An Interlacing Theorem for Reversible Markov Chains

Robert Grone
Department of Mathematics and Statistics
San Diego State University
CA 92182-7720, San Diego, USA

Karl Heinz Hoffmann
Institut für Physik,
Technische Universität Chemnitz,
D-09107 Chemnitz, Germany

Peter Salamon
Department of Mathematics and Statistics
San Diego State University
CA 92182-7720, San Diego, USA

June 14, 2005

Abstract

Reversible Markov chains are an indispensable tool in the modelling of a vast class of physical, chemical, and biological problems. Examples include the master equation descriptions of relaxing physical systems, stochastic optimization algorithms like simulated annealing, and chemical dynamics of protein folding. Very often the large size of the state spaces requires the coarse graining or lumping of micro states into fewer mesoscopic states, and a question of utmost importance for the validity of the physical model is how the eigenvalues of the corresponding stochastic matrix change under this operation. In this paper we prove an interlacing theorem which gives explicit bounds on the eigenvalues of the lumped stochastic matrix.

1 Introduction

Reversible Markov chains are chains that satisfy the detailed balance condition (see equation below). Their importance and introduction to the physics literature dates back to the work of Ludwig Boltzmann (1887) who showed that
this class of Markov chains is the only class to be used for almost all physical processes at the molecular level. Later the principle also became known as the principle of microscopic reversibility. The importance of this class of Markov chains was further increased by the introduction of the Metropolis algorithm (1954) which is the basis for essentially all Monte Carlo simulations of physical systems. This algorithm makes all the “kinetic factors” as large as possible consistent with detailed balance and thus uses a designer Markov chain which has a particular stationary distribution, consistent with the requirement of having a given sparsity matrix and satisfying detailed balance. More recently (1984) this fact has been recognized to be much more generally useful and is the basis of Gibbs sampling exploited in the family of Markov chain Monte Carlo methods (MCMC) now in widespread use for many sorts of statistical simulations [3].

Another manifestation of reversible Markov chains arises from random walks on a simple graph on \( n \) vertices. Let \( A \) denote the adjacency matrix of a graph \( G \). By a random walk on \( G \), we mean a sequence of steps between adjacent vertices where each adjacent vertex is equally likely for the next step. This describes a Markov chain with matrix \( M = AD^{-1} \), where \( D \) is diagonal with \( d_{ii} \) being the degree of vertex \( i \) for all \( i = 1, 2, \ldots, n \). As we state in proposition 2 below, reversible Markov chains have matrices which are diagonally similar to symmetric matrices and thus many facts about symmetric matrices apply. The random walk matrix, \( M = AD^{-1} \), is similar to \( D^{-1/2}AD^{-1/2} \), which is clearly symmetric.

Our main result is an interlacing theorem. Given real numbers \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \) and \( \beta_1 \geq \beta_2 \geq \cdots \geq \beta_{n-1} \), we say the \( \beta \)'s interlace the \( \lambda \)'s if

\[
\lambda_1 \geq \beta_1 \geq \lambda_2 \geq \cdots \lambda_{n-1} \geq \beta_{n-1} \geq \lambda_n.
\] (1)

Perhaps the most well-known example of an interlacing theorem is the Cauchy Interlacing Theorem which states that the eigenvalues of an \((n-1)\)-by-\((n-1)\) principal submatrix of a symmetric matrix \( A \) interlace the eigenvalues of \( A \) [5].

In this paper we consider a reversible Markov chain on \( n \) states with (irreducible) transition matrix \( M \). We describe a “lumped” Markov chain on \( n - 1 \) states obtained by combining two states of \( M \). If \( \hat{M} \) is the transition matrix of the lumped Markov chain, we see that \( \hat{M} \) is reversible and that the eigenvalues of \( \hat{M} \) interlace those of \( M \).

