_1\|_{TV} \leq \epsilon/2.$$

Taking $K = c(\lambda/\epsilon)^{\frac{1}{m}} t_{mix}^{\frac{1}{m}}(\mathbf{P}[\mathbf{1}],\epsilon/2)$ with constant $c >> (-1)^m \frac{\phi^{(m)}(1)}{(m+1)!}$, we obtain

$$\epsilon > -\log(1-\epsilon/2) \geq \lambda \left(\frac{-\gamma(K)}{K^m + \gamma(K)} \right) t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2)$$

and therefore

$$0 \leq S_N \leq 1 - e^{\lambda\left(\frac{\gamma(K)}{K^m+\gamma(K)}\right) t_{mix}(\mathbf{P}[\mathbf{1}],\epsilon/2)} < \epsilon/2.$$

Thus

$$t_{ad}\left(\mathbf{P}[\mathbf{0}], \mathbf{P}[\mathbf{1}], \epsilon\right) = \frac{K t_{mix}(\mathbf{P}[\mathbf{1}], \epsilon/2)}{1 + \frac{\gamma(K)}{K^m}} = \mathcal{O}\left(\left[\frac{\lambda}{\epsilon} \right]^{\frac{1}{m}} t_{mix}^{\frac{m+1}{m}} \left(\mathbf{P}[\mathbf{1}], \epsilon/2\right) \right).$$

Chapter 7

Applications to Queueing

In this chapter, we summarize the results from [27]. In Section 7.1 we develop some interesting facts about continuous-time Markov chains. You can think of this as a combination of the work on continuous-time Markov chains outlined in Chapters 5 and 6 with the work on reversible and birth-death processes from Chapter 4; however, this chapter does not consider an adiabatic evolution through a continuous function, rather, this chapter considers an adiabatic evolution through a step-function. This being the case we modify some of our definitions and objectives. In Section 7.2 we find an application of this kind of adiabatic evolution to a queue and a special queueing policy.

7.1 Background

We begin this section with a review of some facts about the $\|\cdot\|_{\frac{1}{\pi}}$-norm introduced in Chapter 4. For a reversible, irreducible $(K+1)\times(K+1)$ matrix $\mathbf{P}$, we will now use the notation $\mathbf{u}_1, \cdots, \mathbf{u}_{K+1}$ as the orthonormal basis with respect to $< \cdot, \cdot >_{\frac{1}{\pi}}$, where $\mathbf{u}_i$ was associated with the real eigenvalue $\lambda_i(\mathbf{P})$ of $\mathbf{P}$ and we ordered them such that $1 = |\lambda_1(\mathbf{P})| > |\lambda_2(\mathbf{P})| \geq \cdots \geq |\lambda_{K+1}(\mathbf{P})|$.

The following result is used to bound the distance of the continuous chain in terms of a discrete one. It was given in [27].

Proposition 29 *For a continuous time Markov chain on a finite state space* E *with generator matrix* $\mathbf{Q} = q(\mathbf{P} - \mathbb{I})$ *with reversible, irreducible* $\mathbf{P}$*,* q *the largest entry in modulus of* $\mathbf{Q}$ *and stationary distribution* π*, for any probability distribution* ν *on* E*,*

$$\|\nu e^{\mathbf{Q}t} - \pi\|_{\frac{1}{\pi}} \leq \|\nu - \pi\|_{\frac{1}{\pi}} e^{-q(1-|\lambda_2(\mathbf{P})|)t}, \tag{7.1}$$

where $|\lambda_2(\mathbf{P})|$ *is the second largest eigenvalue modulus of* $\mathbf{P}$*.*

We will return to this proposition later in the section. Now we will look at the new type of time-inhomogeneous Markov chains that we are going to apply in this chapter. First, suppose that time is divided into slots of size Δt and the generator matrix changes at these intervals. Furthermore, suppose that the bounded generator matrix $\mathbf{Q}_i$ determines the transition probabilities in the time interval $(i\Delta t, (i+1)\Delta t]$. The method of uniformization gives the corresponding transition probability matrix $\mathbf{P}(i\Delta t, (i+1)\Delta t)$ as it was described for time-homogeneous Markov chains in Section 5.1. The upper bound on departure rates over all states will be denoted q_i for each $\mathbf{Q}_i$. Therefore,

$$\mathbf{P}(i\Delta t, (i+1)\Delta t) = e^{\mathbf{Q}_i \Delta t} = e^{q_i(\mathbf{P}_i - \mathbb{I})\Delta t}, \tag{7.2}$$

where the matrix $\mathbf{P}_i = \mathbb{I} + \frac{1}{q_i}\mathbf{Q}_i$ is irreducible and reversible with stationary distribution π_i.

Note that this evolution is not time-homogeneous for all $t \geq 0$. This is not time-inhomogeneous in the way we described in Chapter 6. Rather than having a continuous function determine our adiabatic evolution, we have a right-continuous, step function function determine out adiabatic evolution. This will result in time-inhomogeneity in the system resulting in a changing $\mathbf{Q}_i$ which is updated at fixed intervals of time. In practice, the time-inhomogeneity can be due to the nature of the underlying process, as we had in Chapter 6 or due to uncertainties in measurements of parameters.

Let ν_n be the distribution of the chain at time $n\Delta t$. When introducing an adiabatic time in the Section 7.2, we will be interested in the distance between ν_n and the stationary distribution π_n corresponding to matrix $\mathbf{P}_n$ at time $n\Delta t$.

The following Theorem gives an upper bound on the distance at time $n\Delta t$ in terms of the distance at $n_0\Delta t$ for integer $n_0 < n$. The proof of this theorem is given [27].

Theorem 20 *For the time-inhomogeneous Markov chain generated by the matrices* $\{\mathbf{Q}_i\}_{i=0}^n = \{q(\mathbf{P}_i - \mathbb{I})\}_{i=0}^n$ *from time* 0 *to time* $n\Delta t$*,*

$$\begin{aligned}
&\|\nu_n - \pi_n\|_{TV} \\
&\quad \le \frac{1}{2}\|\nu_{n_0} - \pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} \prod_{i=n_0}^{n-1} e^{-q_i(1-|\lambda_2(\mathbf{P}_i)|\Delta t} \sqrt{\max_{k\in E} \frac{\pi_i(k)}{\pi_{i+1}(k)}} \\
&\quad + \frac{1}{2}\sum_{i=n_0}^{n-1}\left[\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} \prod_{j=i+1}^{n-1} e^{-q_i(1-|\lambda_2(\mathbf{P}_i)|\Delta t} \sqrt{\max_{k\in E} \frac{\pi_j(k)}{\pi_{j+1}(k)}}\right],
\end{aligned} \tag{7.3}$$

where ν_n *is the distribution at time* $n\Delta t$*,* ν_{n_0} *is the distribution at time* $n_0\Delta t$ *for* $n_0 < n$ *and* $\{\mathbf{P}_i\}_i$ *are irreducible and reversible with respect to the stationary distribution* π_i*.*

Now that we have this result, we apply our knowledge of continuous-time Markov chains to a queueing process in the next section.

