PARALLEL TEMPERING SPEEDS UP MIXING OF MARKOV CHAINS

CHRIS LEE

1. Introduction

Parallel Tempering (PT) is a method of speeding up the convergence of Metropolis-Hasting based Markov Chain Monte Carlo (MCMC) methods. Samberg [1] showed that in practice it performs significantly better on multi-modal distributions. However, there is not yet a generalized proof of this observation. Fischer and Christian [2] give a proof for restricted Boltzman Machines and Woodard, Schmidler, and Huber [3] give conditions for PT algorithms to be torpidly mixing. In this paper we give an upper bound for PT of a birth and death chain on hypercubes by showing the spectral gap λ_{PT}^*, is strictly less than that of independent untempered chains.

2. Background

2.1. **The Metropolis-Hastings Algorithm.** First we review the metropolis algorithm. Let X be a finite set and $\pi(x)$ be a positive probability on X. Let $S(x, y)$ be the transition matrix of our base chain, which we pick to be a symmetric, irreducible Markov chain on X. Then we define $M(x, y) \in X^2$ to be our Metropolis chain. We get it by "thinning down" the base chain by the following:

$$(1) \qquad M(x, y) = \begin{cases} \frac{\pi(y)}{\pi(x)} S(x, y) & \text{if } \frac{\pi(y)}{\pi(x)} < 1 \\ S(x, y) & \text{if } \frac{\pi(y)}{\pi(x)} \geq 1 \text{ and } x \neq y \\ S(x, x) + \sum_{z \neq x} S(x, z) \left(1 - \frac{\pi(x)}{\pi(z)}\right) & \text{if } x = y \end{cases}$$

This describes the process of sampling a possible step via $S(x, y)$. Then if $\frac{\pi(y)}{\pi(x)} \geq 1$ we take the step. If $\frac{\pi(y)}{\pi(x)} < 1$ we flip a coin with probability $\frac{\pi(y)}{\pi(x)}$. If the coin flip succeeds we take the step. In the case where $x = y$ we get the transition probability by observing that in all cases where the ratio $\frac{\pi(z)}{\pi(x)}$ of a proposed step to z is less than one and the coin flip fails we don't step. This resulting chain is irreducible and aperiodic. Furthermore, the chain has stationary distribution π.

2.2. **Parallel Tempering.** Let M_1 and M_2 be parallel independent metropolis chains on a state space X with target distributions π_1 and π_2 respectively. Then let $M \in (X, X)^2$ such that $M = M_1 \otimes M_2$. It is easy to see that M is the joint chain with stationary distribution $\pi(x, y) = \pi_1(x)\pi_2(y)$. The goal of parallel tempering is to swap the states of M_1 and M_1 without changing their joint target stationary

distribution $\pi(x,y)$. This is done by proposing the swap as a Metropolis-Hastings update. That is, we accept a step from $(x,y) \in X^2$ to $(y,x) \in X^2$ with probability

$$\min\left\{1, \frac{\pi(y,x)}{\pi(x,y)}\right\}$$

Equivalently, the swap is accepted with probability 1 if $\frac{\pi_1(x)}{\pi_2(x)} \leq \frac{\pi_1(y)}{\pi_2(y)}$ and accepted with probability $\frac{\pi(y,x)}{\pi(x,y)}$ otherwise. We will denote the transition matrix of this chain $P \in (X,X)^2$.

3. Defintions and Notation

Definition 3.1. *Let $M = M_1 \otimes M_2$ where $M_1, M_2 \in X^2$ are Metropolis chains with stationary distributions $\pi_1 \neq \pi_2$. We then define the "temper step" matrix $T \in (X,X)^2$ in the following way. Let $(x,y) \in (X,X)$. If $y = x$,*

$$(2) \qquad\qquad\qquad T((x,y),(x,y)) = 1$$

If $y \neq x$,

$$(3) \qquad T((x,y),(x,y)) = \begin{cases} 1 - \frac{\pi(y,x)}{\pi(x,y)} & \frac{\pi_1(x)}{\pi_2(x)} \leq \frac{\pi_1(y)}{\pi_2(y)} \\ 0 & \frac{\pi_1(x)}{\pi_2(x)} > \frac{\pi_1(y)}{\pi_2(y)} \end{cases}$$