Physical descriptions that forego many of the details of the real dynamics by lumping together collections of microstates into mesoscopic states is an old idea which is often referred to as coarse graining. Such coarse graining is a crucial ingredient in the modern description of many physical processes [4]. In particular, it is an indispensable tool for the study of complex energy landscapes [6]. Our result shows that such coarse graining can only speed up relaxation on these landscapes.
2 Reversible Markov Chains

Let $M$ be an $n$-by-$n$ irreducible column stochastic matrix, i.e. all entries non-negative and all column sums equal to one and there does not exist a permutation matrix $P$ such that $P^T M P$ is block triangular. The Markov chain with transition matrix $M$ is said to be reversible [1, 2] iff there exists a non-zero vector $v$ such that

$$M_{ij}v_j = M_{ji}v_i, \quad \forall \ i, j \quad (2)$$

If such a vector exists, it follows by summing (2) over $j$ that $v$ must be an eigenvector of $M$ with eigenvalue 1. Thus without loss of generality, we may assume that $v$ has been properly normalized, i.e. that $v > 0$ and $v_1 + \cdots + v_n = 1$. In other words, $v$ corresponds to the stationary distribution of the Markov chain corresponding to $M$. It also follows from equation (2) and our irreducible assumption that the Markov chain corresponding to $M$ is regular and so $\lambda = 1$ is a simple eigenvalue of $M$. As discussed above, condition (2) is also called detailed balance or microscopic reversibility in the physical literature [4]. If the chain induced by $M$ is reversible, then we will also say that $M$ is reversible. The following two propositions are easily shown.

Proposition 1: Let $V = \text{diag}(v_1^{1/2}, \ldots, v_n^{1/2})$. Then $M$ is reversible iff $V^{-1} M V$ is symmetric.

Proposition 2: If $M$ is reversible, then $M$ has all real eigenvalues, say

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n > -1 \quad (3)$$

3 Lumping states

One can form a (lumped) Markov chain on $n-1$ states by consolidating 2 states (say states 1 and 2) of a Markov chain with $n$ states [1]. If we start with the Markov chain with transition matrix $M$, the corresponding column stochastic matrix $\hat{M}$ of the lumped chain obtained from $M$ by lumping the first two states of $M$ into one state is obtained as follows

i. add row 1 of $M$ to row 2 of $M$, and then delete row 1 to form $\hat{M}$.

ii. replace column 2 of $\hat{M}$ by

$$d_1 \text{col}_1(\hat{M}) + d_2 \text{col}_2(\hat{M})$$

where

$$d_1 = \frac{v_1}{v_1 + v_2}, \quad d_2 = \frac{v_2}{v_1 + v_2}$$

and then delete column 1 from $\hat{M}$ to form $\hat{M}$. 

3
It is clear that \( \hat{M} \) is \((n-1)\)-by-(\(n-1\)) and column stochastic. In matrix terms

\[
\hat{M} = CMD
\]  

(4)

where

the \( C \) is the \((n-1)\)-by-\(n\) **collecting matrix**

\[
C = \begin{pmatrix}
1 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
0 & 0 & \ldots & 0 & 1
\end{pmatrix}
\]  

(5)

and \( D \) is the \(n\)-by-(\(n-1\)) **distributing matrix**

\[
D = \begin{pmatrix}
d_1 & 0 & \ldots & 0 \\
d_2 & 0 & \ldots & 0 \\
0 & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \ldots & 0 & 1
\end{pmatrix}
\]  

(6)

where \( d_1 = v_1/(v_1 + v_2) \) and \( d_2 = v_2/(v_1 + v_2) \).

**Proposition 3:**

\( \hat{M} \) is reversible with stationary distribution \( Cv = (v_1 + v_2, v_3, \ldots, v_n)^T \).