7.2 Queueing Processes

A queue is typically defined by its arrival rate, service or departure rate, number of servers and buffer size. In a finite buffer size queue, we are often interested in maintaining a distribution which is more biased towards smaller queue lengths, since otherwise we will have what is called high blocking probabilities. Such a stable queue is achieved by keeping the departure rate strictly above the arrival rate. However, it is physically impossible to have the departure rate arbitrarily higher than the arrival rate due to physical limitations such as the speed of the underlying router circuitry or the amount of power consumption. Furthermore, there might be other constraints on the departure rate due to considerations of the network as a whole. For example, it might not be best to always have a large departure rate that results in traffic bursts, and potentially causes congestion somewhere else in the network. Or in the case of using multiple queues of different kinds of traffic, each of which has to satisfy some pre-specified quality of service (QoS), sending packets from one queue will affect the other queues. As a result, controlling the packet sending rates depending on the types of traffic is desirable to allow for different flows to meet their QoS's requirements. In wireless networks, it is often preferable to maintain a certain departure rate and nothing more, due to power restrictions. Last but not least, in a network of multiple wireless nodes, the number of collisions is an increasing function of the traffic density. If every node sends its packets as fast as it can, then collisions will happen all the time. Hence it becomes necessary to monitor the departure rate and keep it at such a level to maintain a stable queue, and at the same time achieve the desired network objectives.

That said, we consider a queue in which packets arrive at a fixed but unknown rate and we use an estimate of this arrival rate to decide a queueing policy designed to ensure stable queues. In particular, we let the departure rate be always higher than the estimated arrival rate by some fixed multiplicative constant, anticipating that the estimate will be correct in the long-run. Since the estimated arrival rate is changed and is more accurate with time, the departure rate is also changed. As a result, the packets in the queue evolve according to a time inhomogeneous Markov chain dictated by this adaptive departure policy. We study the time required for the queue to reach this stationary distribution using the above outlined adiabatic approach, under suitable estimation and departure policies for the unknown arrival rate.

We begin with a basic definition of a M/M/1/K queue for readers that are unaware.

Definition 37 *A M/M/1/K queue is a stochastic process on the set* $\{0, 1, \cdots, K\}$ *where each element of this set corresponds to the number of members of the queue with the following properties:*

- *Arrivals occur at rate* λ *according to a Poisson process, moving the process from state* i *to state* $i + 1$
- *Departures occur at rate* μ*, according to a Poisson process moving the process from state* i *to state* $i - 1$
- *The buffer is of size* K*, so when in state* 0 *or state* K *you cannot move to a state that is outside of the set* $\{0, 1, \cdots, K\}$*.*

The M/M/1/K model can be described as a finite-state, continuous-time Markov chain.

The generator matrix of a continuous-time M/M/1/K queue is:

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\mu+\lambda) & \lambda & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & \mu & -(\mu+\lambda) & \lambda \\ \cdots & 0 & 0 & \mu & -\mu \end{pmatrix}.$$

In this section we apply the adiabatic evolution model defined in Section 7.1 to a queueing process. In particular, we consider time inhomogeneity due to uncertainty in a parameter. Consider an M/M/1/K queue with unknown packet arrival rate λ per unit time. We estimate λ at time $i\Delta t$ denoted by $\hat{\lambda}_i$ and decide packet departure rate, $\mu_i = f(\hat{\lambda}_i)$ based on this estimate.

Definition 38 *A* <u>*queueing policy*</u> *is defined as the sequence* $\{\hat{\lambda}_i, \mu_i = f(\hat{\lambda}_i)\}_{i\geq 1}$ *where* $f : \mathbb{R}^+ \to \mathbb{R}^+$ *and* μ_i *is applied for time from* $(i\Delta t, (i+1)\Delta t]$*.*

The queueing policy decides the following generator matrix from $(i\Delta t, (i+1)\Delta t]$:

$$\mathbf{Q}_i = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu_i & -(\mu_i+\lambda) & \lambda & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & \mu_i & -(\mu_i+\lambda) & \lambda \\ \cdots & 0 & 0 & \mu_i & -\mu_i \end{pmatrix}$$

In this scenario, the upper bound on the departure rates over all states is $\lambda + \mu_i$. Therefore the corresponding probability transition matrix

$$\mathbf{P}\left(i\Delta t, (i+1)\Delta t\right) = e^{\mathbf{Q}_i \Delta t} = e^{(\lambda+\mu_i)(\mathbf{P}_i - \mathbb{I})\Delta t}, \tag{7.4}$$

where the matrix, $\mathbf{P}_i$ is

$$\mathbf{P}_i = \begin{pmatrix} 1-\beta_i & \beta_i & 0 & 0 & \cdots \\ 1-\beta_i & 0 & \beta_i & 0 & \cdots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \cdots & 0 & 1-\beta_i & 0 & \beta_i \\ \cdots & 0 & 0 & 1-\beta_i & \beta_i \end{pmatrix},$$

where $\beta_i = \frac{\lambda}{\mu_i+\lambda}$. The $\mathbf{P}_i$ are reversible and irreducible with stationary distribution given as in Proposition 23:

$$\pi_i = \frac{1}{\sum_{r=0}^{K} \rho_i^r}[1, \rho_i, \rho_i^2, \cdots, \rho_i^K], \tag{7.5}$$

where $\rho_i = \frac{\beta_i}{1-\beta_i} = \frac{\lambda}{\mu_i}$.

The following proposition was proved in [8] and it gives us a strict measurement of $|\lambda_2(\mathbf{P}_i)|$ for the queueing policy.