$$(4) \qquad T((x,y),(y,x)) = \begin{cases} \frac{\pi(y,x)}{\pi(x,y)} & \frac{\pi_1(x)}{\pi_2(x)} \leq \frac{\pi_1(y)}{\pi_2(y)} \\ 1 & \frac{\pi_1(x)}{\pi_2(x)} > \frac{\pi_1(y)}{\pi_2(y)} \end{cases}$$

Finally, for any $(z,w) \in (X,X)^2$ such that $(z,w) \neq (x,y)$ and $(z,w) \neq (y,x)$,

$$(5) \qquad\qquad\qquad T((x,y),(z,w)) = 0$$

As a sanity check, we note that in each case, $T((x,y),(x,y)) + T((x,y),(y,x)) = 1$. Thus T is stochastic as it should be. The reader may find the following example helpful for visualizing T.

Example 3.2. *Let $M_1, M_2 \in [0,3]^2$ be such that $\pi_1(x) = \theta_1^x/(1+\theta_1)^d$ and $\pi(y) = \theta_2^y/(1+\theta_2)^d$ with $\theta_1 < \theta_2$. Now let $\beta = \frac{\theta_1}{\theta_2}$. Then $T \in ([0,2],[0,2])^2$ is the following matrix,*

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\beta & 0 & \beta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\beta^2 & 0 & 0 & 0 & \beta^2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1-\beta & 0 & \beta & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We will frequently refer to the the points in (X,X) which cause the temper matrix to swap. As such we give the following notation. Let

$$(6) \qquad S = \left\{(x,y) \in X \times X : x \neq y \text{ and } \frac{\pi_1(x)}{\pi_2(x)} \leq \frac{\pi_1(y)}{\pi_2(y)}\right\}.$$

Intuitively, S is the points which do not accept the swap with probability 1. Also note that for all $x \neq y$, if the point $(x,y) \in S$ then the swapped point (y,x) is not.

3.1. Verifying the Temper Step Matrix.
Next, we show that the proposed temper step matrix T satisfies key properties. Namely, that it preserves the stationary distribution π, as promised, and that it is reversible with respect to π.

Lemma 3.3. *Let $M_1, M_2 \in X^2$ be transition matrices of two metropolis chains with stationary distributions π_1 and π_2. Then let T be the temper step matrix as described in definition 3.1. Then the stationary distribution of $M = M_1 \otimes M_2$ and the stationary distribution of $P = MT$ are the same.*

Proof. Let π_M be the stationary distribution of M. Then by properties of the Kronecker product $\pi_M = \pi_1 \otimes \pi_2$. We show that this is the an eigenvector of P with eigenvalue 1. Let $\pi_M(x,y)$ be the (x,y) entry of π_M. Without loss of generality let $(x,y) \in S$. Then observe

$$(\pi_M T)(x,y) = \pi(x,y)\left(1 - \frac{\pi(y,x)}{\pi(x,y)}\right) + \pi(y,x)$$
$$= \pi(x,y) - \pi(y,x) + \pi(y,x)$$
$$= \pi(x,y).$$

If (x,y) not in S then if $x = y$,

$$(\pi_M T)(x,y) = \pi(x,y).$$

If $x \neq y$, then we have

$$(\pi_M T)(y,x) = \pi(x,y)\frac{\pi(y,x)}{\pi(x,y)}$$
$$= \pi(y,x).$$

Thus for any (x,y) we have $(\pi_M MT)(x,y) = (\pi_M T)(x,y) = \pi(x,y)$ and the stationary distribution is preserved. $\square$

Lemma 3.4. *The matrix M is reversible with respect to π_M.*

Proof. Using the reversibility of the independent markov chains M_1 and M_2

$$\pi(x,y)M((x,y),(y,x)) = \pi_1(x)\pi_2(y)M_1(x,y)M_2(y,x)$$
$$= (\pi_1(x)M_1(x,y))(\pi_2(y)M_2(y,x))$$
$$= (\pi_1(y)M_1(y,x))(\pi_2(x)M_2(x,y))$$
$$= \pi(y,x)M((y,x),(x,y)).$$