**Proof:**

It is clear that for \( i, j > 1 \) condition (2) is satisfied for \( \hat{M} \) and \( Cv \). For \( i = 1 \) we have from the reversibility of \( M \) and the definition of \( \hat{M} \) that

\[
\hat{M}_{i,1}(Cv)_1 &= (d_1M_{i,1} + d_2M_{i,2})(v_1 + v_2) \\
&= v_1M_{i,1} + v_2M_{i,2} \\
&= M_{i,1}v_1 + M_{2,i}v_i \\
&= \hat{M}_{1,i}(Cv)_i.
\]  

(7)  

(8)  

(9)  

(10)
4 The Interlacing Theorem

Let $M$ be an $n$-by-$n$ column stochastic matrix corresponding to a reversible Markov chain, and let $\hat{M}$ be formed as previously described. Then $\hat{M}$ corresponds to a reversible Markov chain with stationary distribution $Cv$. It follows that $\hat{M}$ has all real eigenvalues, say

\[ 1 = \beta_1 \geq \beta_2 \geq \cdots \geq \beta_{n-1} > -1. \]  

(11)

Our main result is the following

Theorem:

The eigenvalues of $\hat{M}$ interlace those of $M$

\[ 1 = \lambda_1 = \beta_1 > \lambda_2 \geq \beta_2 \geq \lambda_3 \geq \cdots \geq \beta_{n-1} \geq \lambda_n. \]  

(12)

One implication of our theorem is that $\hat{M}^k$ converges to its limit faster than $M^k$ since convergence is determined by the rates at which $\alpha^k$ and $\hat{\alpha}^k$ converge to zero, where

\[ \alpha = \max\{|\lambda_2|, \ldots, |\lambda_n|\} \]  

(13)

and

\[ \hat{\alpha} = \max\{|\beta_2|, \ldots, |\beta_n|\} \]  

(14)

Proof:

The proof requires some of the same considerations that make the proof of the Cauchy interlacing theorem for symmetric matrices complicated: we first argue that the result (12) holds for generic $M$ having distinct eigenvalues and satisfying an additional genericity condition (??). The general case then follows by a continuity argument.

Since $V^{-1}MV$ is symmetric, we may write

\[ V^{-1}MV = QAQ^T \]  

(15)

where

\[ \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \]  

(16)

and $Q$ is real orthogonal. We assume without loss of generality that $\det(Q) = 1$.

Now

\[ \hat{M} = CMD \]  

(17)

\[ = CVQ\Lambda Q^TV^{-1}D \]  

(18)

and so

\[ \lambda I_{n-1} - \hat{M} = CVQ(\lambda I_n - \Lambda)Q^TV^{-1}D \]  

(19)

since $QQ^T = VV^{-1} = I_n$ and $CD = I_{n-1}$.
If we take the \((n - 1)\)th compound \([5]\) in equation (19) we obtain
\[
\det[\lambda I_{n-1} - \hat{M}] = \mathcal{C}_{n-1}[\lambda I_{n-1} - \hat{M}] = \mathcal{C}_{n-1}[CVQ(\lambda I_n - \Lambda)Q^TV^{-1}D] = \mathcal{C}_{n-1}[CVQ][\lambda I_{n-1} - \Lambda]\mathcal{C}_{n-1}[Q^TV^{-1}D].
\]
(23)

Now for simplicity define
\[
g(\lambda) = \det[\lambda I_{n-1} - \hat{M}] = \prod_{i=1}^{n-1} (\lambda - \beta_i)
\]
(24)
\[
f(\lambda) = \det[\lambda I_n - \Lambda] = \prod_{i=1}^{n} (\lambda - \lambda_i)
\]
(25)
\[
x = \mathcal{C}_{n-1}[CVQ]
\]
(26)
\[
y = \mathcal{C}_{n-1}[Q^TV^{-1}D].
\]
(27)

Let
\[
f_j(\lambda) = \frac{f(\lambda)}{(\lambda - \lambda_j)} = \prod_{i=1,i\neq j}^{n} (\lambda - \lambda_i)
\]
(28)