Proposition 30 *Letting $\rho_i = \frac{\lambda}{\mu_i}$ and now considering π to be the irrational number,*

$$|\lambda_2(\mathbf{P}_i)| = 2\frac{\sqrt{\rho_i}}{(1+\rho_i)}\cos\left(\frac{\pi}{K+1}\right) \tag{7.6}$$

Now we look at a specific queueing policy determined by the time average of number of packets arrived. Let $X_k \sim \text{Poisson}(\lambda\Delta t)$ be the number of packets in the k^{th} slot of duration Δt and let $\delta > 0$ be constant. Suppose

$$\hat{\lambda}_i = \frac{1}{i\Delta t}\sum_{k=1}^{i} X_k, \tag{7.7}$$

$$\mu_i = f(\hat{\lambda}_i) = (1+\delta)\hat{\lambda}_i. \tag{7.8}$$

This particular queueing policy ensures that the departure rate is always higher than the estimated arrival rate and since the estimated arrival rate itself must approach the actual one, we are ensured a stable queue.

With this specific queueing policy, we have the adiabatic evolution by the matrices $\mathbf{Q}_i$. The corresponding $\rho_i = \frac{\lambda}{\mu_i} = \frac{\lambda}{(1+\delta)\hat{\lambda}_i}$. With full knowledge of the arrival rate, the above ratio becomes $\rho = \frac{1}{1+\delta}$ and we will say that the corresponding matrix $\mathbf{P}$ has stationary distribution π. We now make an analogue of the adiabatic time in this setting.

Definition 39 *Given the above transitions generating a continuous-time Markov chain $\mathbf{P}(0, n\Delta t)$, $\epsilon > 0$ and $\gamma < 1$, the expanded adiabatic time, $t_{ad}(\mathbf{P}, \epsilon, \gamma)$ is defined as*

$$\Delta t \inf\{n | \mathbb{P}\left(\|\nu\mathbf{P}(0, n\Delta t) - \pi\|_{TV} < \epsilon\right) > 1 - \gamma\} \tag{7.9}$$

where the infimum is taken over all probability distributions ν

The following theorem gives a sufficient condition on the time we must wait before the distribution of the queue length converges to the desired stationary distribution π. Note that π is decided by δ and can be designed to give a stable stationary distribution.

To understand the proof of the theorem, we first state the following lemma from [26]. This lemma will bound the terms in Theorem 20. The proof of the lemma can be found in [26].

Lemma 3 *For $0 < \epsilon_0 < 1$ and $0 < \gamma_1 < 1$, there exists*

$$n_0 = \frac{2\log\left(\frac{2}{\gamma_1}\right)}{\lambda\Delta t\left(\epsilon_0^2 - \epsilon_0^3\right)}$$

such that

- *For $i \geq n_0$, with probability at least $1-\gamma_1$,*

$$\left|\hat{\lambda}_i - \lambda\right| \leq \lambda\epsilon_0,$$
$$e^{-(\lambda+\mu_i)(1-|\lambda_2(\mathbf{P}_i)|)\Delta t} < A,$$
$$\|\pi_i - \pi\|_{TV} < \frac{\epsilon_0(1+\delta)}{2\left(\delta - \epsilon_0(1+\delta)\right)}$$

where $A = e^{-\lambda\Delta t\left(\sqrt{(1+\delta)(1-\epsilon_0)}-1\right)^2}$.

- *For* $\epsilon_1 = \frac{1}{n_0}\left(\frac{1}{\sqrt{\lambda\Delta t\gamma_2}} + \epsilon_0\right)$ *and* $0 < \gamma_2 < 1$ *and for* $i \geq n_0$*, with probability at least* $1-\gamma = 1-\gamma_1-\gamma_2$

$$\left|\hat{\lambda}_{i+1} - \hat{\lambda}_i\right| < \lambda\epsilon_1,$$
$$\sqrt{\max_{k\in\{0,1,\cdots,K\}} \frac{\pi_i(k)}{\pi_{i+1}(k)}} < B,$$
$$\|\pi_i - \pi_{i+1}\|_{\frac{1}{\pi_{i+1}}} < C,$$

where

$$B = \sqrt{\frac{[(1+\delta)(1-\epsilon_0+\epsilon_1)]^{K+1}-1}{[(1+\delta)(1-\epsilon_0)]^{K+1}-1}}$$

and

$$C = \frac{\sqrt{(1+\delta)(1-\epsilon_0)}\epsilon_1}{|(1-\epsilon_0)^2(1+\delta)-(1-\epsilon_0+\epsilon_1)|}.$$

- *With probability at least $1-\gamma_1$,*

$$\|\nu_{n_0} - \pi_{n_0}\|_{\frac{1}{\pi_{n_0}}} < \frac{[(1+\delta)(1+\epsilon_0)]^{\frac{K}{2}+1}}{\delta}.$$

Now that we have this important lemma, we use it in the proof of the following theorem.

Theorem 21 *Given $0 < \epsilon 1$, $0 < \gamma < 1$ and λ, the unknown arrival rate, the queueing policy described above with $\delta > 0$ for a M/M/1/K queue has*

$$t_{ad}(\mathbf{P}, \epsilon, \gamma) \\ \leq \frac{2\log\left(\frac{2}{\gamma_1}\right)}{\lambda\left(\epsilon_0^2 - \epsilon_0^3\right)} + \frac{\log\left(2[(1+\epsilon_0)(1+\delta)]^{\frac{k}{2}+1}\right) - \log\left(\epsilon\delta\right)}{\frac{1}{2\Delta t}\log\left(\frac{[(1+\delta)(1-\epsilon_0)]^{K+1}-1}{[(1+\delta)(1-\epsilon_0+\epsilon_1)]^{K+1}-1}\right) + \lambda\left(\sqrt{(1+\delta)(1-\epsilon_0)} - 1\right)^2} \tag{7.10}$$

where ϵ_0 satisfies

$$e^{-\lambda\Delta t\left(\sqrt{(1+\delta)(1-\epsilon_0)}-1\right)^2}\sqrt{\frac{[(1+\delta)(1-\epsilon_0+\epsilon_1)]^{K+1}-1}{[(1+\delta)(1-\epsilon_0)]^{K+1}-1}} < 1,$$

$$\frac{\frac{\sqrt{(1+\delta)(1-\epsilon_0)}\epsilon_1}{|(1-\epsilon_0)^2(1+\delta)-(1-\epsilon_0+\epsilon_1)|}}{1 - e^{-\lambda\Delta t\left(\sqrt{(1+\delta)(1-\epsilon_0)}-1\right)^2}\sqrt{\frac{[(1+\delta)(1-\epsilon_0+\epsilon_1)]^{K+1}-1}{[(1+\delta)(1-\epsilon_0)]^{K+1}-1}}} \leq \frac{\epsilon}{2},$$

$$\epsilon_0 \leq \frac{\epsilon\delta}{(1+\delta)(1+\epsilon)}$$

and $\epsilon_1 = \frac{\lambda\Delta t\left(\epsilon_0^2-\epsilon_0^3\right)}{2\log\left(\frac{2}{\gamma_1}\right)}\left(\frac{1}{\sqrt{\lambda\Delta t\gamma_2}} + \epsilon_0\right)$, $0 < \gamma_1 < \gamma$, $\gamma_2 = \gamma - \gamma_1$.