$\square$

Lemma 3.5. *The matrix T is reversible with respect to π_M.*

Proof. Let $(x,y),(y,x) \in (X,X)^2$ be an element of T. If $T((x,y),(y,x)) = 0$ then $T((y,x),(x,y)) = 0$ as needed. If $T((x,y),(y,x))$ is on the diagonal then clearly this element is reversible, as the diagonals are not affected by transposition. Finally, let

$T((x, y), (y, x))$ be a nonzero, nondiagonal element of T. Then, if without loss of generality let $T((x, y), (y, x))$ below the diagonal. Now,

$$\pi(x, y)T((x, y), (y, x)) = \pi(x, y)$$
$$= \pi(x, y)\left(\frac{\pi(y, x)}{\pi(y, x)}\right)$$
$$= \pi(y, x)\left(\frac{\pi(x, y)}{\pi(y, x)}\right)$$
$$= \pi(y, x)T((y, x), (x, y)).$$

$\square$

Proposition 3.6. *The matrix $P = MT$ is reversible.*

Proof. By Lemma 3.3 we know that $P = MT$ has stationary distribution π_M. Then by Lemma 3.4 and Lemma 3.5 we have,

$$\pi(x, y) \cdot MT((x, y), (y, x)) = \pi(x, y) \sum_{(z,w)\in(X,X)} M((x, y), (z, w)) \cdot T((z, w), (y, x))$$
$$= \sum_{(z,w)\in(X,X)} \pi(x, y)M((x, y), (z, w)) \cdot T((z, w), (y, x))$$
$$= \sum_{(z,w)\in(X,X)} \pi(z, w)M((z, w), (x, y)) \cdot T((z, w), (y, x))$$
$$= \sum_{(z,w)\in(X,X)} M((z, w), (x, y)) \cdot \pi(z, w)T((z, w), (y, x))$$
$$= \sum_{(z,w)\in(X,X)} M((z, w), (x, y)) \cdot \pi(y, x)T((y, x), (z, w))$$
$$= \pi(y, x) \sum_{(z,w)\in(X,X)} M((z, w), (x, y)) \cdot T((y, x), (z, w))$$
$$= \pi(y, x) \cdot MT((y, x), (x, y)).$$

Thus $\pi(x, y)P((x, y), (y, x)) = \pi(y, x) \cdot P((y, x), (x, y))$ which implies P is reversible.
$\square$

3.2. **Reversible Matrices have a Symmetric Conjugate.** In order to analyze the eigenvalues and eigenvectors of P we want to be able to take spectral decompositions. In order to do this we need the matrices to be symmetric. Luckily, reversible markov chains become symmetric when conjugated by the matrix $D_\pi^{1/2}$.

Definition 3.7. *Let $D_\pi^{1/2}$ be a diagonal matrix with entries $\sqrt{\pi(x, y)}$. Similarly, let $D_\pi^{-1/2}$ be it's inverse, a diagonal matrix with entries $\frac{1}{\sqrt{\pi(x,y)}}$.*

Corollary 3.8. *The matrices $\mathcal{P} = D_\pi^{1/2}PD_\pi^{-1/2}$, $\mathcal{M} = D_\pi^{1/2}MD_\pi^{-1/2}$, and $\mathcal{T} = D_\pi^{1/2}TD_\pi^{-1/2}$ are symmetric.*

Proof. By Lemmas 3.4 and 3.5 and Proposition 3.6, each of the matrices are reversible. It then follows from simple computation (recalling the defintion of $D_\pi^{1/2}$ and $D_\pi^{-1/2}$) that these are symmetric. $\square$

Additionally, we give some basic results about the properties of these matrices.