Since \(\lambda I_n - \Lambda\) is diagonal, it is clear that
\[
\mathcal{C}_{n-1}[\lambda I_n - \Lambda] = \begin{pmatrix}
  f_n(\lambda) & 0 & \cdots & 0 \\
  0 & f_{n-1}(\lambda) & \cdots & 0 \\
  \vdots & \vdots & \ddots & 0 \\
  0 & 0 & \cdots & f_1(\lambda)
\end{pmatrix}
\]
(29)

We then have
\[
g(\lambda) = x \begin{pmatrix}
  f_n(\lambda) & 0 & \cdots & 0 \\
  0 & f_{n-1}(\lambda) & \cdots & 0 \\
  \vdots & \vdots & \ddots & 0 \\
  0 & 0 & \cdots & f_1(\lambda)
\end{pmatrix} y,
\]
so
\[
g(\lambda) = \sum_{i=1}^{n} x_i y_i f_{n+1-i}(\lambda).
\]
(31)

To continue, we make use of the following lemmas

**Lemma 1:** \(x = ky^T\) for \(k = \frac{v_2\cdots v_n}{v_1+v_2} > 0\).

**Lemma 2:** \(x_n = 0\).

**Lemma 3:** Let \(0 \neq w \in W\), where \(W\) is a real \(m\)-dimensional inner product space. Then there exists an orthonormal basis of \(W\), say \(\{w_1, \ldots, w_m\}\) such that \((w, w_i) \neq 0\), all \(i = 1, \ldots, m\).
Lemma 4: It suffices to prove the theorem under the genericity assumption that
\(x_i \neq 0\), all \(i = 1, \ldots, n-1\).

Lemma 5: It suffices to prove the theorem under the genericity assumption that
\(\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_n\).

We leave the proof of these lemmas till last and proceed with the proof of
the theorem. Using Lemmas 1 and 2, equation (31) becomes

\[
g(\lambda) = \sum_{i=1}^{n-1} k x_i^2 f_{n+1-i}(\lambda)
\]

so \(g(\lambda)\) is a positive linear combination of \(f_2(\lambda), \ldots, f_n(\lambda)\). Clearly,

\[
f_i(\lambda_j) = 0, \quad \text{all } i \neq j
\]

and

\[
f_i(\lambda_i) = \prod_{j=1, j \neq i}^{n} (\lambda_i - \lambda_j)
\]

From this last fact and the strict inequality among the \(\lambda\)'s in Lemma 5, we have

\[
f_1(\lambda_1) > 0
\]
\[
f_2(\lambda_2) < 0
\]
\[
f_3(\lambda_3) > 0
\]
\[
f_4(\lambda_4) < 0
\]
\[
\ldots
\]

Since \(g(\lambda)\) is a positive linear combination of \(f_2(\lambda), \ldots, f_n(\lambda)\), this yields

\[
g(\lambda_1) = 0
\]
\[
g(\lambda_2) < 0
\]
\[
g(\lambda_3) > 0
\]
\[
g(\lambda_4) < 0
\]
\[
\ldots
\]

By the Intermediate Value Theorem, \(g(\lambda)\) must have at least one zero in each
of the intervals

\[
(\lambda_n, \lambda_{n-1}), (\lambda_{n-1}, \lambda_{n-2}), \ldots, (\lambda_3, \lambda_2).
\]

Since \(\deg(g(\lambda)) = n - 1\), and \(\beta_1 = 1\) is known, this accounts for all the other
\(n - 2\) roots of \(g(\lambda)\), and we have shown that

\[
1 = \lambda_1 = \beta_1 > \lambda_2 > \beta_2 > \lambda_3 > \cdots > \beta_{n-1} > \lambda_n
\]

This completes the proof of the theorem. We now present proofs of the
lemmas.
Proof of Lemma 1: $x = ky^T$