Hence, for given small ϵ and γ, the theorem gives the sufficient amount of time to converge to a stable distribution within ϵ with high probability of $1-\gamma$. The choice of ϵ_0 to be the largest which satisfies all three conditions will give the lowest lower bound in the theorem. At large enough time, the estimated arrival rate must approach the actual arrival rate and the difference can be bounded by ϵ_0 with a high probability. Furthermore, two consecutive estimates, can differ only by a maximum of ϵ_1.

Chapter 8

Applications to an Ising Model with Glauber Dynamics

The second application of the asymptotic bound of the adiabatic time for continuous-time Markov chains is to a statistical mechanical model called the Ising model. We will define the Ising model and Glauber dynamics in Section 8.1 and then we will apply Theorem 19 to an Ising Model with Glauber dynamics on three different graph structures. In Section 8.2 we will consider a one-dimensional torus, in Section 8.3 we will consider a two-dimensional torus, and in Section 8.4 we will look at a general d-dimensional torus.

8.1 The Ising Model with Glauber Dynamics

In this section we will introduce the concept of a nearest-neighbor Ising model. Before we do this we first consider a graph $G_n = (V_n, E_n)$ with a finite number of vertices, $V_n = \{1, \cdots, n\}$, and a collection of edges that connect pairs of vertices, E_n. We will create a matrix to describe how the edges connect the vertices of the graph.

Definition 40 *A* <u>*nearest-neighbor communication matrix*</u> *of the graph* G_n *is a symmetric matrix* $\mathbf{M}$ *so that*

$$\mathbf{M}_{ij} = \begin{cases} 1 & \text{if there is an edge connecting vertex } i \text{ to vertex } j \\ 0 & \text{if there is not an edge connecting vertex } i \text{ to vertex } j. \end{cases}$$

The nearest-neighbor Ising model is a probability distribution defined on $\{-1,1\}^{V_n} = \{-1,1\}^n$. We will use the following definition to introduce the nomenclature for the vector elements of this set and the entries of the vectors.

Definition 41 *For* $\mathbf{x} \in \{-1,1\}^{V_n}$ *and* $1 \le i \le n$ *we call* $\mathbf{x}(i)$ *a* <u>*spin*</u>*. We call* $\mathbf{x}$ *a* <u>*configuration of spins*</u> *on* V_n*.*

Neighboring spins 'interact' multiplicatively and for each location j, one can measure the energy contribution of these interactions for all the spins neighboring j. This energy depends on a parameter β, which is an interaction strength coefficient having a physical interpretation as the inverse temperature. I summarize this in the following definition.

Definition 42 *For a given configuration of spins* $\mathbf{x}$*, a real-valued energy function, called a* <u>*local Hamiltonian*</u>*, sends a location* $j \in V_n$ *to*

$$\mathcal{H}^{loc}(\mathbf{x}(j)) = -\beta \sum_{i \neq j} \mathbf{M}_{i,j} \mathbf{x}(i)\mathbf{x}(j). \tag{8.1}$$

From this we define a real-valued energy function on $\{-1,1\}^{V_n}$. This function will prescribe to each configuration of spins a total energy.

Definition 43 *A real-valued energy function, called a* <u>*Hamiltonian*</u>*, sends a configuration of spins* $\mathbf{x}$ *to*

$$\mathcal{H}(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{n}\mathcal{H}^{loc}(\mathbf{x}(j)). \tag{8.2}$$

This Hamiltonian measurement will be crucial in the creation of a probability distribution on the configuration of spins. We now define the nearest-neighbor Ising model on G_n

Definition 44 *Letting*

$$Z(\beta) = \sum_{\mathbf{x}\in\{-1,1\}^{V_n}} e^{\mathcal{H}(\mathbf{x})}$$

we define the <u>*nearest-neighbor Ising model*</u> *as a probability distribution* μ *on* $\{-1,1\}^{V_n}$ *dependent on the parameter* β *given by*

$$\mu(\mathbf{x};\beta) = Z(\beta)^{-1}e^{\mathcal{H}(\mathbf{x})}. \tag{8.3}$$

We now define Glauber dynamics for this probability distribution μ as a continuous-time, time-homogeneous Markov chain $X(t)$ on $\{-1,1\}^{V_n}$. We number the configurations $\mathbf{x}_1,\cdots,\mathbf{x}_{2^{|V_n|}}$ and define the entries of $\mathbf{P}(t)$ as $p(t)_{kl} = \mathbb{P}\left(X(t)=\mathbf{x}_l \middle| X(0)=\mathbf{x}_k\right)$.

One can describe how Glauber dynamics work in the case where each vertex of a connected graph is of the same degree. In this case, for each location j, we have an independent exponential clock with parameter one associated with it. Suppose we initially have configuration of spins $\mathbf{x}$ on the graph. When the clock at location j rings, the spin $\mathbf{x}(j)$ is reselected with the following probability:

$$\mathbb{P}\left(\mathbf{x}(j)=1\right) = \frac{e^{-\mathcal{H}^{loc}(\mathbf{x}_+(j))}}{e^{-\mathcal{H}^{loc}(\mathbf{x}_-(j))}+e^{-\mathcal{H}^{loc}(\mathbf{x}_+(j))}} \tag{8.4}$$

where $\mathbf{x}_+(i) = \mathbf{x}_-(i) = \mathbf{x}(i)$ for $i \neq j$ and $\mathbf{x}_+(j) = 1$ and $\mathbf{x}_-(j) = -1$. Notice that $\mathcal{H}^{loc}(\mathbf{x}_-(j)) = -\mathcal{H}^{loc}(\mathbf{x}_+(j))$. This would imply that

$$\mathbb{P}(\mathbf{x}(j) = 1) = \frac{1}{2}\left(1 - \tanh\left(\mathcal{H}^{loc}(\mathbf{x}_+(j))\right)\right)$$
$$\mathbb{P}(\mathbf{x}(j) = -1) = \frac{1}{2}\left(1 + \tanh\left(\mathcal{H}^{loc}(\mathbf{x}_+(j))\right)\right).$$

In the case of G_n being a connected graph with each vertex having the same degree, this process describes a continuous-time, time-homogeneous Markov chain on the space of spin configurations. The stationary distribution of this Markov chain is μ.