Lemma 3.9. *Let A be a matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and associated eigenvectors $\Psi_1, \ldots, \Psi_n$. If $\mathcal{A} = D_\pi^{1/2} A D_\pi^{-1/2}$ then $\mathcal{A}$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ with associated eigenvectors $\Psi_i = D_\pi^{1/2} \Psi_i$ for $1 \leq i \leq n$.*

Proof. Because $D_\pi^{1/2} D_\pi^{-1/2} = I$, where I is the identity matrix, $\mathcal{A}$ is similar to A. Thus the eigenvalues are the same. Then, we can simply compute the eigenvectors as follows:

$$
\begin{aligned}
D_\pi^{1/2} A D_\pi^{-1/2} \Psi &= D_\pi^{1/2} A D_\pi^{-1/2} \left(D_\pi^{1/2} \Psi \right) \\
&= D_\pi^{1/2} A \Psi \\
&= D_\pi^{1/2} \lambda \Psi \\
&= \lambda \Psi.
\end{aligned}
$$

We note that $D_\pi^{-1/2} B D_\pi^{1/2} = A$ so we can use the same method to deduce the eigensystem of A from $\mathcal{A}$. $\qquad\square$

Finally, because π is the stationary distribution of P we get one more property that will be useful later.

Lemma 3.10. *A vector x is orthogonal to $D_\pi^{1/2} \mathbb{1}$ if and only if the vector $y = \mathcal{P}x = D_\pi^{1/2} M T D_\pi^{-1/2} x$ is orthogonal to $D_\pi^{1/2} \mathbb{1}$.*

Proof. Observe,

$$
\begin{aligned}
(D_\pi^{1/2} \mathbb{1})^\top y &= \mathbb{1}^\top D_\pi^{1/2} (D_\pi^{1/2} M T D_\pi^{-1/2} x) \\
&= \mathbb{1}^\top D_\pi^1 M T D_\pi^{-1/2} x) \\
&= \pi M T D_\pi^{-1/2} x) \\
&= (\pi D_\pi^{-1/2}) x \\
&= \mathbb{1}^\top D_\pi^{1/2} x \\
&= (D_\pi^{1/2} \mathbb{1})^\top x
\end{aligned}
$$

Therefore, $(D_\pi^{1/2} \mathbb{1})^\top y = 0$ if and only if $(D_\pi^{1/2} \mathbb{1})^\top x = 0$. It follows that y is orthogonal to $D_\pi^{1/2} \mathbb{1}$ if and only if x is orthogonal to $D_\pi^{1/2} \mathbb{1}$. $\qquad\square$

4. The Spectrum of the Temper Step Matrix

To prove the main result we need a few Lemmas about the spectrum of T and it's relationship to the sprectrum of M. We also state the Rayleigh Quotient which we will use to bound eigenvalues.

Lemma 4.1. *Let T be as defined in 3.1. Then the eigenvalues of T are 1 and $-\frac{\pi(y,x)}{\pi(x,y)}$ where $(x,y) \in S$. The eigenspace $\Lambda_1 = \{\Psi_T : T\Psi_T = \Psi_T\}$ has an orthogonal basis of eigenvectors of the form either:*

$$
\Psi = e_{(z,z)}
$$

where $k \in X$, or

$$
\Psi = e_{(x,y)} + e_{(y,x)}
$$

where $x \neq y$. Note that $e_{(i,j)}$ is the standard basis vector corresponding to the row $(i,j) \in (X,X)$. The remaining eigenspaces have dimension 1. Let $(x,y) \in S$. Then $-\frac{\pi(y,x)}{\pi(x,y)}$ is an eigenvalue with multiplicity 1 and unique eigenvector

$$\Psi = -\frac{\pi(y,x)}{\pi(x,y)}e_{(x,y)} + e_{(y,x)}$$

Proof. We first prove that $\Psi = e_{(z,z)} \in \Lambda_1$ is an eigenvector. Let $z \in X$. Then $T\Psi$ is the (z,z) column of T. Observe that by defintion 3.1, we know that $T((z,z),(z,z)) = 1$. Additionally, for any other element $(x,y) \in (X,X)$ we have $T((x,y),(z,z)) = 0$. Thus,