We have

$$x = C_{n-1}[CVQ]$$  \hspace{1cm} (47)

$$= C_{n-1}[CV][C_{n-1}[Q]]$$  \hspace{1cm} (48)

$$= C_{n-1}\begin{pmatrix} v_1^{1/2} & v_2^{1/2} & 0 & \cdots & 0 \\ 0 & 0 & v_3^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & v_n^{1/2} \end{pmatrix} C_{n-1}[Q]$$  \hspace{1cm} (49)

$$= (0, \ldots, 0, w_{n-1}, w_n) C_{n-1}[Q]$$  \hspace{1cm} (50)

where

$$w_{n-1} = v_1^{1/2}(v_3 \ldots v_n)^{1/2}$$  \hspace{1cm} (52)

and

$$w_n = v_2^{1/2}(v_3 \ldots v_n)^{1/2}$$  \hspace{1cm} (53)

Similarly,

$$y = C_{n-1}[Q^TV^{-1}D]$$  \hspace{1cm} (54)

$$= C_{n-1}[Q^T][C_{n-1}[V^{-1}D]]$$  \hspace{1cm} (55)

$$= C_{n-1}[Q^T]\begin{pmatrix} d_1 v_1^{-1/2} & 0 & \cdots & 0 \\ d_2 v_2^{-1/2} & 0 & \cdots & 0 \\ 0 & v_3^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n^{-1/2} \end{pmatrix}$$  \hspace{1cm} (56)

$$= C_{n-1}[Q^T](0, \ldots, 0, z_{n-1}, z_n)^T$$  \hspace{1cm} (57)

where

$$z_{n-1} = d_1 v_1^{-1/2}(v_3 \ldots v_n)^{-1/2} = \frac{v_1^{1/2}}{v_1 + v_2} (v_3 \ldots v_n)^{-1/2}$$  \hspace{1cm} (58)

and

$$z_n = d_2 v_2^{-1/2}(v_3 \ldots v_n)^{-1/2} = \frac{v_2^{1/2}}{v_1 + v_2} (v_3 \ldots v_n)^{-1/2}$$  \hspace{1cm} (59)

If we let

$$k = \frac{v_3 \ldots v_n}{v_1 + v_2}$$  \hspace{1cm} (60)
we see $kz = w^T$. Now

$$
x^T = (wC_{n-1}[Q])^T \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad MATHEMATICAL

\text{Proof of Lemma 2: } x_n = 0

Recall that $x = C_{n-1}[CVQ]$, so $x = C_{n-1}(C) C_{n-1}(V) C_{n-1}(Q)$. It is easy to see that

$$C_{n-1}(C) = (0, \ldots, 0, 1, 1)$$

and

$$C_{n-1}(V) = (v_1 v_2 \ldots v_n)^{1/2}$$

Also, from fact 5 in our appendix,

$$C_{n-1}(Q) = P^T Q^T P,$$

which has the form

$$
\begin{bmatrix}
\pm v_n^{1/2} \\
\mp v_n^{1/2} \\
\vdots \\
\star \\
\pm v_{n-1}^{1/2} \\
- v_2^{1/2} \\
\cdots \\
- v_2^{1/2} \\
v_1^{1/2}
\end{bmatrix}
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad MATHEMATICAL
Now

\[
x = C_{n-1}(C) C_{n-1}(V) C_{n-1}(Q)
\]

\[
= (v_1 v_2 \ldots v_n)^{1/2} (0, \ldots, 0, 1, 1)
\begin{bmatrix}
  v_n^{-1/2} & 0 & 0 \\
  0 & v_{n-1}^{-1/2} & 0 \\
  \vdots & \ddots & \vdots \\
  0 & \cdots & v_1^{-1/2}
\end{bmatrix} C_{n-1}(Q)
\]

\[
= (v_1 v_2 \ldots v_n)^{1/2} (0, \ldots, v_2^{-1/2}, v_1^{-1/2}) C_{n-1}(Q)
\]

\[
= (v_1 v_2 \ldots v_n)^{1/2} (0, \ldots, 0, v_2^{-1/2}, v_1^{-1/2})
\begin{bmatrix}
  \pm v_n^{1/2} \\
  \mp v_{n-1}^{1/2} \\
  \vdots \\
  v_2^{1/2} \\
  v_1^{1/2}
\end{bmatrix}
\]

\[
= (v_1 v_2 \ldots v_n)^{1/2} (*, *, \ldots, *, 0).
\]