Now consider a linear adiabatic evolution of Hamiltonians. Let $\mathcal{H}_0$ be the initial Hamiltonian with thermodynamic parameter β_0 and let $\mathcal{H}_1$ be the final Hamiltonian with thermodynamic parameter β_1. We have for $t \in [0, 1]$

$$\mathcal{H}[t] = (1-t)\mathcal{H}_0 + t\mathcal{H}_1. \tag{8.5}$$

We can also define the linear adiabatic evolution of local Hamiltonians accordingly

$$\mathcal{H}[t]^{loc} = (1-t)\mathcal{H}_0^{loc} + t\mathcal{H}_1^{loc}. \tag{8.6}$$

We can use local Hamiltonians to define linear adiabatic Glauber dynamics. This process is similar to regular Glauber dynamics when G_n is a connected graph with each vertex having the same degree: for each location j, we have an independent exponential clock with parameter one associated with it. Suppose we initially have configuration of spins $\mathbf{x}$ on the graph. When the clock at location j rings, say for example at time $t \in [0, 1]$, the spin $\mathbf{x}(j)$ is reselected with the following probability:

$$\mathbb{P}_t\left(\mathbf{x}(j)=1\right)=\frac{1}{2}\left(1-\tanh\left(\mathcal{H}[t]^{loc}\left(\mathbf{x}_+(j)\right)\right)\right)$$
$$\mathbb{P}_t\left(\mathbf{x}(j)=-1\right)=\frac{1}{2}\left(1+\tanh\left(\mathcal{H}[t]^{loc}\left(\mathbf{x}_+(j)\right)\right)\right).$$

Recall that neighboring configurations $\mathbf{x}$ and $\mathbf{y}$ have the same spins at all locations except for one location, for example location j. Here $\mathbf{y}(j)=-\mathbf{x}(j)$. The transition rates of our continuous-time, time-inhomogeneous Markov chain evolve according to linear adiabatic Glauber dynamics rules and the transition rates can be represented as

$$q_{\mathbf{xy}}[t]=\left(1-\phi_{\mathbf{xy}}(t)\right)q_{\mathbf{xy}}^{(0)}+\phi_{\mathbf{xy}}(t)q_{\mathbf{xy}}^{(1)} \tag{8.7}$$

where the functions $\phi_{\mathbf{xy}}(t)$ for neighbors $\mathbf{x}$ and $\mathbf{y}$ depend entirely on the spins around the discrepancy site j.

In the following sections of this chapter, we will find these functions on different dimensional tori and apply Theorem 19 to the linear adiabatic Glauber dynamics to find an appropriate bound on the adiabatic time for these continuous-time, time-inhomogeneous Markov chains.

8.2 One Dimensional Torus

We first consider adiabatic Glauber dynamics of an Ising model on $\mathbb{Z}/n\mathbb{Z}$. In this scenario every location j will have two neighbors. Suppose we begin with a configuration of spins $\mathbf{x}$. Given that we remove the spin $\mathbf{x}(j)$ at location j and replace it with a positive spin, we will have three possible scenarios for our local Hamiltonians at location j.

We first consider the scenario when the two neighbors of location j have opposite spins. If we calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location j, we find that $\mathcal{H}_0^{loc}(x_+(j)) = \mathcal{H}_1^{loc}(x_+(j)) = 0$. This will imply that the linear adiabatic evolution of local Hamiltonians $\mathcal{H}[t]^{loc} = 0$ for $t \in [0,1]$. This will imply that $\mathbb{P}_t\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}$ and $\mathbb{P}_t\left(\mathbf{x}(j) = -1\right) = \frac{1}{2}$ for all $t \in [0,1]$. We now have a nonlinear adiabatic evolution of continuous-time, Markov chains. For neighboring configurations $\mathbf{x}$ and $\mathbf{y}$, recall that we seek function $\phi_{\mathbf{xy}}$ such that

$$q_{\mathbf{xy}}[t] = \left(1 - \phi_{\mathbf{xy}}(t)\right) q_{\mathbf{xy}}^{(0)} + \phi_{\mathbf{xy}}(t) q_{\mathbf{xy}}^{(1)}.$$

In this scenario for any time $t \in [0,1]$

$$\begin{aligned} q_{\mathbf{xy}}[t] = \mathbb{P}_t\left(\mathbf{x}(j) = 1\right) &= \frac{1}{2} \\ q_{\mathbf{xy}}^{(0)} = \mathbb{P}_0\left(\mathbf{x}(j) = 1\right) &= \frac{1}{2} \\ q_{\mathbf{xy}}^{(1)} = \mathbb{P}_1\left(\mathbf{x}(j) = 1\right) &= \frac{1}{2} \end{aligned}$$

so we see that any function $\phi_{\mathbf{xy}}$ satisfies the above equation.

Next we consider the scenario when the two neighbors of location j have positive spins. If we again calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location j, we now find that $\mathcal{H}_0^{loc}(x_+(j)) = -2\beta_0$ and $\mathcal{H}_1^{loc}(x_+(j)) = -2\beta_1$. This will imply that the linear adiabatic evolution of local Hamiltonians $\mathcal{H}[t]^{loc} = (1-t)(-2\beta_0) + t(-2\beta_1)$ for $t \in [0,1]$. This will imply that

$$\begin{aligned} \mathbb{P}_t\left(\mathbf{x}(j) = 1\right) &= \frac{1}{2}\left(1 - \tanh\left(-2\beta_0(1-t) - 2\beta_1 t\right)\right) \\ \mathbb{P}_t\left(\mathbf{x}(j) = -1\right) &= \frac{1}{2}\left(1 + \tanh\left(-2\beta_0(1-t) - 2\beta_1 t\right)\right) \end{aligned}$$

for $t \in [0,1]$.