$$T\Psi = e_{(z,z)} = \Psi$$

and $\Psi = e_{(z,z)}$ is an eigenvector with eigenvalue $\lambda = 1$. Additionally, there are $|X|$ distinct choices of z. Thus there are $|X|$ many eigenvectors of this form. Now we prove $\Psi = e_{(x,y)} + e_{(y,x)}$ is in eigenvector with eigenvalue $\lambda = 1$. Let $(x,y) \in S$. Then using condition 5 of defintion 3.1 we get that the only possible nonzero values of $T\Psi$ are

$$T\Psi(x,y) = T((x,y),(x,y)) + T((x,y),(y,x))$$

and

$$T\Psi(y,x) = T((y,x),(x,y)) + T((y,x),(y,x)).$$

Because $(x,y) \in S$ it must be true that $T((x,y),(x,y)) = 1 - \frac{\pi(y,x)}{\pi(x,y)}$ and $T((x,y),(y,x)) = \frac{\pi(y,x)}{\pi(x,y)}$. Thus,

$$T\Psi(x,y) = 1 - \frac{\pi(y,x)}{\pi(x,y)} + \frac{\pi(y,x)}{\pi(x,y)} = 1.$$

Additionally, because $(x,y) \in S$ it must be that $(y,x) \notin S$. As such,

$$\frac{\pi_1(y)}{\pi_2(y)} > \frac{\pi_1(x)}{\pi_2(x)}.$$

By defintion 3.1 it must be the case that $T((y,x),(y,x)) = 0$ and $T((y,x),(x,y)) = 1$. So,

$$T\Psi(y,x) = 1.$$

Thus we have that $T\Psi = e_{(x,y)} + e_{(y,x)} = 1 \cdot \Psi$. We have one eigenvector of this form for each $(x,y) \in S$. Recalling the defintion of S, it follows from symmetry that half of the $|X|(|X|-1)$ pairs (x,y) with $x \neq y$ are in S. Thus there are $|X|(|X|-1)/2$ eigenvectors of the form $T\Psi = e_{(x,y)} + e_{(y,x)}$, each with eigenvalue $\lambda = 1$. Lastly, because each $e_{(x,y)}$ is orthgonal to every other basis vector, the eigenvectors in Λ_1 are orthogonal.

Now we show that $\Psi = -\frac{\pi(y,x)}{\pi(x,y)}e_{(x,y)} + e_{(y,x)}$ is an eigenvector with eigenvalue $\lambda = -\frac{\pi(y,x)}{\pi(x,y)}$. Again, let $(x,y) \in S$ using condition 5 of defintion 3.1 we get that the only possible nonzero values of $T\Psi$ are

$$-\frac{\pi(y,x)}{\pi(x,y)}T((x,y),(x,y)) + T((x,y),(y,x))$$

and

$$-\frac{\pi(y,x)}{\pi(x,y)}T((y,x),(x,y)) + T((y,x),(y,x)).$$

From defintion 3.1 we have $T((y,x),(y,x)) = 0$ and $T((y,x),(x,y)) = 1$ so

$$T\Psi(x,y) = -\frac{\pi(y,x)}{\pi(x,y)}\left(1 - \frac{\pi(y,x)}{\pi(x,y)}\right) + \frac{\pi(y,x)}{\pi(x,y)} = \left(\frac{\pi(y,x)}{\pi(x,y)}\right)^2.$$

Likewise,

$$T\Psi(y,x) = -\frac{\pi(y,x)}{\pi(x,y)}(1) + \frac{\pi(y,x)}{\pi(x,y)(0)} = -\frac{\pi(y,x)}{\pi(x,y)}.$$

Then we can compute

$$T\Psi = \left(\frac{\pi(y,x)}{\pi(x,y)}\right)^2 e_{(x,y)} - \frac{\pi(y,x)}{\pi(x,y)}e_{(y,x)} = -\frac{\pi(y,x)}{\pi(x,y)}\Psi.$$