Proof of Lemma 3:

We may assume without loss of generality that \( \|w\| = 1 \). If we extend \( w \) to an arbitrary orthonormal basis of \( W \), say \( \{y_1, y_2, \ldots, y_m\} \) with \( w = y_1 \), we may assume that \( W = \mathbb{R}^m \), and that \( w = e_1 = (1, 0, \ldots, 0)^T \). Now let \( P \) be any \( m \times m \) orthogonal matrix with no zero entries in the first column. If the rows of \( P \) are \( r_1, \ldots, r_m \), we may let \( y_i = r_i^T \), for all \( i = 1, \ldots, m \).

Proof of Lemma 4:

We saw in the proof of lemma 3 that \( x \) is a positive multiple of

\[
(0, \ldots, 0, v_2^{-1/2}, v_1^{-1/2}) P^T Q^T P,
\]

where \( P^T Q^T P \) is an \( n \times n \) orthogonal matrix with last column

\[
y_n = (\pm v_n^{1/2}, \ldots, v_3^{1/2}, -v_2^{1/2}, v_1^{1/2})^T.
\]

In lemma 2, this was used to show that \( x_n = 0 \), or that \( (0, \ldots, 0, v_2^{-1/2}, v_1^{-1/2})^T \) is orthogonal to \( y_n \). Now apply lemma 3 to \( W = < y_n >^T \) and \( w = (0, \ldots, 0, v_2^{-1/2}, v_1^{-1/2})^T \in W \). We deduce that there exists an orthonormal basis of \( W \), say \( \{y_1, \ldots, y_{n-1}\} \) where \( (w, y_i) \neq 0 \), for all \( i = 1, \ldots, n-1 \). Since \( \{y_1, \ldots, y_n\} \) is orthonormal, the matrix \( Y = [y_1 \ y_2 \ldots \ y_n] \) is orthogonal. If we define \( \hat{Q} = P^T Y^T P \), then \( \hat{Q} \) is an \( n \times n \) orthogonal matrix sharing the first column of \( Q \), but for which

\[
\hat{x} = C_{n-1}(CV \hat{Q})
\]

\[
= (\hat{x}_1, \ldots, \hat{x}_{n-1}, 0),
\]

where

\[
\hat{x}_i = (w, y_i) \neq 0, \quad i = 1, \ldots, n-1.
\]
Now let $\hat{Q}$ denote any orthogonal matrix having first column the same as $Q$ but with $\hat{x}_i \neq 0$, all $i = 1, \ldots, n - 1$, and define

$$\hat{M} = V \hat{Q} \Lambda \hat{Q}^T V^{-1}. \quad (75)$$

Then $\hat{M}$ has the same eigenvalues as $M$, and has the same eigenvector corresponding to $\lambda_1 = 1$. (Note that $\hat{M}$ may have negative entries, but its columns still add to 1). Among this set of matrices, the set of $\hat{M}$ for which $\hat{x}_i \neq 0$, all $i = 1, \ldots, n - 1$ is a dense subset. Thus it suffices to prove our result under the generic assumption $x_i \neq 0$, all $i = 1, \ldots, n - 1$.