We again have a nonlinear adiabatic evolution of continuous-time, Markov chains. For neighboring configurations $\mathbf{x}$ and $\mathbf{y}$, recall that we seek function $\phi_{\mathbf{xy}}$ such that

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}.$$

In this scenario for any time $t \in [0, 1]$

$$\begin{aligned}
q_{\mathbf{xy}}[t] &= \mathbb{P}_t\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(-2\beta_0(1-t) - 2\beta_1 t\right)\right) \\
q^{(0)}_{\mathbf{xy}} &= \mathbb{P}_0\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(-2\beta_0\right)\right) \\
q^{(1)}_{\mathbf{xy}} &= \mathbb{P}_1\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(-2\beta_1\right)\right).
\end{aligned}$$

It takes a bit of algebra, but one can solve this problem for $\phi_{\mathbf{xy}}$. We see that

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(-2\beta_1)\sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1)\cosh(-2\beta_0 + t(2\beta_0 - 2\beta_1))}.$$

Finally we consider the scenario when the two neighbors of location j have negative spins. If we again calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location j, we now find that $\mathcal{H}^{loc}_0(x_+(j)) = 2\beta_0$ and $\mathcal{H}^{loc}_1(x_+(j)) = 2\beta_1$. This will imply that the linear adiabatic evolution of local Hamiltonians $\mathcal{H}[t]^{loc} = (1-t)(2\beta_0) + t(2\beta_1)$ for $t \in [0, 1]$. This will imply that

$$\begin{aligned}
\mathbb{P}_t\left(\mathbf{x}(j) = 1\right) &= \frac{1}{2}\left(1 - \tanh\left(2\beta_0(1-t) + 2\beta_1 t\right)\right) \\
\mathbb{P}_t\left(\mathbf{x}(j) = -1\right) &= \frac{1}{2}\left(1 + \tanh\left(2\beta_0(1-t) + 2\beta_1 t\right)\right)
\end{aligned}$$

for $t \in [0, 1]$.

We again have a nonlinear adiabatic evolution of continuous-time, Markov chains. For neighboring configurations $\mathbf{x}$ and $\mathbf{y}$, recall that we seek function

$\phi_{\mathbf{xy}}$ such that

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}.$$

In this scenario for any time $t \in [0, 1]$

$$\begin{aligned} q_{\mathbf{xy}}[t] &= \mathbb{P}_t\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(2\beta_0(1-t) + 2\beta_1 t\right)\right) \\ q^{(0)}_{\mathbf{xy}} &= \mathbb{P}_0\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(2\beta_0\right)\right) \\ q^{(1)}_{\mathbf{xy}} &= \mathbb{P}_1\left(\mathbf{x}(j) = 1\right) = \frac{1}{2}\left(1 - \tanh\left(2\beta_1\right)\right). \end{aligned}$$

Performing some algebra again, we solve this problem for $\phi_{\mathbf{xy}}$. We see that

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(2\beta_1)\sinh(t(2\beta_1 - 2\beta_0))}{\sinh(2\beta_1 - 2\beta_0)\cosh(2\beta_1 + t(2\beta_1 - 2\beta_0))}.$$

Because hyperbolic cosines are even functions and hyperbolic sines are odd functions, we have then that the functions $\phi_{\mathbf{xy}}$ are the same if the neighboring spins are both positive or both negative. We have that

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}$$

has solution

$$\phi(t) = \frac{\cosh(-2\beta_1)\sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1)\cosh(-2\beta_0 + s(2\beta_0 - 2\beta_1))} \tag{8.8}$$

regardless of the configuration of spins. We would therefore have that

$$\mathbf{Q}[t] = (1 - \phi(t))\mathbf{Q}_0 + \phi(t)\mathbf{Q}_1.$$

We see that $\phi'(1) \neq 0$ when $\beta_0 \neq \beta_1$, so we can apply Theorem 19 to get an asymptotic bound for the adiabatic time with respect to the mixing time and we can apply Theorem 15.1 from [18] to find a bound for the mixing time for Glauber dynamics of an Ising Model on a one dimensional torus. We have the following asymptotic for our adiabatic time

$$\begin{aligned} & t_{ad}\left(\mathbf{P}(0),\mathbf{P}(1),\epsilon\right) \\ & = \mathcal{O}\left(\frac{n}{\epsilon}(2\beta_0 - 2\beta_1)\left[\coth(2\beta_0 - 2\beta_1) - \tanh(-2\beta_1)\right]\left[\frac{\log(n)+\log(\frac{2}{\epsilon})}{1-\tanh(2\beta_1)}\right]^2\right). \end{aligned} \tag{8.9}$$

8.3 Two Dimensional Torus

We now consider adiabatic Glauber dynamics of an Ising model on $(\mathbb{Z}/n\mathbb{Z})^2$. In this scenario every location (i,j) will have four neighbors. Again suppose that we begin with a configuration of spins $\mathbf{x}$. Given that we remove spin $\mathbf{x}(i,j)$ at location (i,j) and replace it with a positive spin, we will have five possible scenarios for our local Hamiltonians at location (i,j).

We first consider the scenario when two neighbors of location (i,j) have positive spins and two neighbors of location (i,j) have negative spins. One can visualize this with the following example diagram:

$$\begin{array}{ccccc} & & -1 & & \\ & & | & & \\ +1 & -- & (i,j) & -- & +1 \\ & & | & & \\ & & -1 & & \end{array}$$

If we calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location (i,j), we find that $\mathcal{H}_0^{loc}(x_+(i,j)) = \mathcal{H}_1^{loc}(x_+(i,j)) = 0$. If we recall the scenario from the one dimensional torus where both local Hamiltonians are 0, we see that any function $\phi_{\mathbf{xy}}$ satisfies the equation

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}.$$

Next we consider the scenario when three neighbors of location (i, j) have positive spins and one neighbor of location (i, j) has a negative spin. One can visualize this with the following example diagram:

$$\begin{array}{ccccc} & & -1 & & \\ & & | & & \\ +1 & -- & (i,j) & -- & +1 \\ & & | & & \\ & & +1 & & \end{array}$$

If we again calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location (i, j), we now find that $\mathcal{H}^{loc}_0(x_+(i, j)) = -2\beta_0$ and $\mathcal{H}^{loc}_1(x_+(i, j)) = -2\beta_1$. If we recall the scenario from the one dimensional torus where we had these local Hamiltonians, we see that