Therefore $\Psi = -\frac{\pi(y,x)}{\pi(x,y)}e_{(x,y)} + e_{(y,x)}$ is an eigenvector with eigenvalue $\lambda = -\frac{\pi(y,x)}{\pi(x,y)}$. By the same symmetry argument as above there are $|X|(|X|-1)/2$ eigenvectors of this form. It is clear that $\Psi = -\frac{\pi(y,x)}{\pi(x,y)}e_{(x,y)} + e_{(y,x)}$ is not in the span of any standard basis vector. Additionally, the $e_{(x,y)}$ component of Ψ has opposite sign of the $e_{(y,x)}$ component. Thus it is not in the span of eigenvectors of the form $e_{(x,y)} + e_{(y,x)}$. Thus each $\Psi = -\frac{\pi(y,x)}{\pi(x,y)}e_{(x,y)} + e_{(y,x)}$ is a distinct eigenvector. Since we have constructed $|X|(|X|-1)/2 + |X|(|X|-1)/2 + |X| = |X|^2$ linearly independent eigenvectors we know that there are no more. $\square$

5. Bounding the Spectral Gap of Parallel Tempering

Lemma 5.1. *Let A and B be trasition matrices of metropolis chains with eigenvectors $\Psi_{A,i}$ and $\Psi_{B,i}$ corresponding to the ith largest eigenvalues $\lambda_{A,i}$ and $\lambda_{B,i}$, respectively. Let M and T be such that $MT = (A \otimes B)T$, the transition matrix of P. Additionally, let $\Psi_{P,i}$ be the eigenvector corresponding to the ith largest eigenvalue $\lambda_{P,i}$ of P. Finally, let $\Lambda_1 = \{\Psi_T : T\Psi_T = \Psi_T\}$. Then*

$$(\Psi_{A,1} \otimes \Psi_{B,2}) \notin \Lambda_1$$

and

$$(\Psi_{A,2} \otimes \Psi_{B,1}) \notin \Lambda_1$$

Proof. Assume for contradition $\Psi_{P,2} = \Psi_{A,1} \otimes \Psi_{B,2}$. Observe that because $M = A \otimes B$, by properties of the Kronecker product $\Psi_{A,1} \otimes \Psi_{B,2}$ is an eigenvector of M. Thus is suffices to show that for all x orthogonal to $\Psi_{P,1} = \mathbb{1}$, the vector $Tx \neq \Psi_{A,1} \otimes \Psi_{B,2}$. Because B is stochastic $\Psi_{B,2} \neq \mathbb{1}$ and there is $i,j \in X$ with $i \neq j$ such that

$$\Psi_{B,2}(i) \neq \Psi_{B,2}(j).$$

Then, because A is also stochastic $\Psi_{P,2}(\cdot, i) \neq \Psi_{P,2}(\cdot, j)$. In particular, we have that

$$\Psi_{P,2}(j,i) \neq \Psi_{P,2}(i,j).$$

However, by Lemma 4.1 these elements are not in the span of any vector of Λ_1 as the only eigenvectors in Λ_1 that with either $\Psi_{P,2}(j,i) \neq 0$ or $\Psi_{P,2}(i,j) \neq 0$ have that $\Psi_{P,2}(j,i) = \Psi_{P,2}(i,j)$. Thus $(\Psi_{A,1} \otimes \Psi_{B,2}) \notin \Lambda_1$. We now show $(\Psi_{A,2} \otimes \Psi_{B,1}) \notin \Lambda_1$

in the same way. Because A is stochastic $\Psi_{A,2} \neq \mathbb{1}$ and there is $i, j \in X$ with $i \neq j$ such that

$$\Psi_{A,2}(i) \neq \Psi_{A,2}(j).$$

Then, because B is also stochastic $\Psi_{P,2}(i, \cdot) \neq \Psi_{P,2}(j, \cdot)$. In particular, we have that

$$\Psi_{P,2}(i, j) \neq \Psi_{P,2}(j, i).$$

However, by Lemma 4.1 these elements are not in the span of any vector of Λ_1 as the only eigenvectors in Λ_1 that with either $\Psi_{P,2}(j, i) \neq 0$ or $\Psi_{P,2}(i, j) \neq 0$ have that $\Psi_{P,2}(j, i) = \Psi_{P,2}(i, j)$. Thus $(\Psi_{A,2} \otimes \Psi_{B,1}) \notin \Lambda_1$.   □

The last thing we need before we can prove the main result is the Min-Max Theorem of Rayleigh Quotients. This was proved in [?]. We state the result here.