Proof of Lemma 5:

Recall that $M = VQA^T V^{-1}$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $\lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n$. For any $\epsilon < \frac{\lambda_1 - \lambda_2}{n-1}$, let

$$\Lambda_\epsilon = \text{diag}(\lambda_1, \lambda_2 + (n-1)\epsilon, \lambda_3 + (n-2)\epsilon, \ldots, \lambda_n + \epsilon) \quad (76)$$

and

$$M_\epsilon = V \Lambda_\epsilon \hat{Q}^T V^{-1}. \quad (77)$$

Since $M_\epsilon$ has distinct eigenvalues for all $\epsilon \in (0, \frac{\lambda_1 - \lambda_2}{n-1})$ and $M = M_0$, it suffices by continuity to prove the theorem for $M_\epsilon$, where $0 < \epsilon < \frac{\lambda_1 - \lambda_2}{n-1}$.

Corollary: When we lump $k$ states of a reversible Markov chain, the eigenvalues of the lumped chain interlace the eigenvalues of the original chain according to

$$\lambda_1 = \beta_1 = 1 \geq \lambda_2 \geq \beta_2 \geq \lambda_3 \geq \beta_3 \geq \lambda_4 \geq \beta_4 \geq \cdots \geq \lambda_{2+k-1} \geq \beta_{2+k-1} \geq \lambda_{3+k-1} \geq \beta_{3+k-1} \geq \cdots \quad (78)$$

The proof follows from the fact that lumping $k$ states at once gives the same chain as iterated lumping.

5 Conclusions

The main result of this work is an interlacing theorem for reversible Markov chains that results by lumping two states of a bigger chain. The theorem provides a solid tool for the analysis of a large class of Markov chain models in physics and other areas. It gives strict bounds on the relaxation dynamics for a multitude of processes which can be used as coarse grained or lumped approximations of the underlying process. The realm of irreversible decay towards an equilibrium or stationary state rests on the use of Markov chain descriptions. With the typical $10^{23}$ particles per cubic centimeter and the corresponding number of degrees of freedom, it is clear that coarse grained descriptions which reduce the dimensionality of the problem are necessary. Our theorem provides
direct bounds on the eigenvalues of the lumped problem, and thus provides a rigorous proof of limits for the underlying dynamics. Note that our theorem bounds eigenvalues of the lumped dynamics from above and below. Depending on the structure of the original system these could be quite restrictive thus giving a small range of possible time scales for the lumped dynamics.

For large numbers of sequential lumping steps, the relevant eigenvalues might change considerably. On the other hand, for cases where symmetries in the system lead to the dynamical equivalence of states the eigenvalues might change very little. In some disordered physical systems Markov chain techniques have revealed certain “dynamical degeneracies” of states [7]. The behavior of such systems under lumping is a highly interesting yet unsolved field. While the present interlacing theorem provides partial information regarding these open problems stronger interlacing theorems along the lines of Weyl’s theorem bounding the eigenvalues of the unlumped chain to the eigenvalues of the dynamic components will be explored in a future effort.

Finally we mention that for the designer Markov chains developed in different areas as tools for a variety of problems like finding global minima of multi-minima functions the interlacing theorem provides a tool to develop fast converging algorithms. Depending on the structure, clever coarse graining of the state space might allow much faster convergence compared to standard methods.

6 Appendix

Let $A$ and $B$ be $m$-by-$n$ matrices, and let $1 \leq k \leq \min\{m, n\}$. The $k$th minor (compound) of $A$ is the $\binom{m}{k}$-by-$\binom{n}{k}$ matrix of $k$-by-$k$ subdeterminants of $A$. They have the following properties [5]:

1. $C_1[A] = A$
2. If $m = n$, $C_n[A] = \det A$
5. If $A$ is $n$-by-$n$ and invertible, then

$$C_{n-1}(A) = \det(A)P^TA^{-1}P,$$

where

$$P = \begin{bmatrix} 0 & \ldots & 0 & 0 & -1 \\ 0 & \ldots & 0 & +1 & 0 \\ 0 & \ldots & -1 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \pm 1 & 0 & \ldots & \ldots & 0 \end{bmatrix}.$$
References