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(-2\beta_1) \sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1) \cosh(-2\beta_0 + t(2\beta_0 - 2\beta_1))}$$

satisfies the equation

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}.$$

Finall we consider the scenario when all neighbors of location (i, j) have positive spins. If we calculate both the initial and final local Hamiltonians of the configuration $\mathbf{x}$ at location (i, j), we now find that $\mathcal{H}^{loc}_0(x_+(i, j)) = -4\beta_0$ and $\mathcal{H}^{loc}_1(x_+(i, j)) = -4\beta_1$. This will imply that the linear adiabatic evolution of local Hamiltonians $\mathcal{H}[t]^{loc} = (1 - t)(-4\beta_0) + t(-4\beta_1)$ for $t \in [0, 1]$. This will imply that

$$\mathbb{P}_t\left(\mathbf{x}(i,j)=1\right) = \frac{1}{2}\left(1-\tanh\left(-4\beta_0(1-t)-4\beta_1 t\right)\right)$$
$$\mathbb{P}_t\left(\mathbf{x}(i,j)=-1\right) = \frac{1}{2}\left(1+\tanh\left(-4\beta_0(1-t)-4\beta_1 t\right)\right)$$

for $t \in [0,1]$.

We again have a nonlinear adiabatic evolution of continuous-time, Markov chains. For neighboring configurations $\mathbf{x}$ and $\mathbf{y}$, recall that we seek function $\phi_{\mathbf{xy}}$ such that

$$q_{\mathbf{xy}}[t] = \left(1-\phi_{\mathbf{xy}}(t)\right) q_{\mathbf{xy}}^{(0)} + \phi_{\mathbf{xy}}(t) q_{\mathbf{xy}}^{(1)}.$$

In this scenario for any time $t \in [0,1]$

$$q_{\mathbf{xy}}[t] = \mathbb{P}_t\left(\mathbf{x}(i,j)=1\right) = \frac{1}{2}\left(1-\tanh\left(-4\beta_0(1-t)-4\beta_1 t\right)\right)$$
$$q_{\mathbf{xy}}^{(0)} = \mathbb{P}_0\left(\mathbf{x}(i,j)=1\right) = \frac{1}{2}\left(1-\tanh\left(-4\beta_0\right)\right)$$
$$q_{\mathbf{xy}}^{(1)} = \mathbb{P}_1\left(\mathbf{x}(i,j)=1\right) = \frac{1}{2}\left(1-\tanh\left(-4\beta_1\right)\right).$$

After some algebra, one can solve this problem for $\phi_{\mathbf{xy}}$. We see that

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(-4\beta_1)\sinh(t(4\beta_0-4\beta_1))}{\sinh(4\beta_0-4\beta_1)\cosh(-4\beta_0+t(4\beta_0-4\beta_1))}.$$

Due to the symmetry of the local Hamiltonians and the functions $\phi_{\mathbf{xy}}$, the other two scenarios will result in similar solutions to the equation

$$q_{\mathbf{xy}}[t] = \left(1-\phi_{\mathbf{xy}}(t)\right) q_{\mathbf{xy}}^{(0)} + \phi_{\mathbf{xy}}(t) q_{\mathbf{xy}}^{(1)}.$$

If all neighbors of location (i,j) have negative spins, the function would be

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(-4\beta_1)\sinh(t(4\beta_0-4\beta_1))}{\sinh(4\beta_0-4\beta_1)\cosh(-4\beta_0+t(4\beta_0-4\beta_1))}$$

and if three of the neighbors of location (i,j) have negative spins while one of the neighbors of location (i,j) has a positive spin, we see that

$$\phi_{\mathbf{xy}}(t) = \frac{\cosh(-2\beta_1)\sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1)\cosh(-2\beta_0 + t(2\beta_0 - 2\beta_1))}.$$

Depending on which neighbors we have locally at (i,j), we have one of two functions as a possibility to to the solution of

$$q_{\mathbf{xy}}[t] = (1 - \phi_{\mathbf{xy}}(t))\, q^{(0)}_{\mathbf{xy}} + \phi_{\mathbf{xy}}(t) q^{(1)}_{\mathbf{xy}}.$$

These two functions are

$$\begin{aligned}\phi_1(t) &= \frac{\cosh(-2\beta_1)\sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1)\cosh(-2\beta_0 + t(2\beta_0 - 2\beta_1))}\\ \phi_2(t) &= \frac{\cosh(-4\beta_1)\sinh(t(4\beta_0 - 4\beta_1))}{\sinh(4\beta_0 - 4\beta_1)\cosh(-4\beta_0 + t(4\beta_0 - 4\beta_1))}.\end{aligned}$$

Assuming that $\tanh(2\beta_1) \leq \frac{1}{2}$, we see that $\phi_1(t) \leq \phi_2(t)$ for all $t \in [0,1]$. We see that $\phi_1'(1) \neq 0$ when $\beta_0 \neq \beta_1$, so we can again apply Theorem 19 to find an asymptotic bound for the adiabatic time with respect to the mixing time and we can agian apply Theorem 15.1 from [18] to find a bound for the mixing time for Glauber dynamics of an Ising Model on a two dimensional torus. We have the following asymptotic for our adiabatic time

$$\begin{aligned}&t_{ad}\left(\mathbf{P}(0), \mathbf{P}(1), \epsilon\right)\\ &= \mathcal{O}\left(\frac{n^2}{\epsilon}(2\beta_0 - 2\beta_1)\left[\coth(2\beta_0 - 2\beta_1) - \tanh(-2\beta_1)\right]\left[\frac{\log(n) + \log(\frac{2}{\epsilon})}{1 - \tanh(2\beta_1)}\right]^2\right).\end{aligned} \tag{8.10}$$

8.4 General Dimensional Torus

The adiabatic Glauber dynamics of an Ising model on a d-dimensional torus $(\mathbb{Z}/n\mathbb{Z})^d$ solves similarly. In this setting the minimum function $\phi(t)$ that we use in Theorem 19 is the same as that of the two dimensional torus

$$\phi(t) = \frac{\cosh(-2\beta_1)\sinh(t(2\beta_0 - 2\beta_1))}{\sinh(2\beta_0 - 2\beta_1)\cosh(-2\beta_0 + t(2\beta_0 - 2\beta_1))}.$$