Theorem 5.2. *Let A be Hermitian with eigenvalues $\lambda_1, \ldots, \lambda_k, \ldots, \lambda_n$. Then,*

$$\lambda_k = \min_{\substack{U \\ \dim(U)=k}} \max_{\substack{x \in U \\ x \neq 0}} \frac{||Ax||}{||x||}$$

With this we can now prove the main result of the paper.

5.1. **Main Result.**

Theorem 5.3. *Let λ_P^* be the second largest eigenvalue of the transition matrix of the parallel tempered chain P. Let λ_M^* be the second largest eigenvalue of the transition matrix M, the joint chain of two independent Metropolis chains. Then,*

$$\lambda_P^* < \lambda_M^*$$

Proof. By proposition 3.6 we know P is reversible. Thus it is similar to the matrix $\mathcal{P} = D_\pi^{1/2} P D_\pi^{-1/2}$ which is symmetric. By 3.9 and Theorem 5.2 we know

$$(7) \qquad\qquad \lambda_P^* = \max_{x \perp D_\pi^{1/2} \Psi_{P,1}} \frac{||D_\pi^{1/2} P D_\pi^{-1/2} x||}{||x||}.$$

Then we can multiply by the identity I as follows:

$$\begin{aligned} D_\pi^{1/2} P D_\pi^{-1/2} &= D_\pi^{1/2} MT D_\pi^{-1/2} \\ &= D_\pi^{1/2} MIT D_\pi^{-1/2} \\ &= D_\pi^{1/2} M D_\pi^{-1/2} D_\pi^{1/2} T D_\pi^{-1/2} \\ &= \mathcal{M}\mathcal{T} \end{aligned}$$

By Corollary 3.8, $\mathcal{M}$ and $\mathcal{T}$ are symmetric. Thus there norms are maximized by eigenvectors. In particular, we have

$$(8) \qquad\qquad \lambda_M^* = \max_{x \perp D_\pi^{1/2} \Psi_{M,1}} \frac{||\mathcal{M}x||}{||x||}.$$

Then, it follows from Lemma 5.1 that for all x one of the following is true. Either

$$||\mathcal{M}(\mathcal{T}x)|| = ||\mathcal{T}x|| \text{ and } ||\mathcal{T}x|| < ||x||$$

Or,

$$||\mathcal{T}x|| = ||x|| \text{ and } ||\mathcal{M}(\mathcal{T}x)|| < ||\mathcal{T}x||$$

In all cases,

$$||\mathcal{P}x|| = ||\mathcal{M}(\mathcal{T}x)|| < ||x|| = ||\mathcal{M}x||$$

It then follows from equations 7 and 8 that

$$\lambda_P^* < \lambda_M^*$$

$\square$

References

[1] RT Journal Article A1 Sambridge, Malcolm T1 A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization JF Geophysical Journal International JO Geophys J Int YR 2013 DO 10.1093/gji/ggt342 VO 196 IS 1 SP 357 OP 374 SN 0956-540X RD 5/24/2022 UL https://doi.org/10.1093/gji/ggt342

[2] Asja Fischer, Christian Igel, A bound for the convergence rate of parallel tempering for sampling restricted Boltzmann machines, Theoretical Computer Science, Volume 598, 2015, Pages 102-117, ISSN 0304-3975, https://doi.org/10.1016/j.tcs.2015.05.019. (https://www.sciencedirect.com/science/article/pii/S0304397515004235)

[3] Dawn Woodard. Scott Schmidler. Mark Huber. "Sufficient Conditions for Torpid Mixing of Parallel and Simulated Tempering." Electron. J. Probab. 14 780 - 804, 2009. https://doi.org/10.1214/EJP.v14-638