Here if we assume that $\tanh(2\beta_1) \leq \frac{1}{d}$ we can similarly asymptotically bound the adiabatic time by

$$\begin{aligned} &t_{ad}(\mathbf{P}(0), \mathbf{P}(1), \epsilon) \\ &= \mathcal{O}\left(\frac{n^d}{\epsilon}(2\beta_0 - 2\beta_1)\left[\coth(2\beta_0 - 2\beta_1) - \tanh(-2\beta_1)\right]\left[\frac{\log(n) + \log(\frac{2}{\epsilon})}{1 - \tanh(2\beta_1)}\right]^2\right). \end{aligned} \tag{8.11}$$

Chapter 9

Final Remarks

This summarizes my graduate body of work related to adiabatic and stable adiabatic times. Throughout this manuscript, we have considered many different types of adiabatic evolutions and we have derived many adiabatic theorems. You may argue that some of the results are redundant, but progress in this area of research was made incrementally and I value the mathematical process involved with each case. We have seen the usefulness in applying an adiabatic evolution to problems in networking and statistical mechanics. Our ultimate goal is to find a stable adiabatic result for a general adiabatic transition, however, we would be content with finding this result for Lipschitz continuous matrix-valued functions.

We were able to show that the bounds on the adiabatic time were the best in each setting (linear, nonlinear, general), but we were not able to show that the bounds on the stable adiabatic time were the best in each setting(linear, piecewise bi-Lipschitz nonlinear). Our second goal is finding the lowest bound of the stable adiabatic time. My final goal is to find a strict bound, rather than an asymptotic bound of the stable adiabatic time with respect to the largest mixing time.

We emphasize throughout this paper that the use of the mixing time, rather than the spectral gap, in this research was crucial. The results from Chapter 4 can be used beyond the content of this dissertation.

With the generalizations that we made to matrix-valued functions, I am positive that many more applications await. I look forward to exploring the possibilities of adiabatic transition in the future.

Bibliography

[1] D. Aldous and J.A. Fill, *Reversible Markov chains*, stat-www.berkeley.edu, 2002.

[2] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd and O. Regev, *Adiabatic Quantum Computation is Equivalent to Standard Computation*, SIAM review 50 (4), 755 - 787.

[3] A. Ambainis and O. Regev, *An Elementary Proof of the Quantum Adiabatic Theorem*, arXiv:quant-ph/0411152v2

[4] K. Bradford and Y. Kovchegov, *Adiabatic times for Markov chains and their applications*, Journal of Statistical Physics, Vol. **143**, 2011, pp. 955-969.

[5] K. Bradford, Y. Kovchegov and T. Nguyen, *Stable Adiabatic Times for Markov Chains*, submitted to Stochastic Processes and their Applications.

[6] P. Brémaud, *Markov Chains: Gibbs fields, Monte Carlo Simulation, and Queues*, Springer Science+Business Media Inc. Texts in Applied Mathematics, Vol. **31**, 2010.

[7] P. Diaconis, *Group Representation in Probability and Statistics*, Lecture Notes–Monograph Series, Volume 11, Hayward, CA: Institute of Mathematical Statistics, 1988.

[8] W. Feller, *An introduction to probability theory and its applications, volume 1, 2nd ed.*, New York: John Wiley & Sons, 1959.

[9] V.A. Fock, *Selected Works: Quantum Mechanics and Quantum Field Theory*, Chapman & Hall/CRC, 2004.

[10] S.H. Friedberg, A.J. Insel and L.E. Spence, *Linear Algebra – 4th ed.*, Pearson Education Inc., 2003.

[11] F.R. Gantmacher, *The Theory of Matrices: Volume II*, AMS Chelsea Publishing 1959.

[12] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, 2001.

[13] D.L. Isaacson and R.W. Madsen, *Markov Chains: Theory and Applications*, John Wiley, New York, 1976.

[14] S. Karlin and H.M. Taylor, *A first course in Stochastic Processes*, Academic Press, 1975.

[15] T. Kato, *On the Adiabatic Theorem of Quantum Mechanics*, Journal of t he Physical Society of Japan, Vol. **5**, pp. 435 - 439, 1950.

[16] Y. Kovchegov, *A note on adiabatic theorem for Markov chains*, Statistics and Probability Letters, Vol. **80**, 2010, pp. 186-190.

[17] H. Krovi, M. Ozols and J. Roland, *Adiabatic condition and the quantum hitting time of Markov chains*, Physical Review A **82**, 022333, 2010.

[18] D. Levin, Y. Peres and E. Wilmer, *Markov Chains and Mixing Times*, American Mathematical Society, 2009.

[19] S. Rajagopalan, D. Shah and J. Shin, *Network Adiabatic Theorem: An efficient Randomized Protocol for Contention Resolution*, ACM 978-1-60558-511-6/09/06, 2009.

[20] S. Ross, *Simulation: Fourth Ed.*, Elsevier Academic Press, 2006.

[21] H. Kesten (editor), D. Aldous, G. R. Grimmett, C. D. Howard, F. Martinelli, L. Saloff-Coste and J. M. Steele, *Probability on Discrete Structures (Encyclopedia of Mathematical Sciences)*, Springer, 2010.

[22] L. Saloff-Coste and J. Zúñiga, *Covergence of some time inhomogeneous Markov chains via spectral techniques*, Stochastic Processes and their Applications, Vol. **117**, 2007, pp. 961 - 979.

[23] L. Saloff-Coste and J. Zúñiga, *Merging for time-inhomogeneous finite Markov chains, Part I: singular values and stability*, Electronic Journal of Probability, Vol. **14**, 2009, pp. 1456-1494.

[24] L. Saloff-Coste and J. Zúñiga, *Merging for time-inhomogeneous finite Markov chains, Part II: Nash and log-Sobolev inequalities*, Annals of Probability, Vol. **39** No. **3**, 2011, pp. 1161 - 1203.

[25] L.N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, 1997.

[26] L. Zacharias, *An adiabatic approach to analysis of time inhomogeneous Markov chains: a queueing policy application*, Master's thesis, Oregon State University, Corvallis, OR, USA, September 2011.

[27] L. Zacharias, T. Nguyen, Y. Kovchegov, K. Bradford, *Analysis of Adaptive Queueing Policies via Adiabatic Approach*, Proceedings of the 2013 International Conference on Computing, Networking and Communications (ICNC), 2013, pp. 1053 - 1057.

[28] L. Zacharias, T. Nguyen, Y. Kovchegov and K. Bradford, *Analysis of Adaptive Queueing Policies via Adiabatic Approach*, Submitted.