An interlacing theorem for reversible Markov chains

FAST TRACK COMMUNICATION

An interlacing theorem for reversible Markov chains

Robert Grone[1], Karl Heinz Hoffmann[2] and Peter Salamon[1]

[1] Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182-7720, USA
[2] Institut für Physik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany

Abstract
Reversible Markov chains are an indispensable tool in the modeling of a vast class of physical, chemical, biological and statistical problems. Examples include the master equation descriptions of relaxing physical systems, stochastic optimization algorithms such as simulated annealing, chemical dynamics of protein folding and Markov chain Monte Carlo statistical estimation. Very often the large size of the state spaces requires the coarse graining or lumping of microstates 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.

PACS numbers: 02.50.Ga, 02.70.Tt, 02.10.Yn, 82.20.Wt

1. Introduction

Reversible Markov chains are chains that satisfy the detailed balance condition (see equation (3) 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, through the work of Lars Onsager (1931), the condition also became known as the *principle of microscopic reversibility* [1, 2]. The importance of this class of Markov chains was further increased by the introduction of the Metropolis algorithm (1953) which is the basis for essentially all Monte Carlo simulations of physical systems [3, 4]. This algorithm makes all the 'kinetic factors' [5, 6] as large as possible consistent with detailed balance and a given sparsity matrix, thereby creating a designer Markov chain which has a desired stationary distribution and relaxes as quickly as possible. More recently (1984) this fact has been recognized to be much more generally useful and is the basis of Gibbs sampling and the Metropolis–Hastings algorithm exploited in the family of Markov chain Monte Carlo methods (MCMC) now in widespread use for many types of statistical simulations [7, 8].

J. Phys. A: Math. Theor. **41** (2008) 212002   Fast Track Communication

IOP FTC ▶▶▶

Another manifestation of reversible Markov chains arises from random walks on a simple graph of 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 transition probability 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.

As a final example, we consider the case when the transition matrix is symmetric. This also corresponds to a Markov chain which is reversible as is immediate from condition (3) below. In this case the matrix is bistochastic (rows and columns sum to one) and the stationary distribution is uniform.

Our main result is an interlacing theorem. Given real numbers $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ and $\beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_m$, we say β's *interlace* λ's if

$$\lambda_i \geqslant \beta_i \geqslant \lambda_{n-m+i}, \qquad \text{for} \quad i = 1, \ldots, m. \tag{1}$$

The choice of the nomenclature is clearest for the case $m = n - 1$, for which equation (1) assumes the form

$$\lambda_1 \geqslant \beta_1 \geqslant \lambda_2 \geqslant \beta_2 \geqslant \cdots \lambda_{n-1} \geqslant \beta_{n-1} \geqslant \lambda_n. \tag{2}$$

The most well-known example of an interlacing theorem is the Cauchy interlacing theorem which states that the eigenvalues of an $m \times m$ principal submatrix of a symmetric $n \times n$ matrix A interlace the eigenvalues of A [9–11].

In this paper we consider a reversible Markov chain on n states with (irreducible) transition matrix M. We prove an interlacing theorem between the eigenvalues of M and the *lumped* Markov chain obtained by partitioning the states of M into disjoint subsets and considering transitions among these subsets [12–15]. Such lumped chains associated with a given Markov chain and a partition of its states have surfaced in a number of contexts recently under various names. Jerrum *et al* [16] refer to the lumped chain as the *projection* chain and combine Poincaré and log-Sobolev constants of the lumped chain with restriction chains on each of the sets in the partition to find Poincaré and log-Sobolev constants of the unlumped chain. These constants characterize non-asymptotic rates of convergence associated with the degree of mixing. Lumped chains have also been called *aggregate* chains [17] in the context of nearly completely decomposable (NCD) chains. Haemers [11] considers a closely related concept called the *quotient* matrix of a symmetric matrix and a partition in the context of interlacing theorems for combinatorics on graphs. In the physical literature, lumped processes have a long history under the name of coarse graining. Physical descriptions that forego many of the details of the real dynamics by lumping together collections of *microstates* into *mesoscopic* states are 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 [18]. In particular, it is an indispensable tool for the study of complex energy landscapes [19]. Our result shows that such coarse graining can only speed up relaxation on these landscapes.

Our interlacing theorem follows from the special property of reversible Markov chains mentioned above that their transition probability matrices are diagonally similar to a symmetric matrix. This fact enables us to map our problem onto the Cauchy interlacing theorem. Thus the resulting interlacing theorem for reversible Markov chains may be viewed as a corollary to the strengthened form of the Cauchy interlacing theorem found in [10, 11].

There is a large literature concerning bounds on the second largest modulus of an eigenvalue in a Markov chain [20]. This modulus is known as the coefficient of ergodicity of the chain and is given by $\xi = \max\{|\lambda_i|, i = 2, \ldots, n\} = \max\{|\lambda_2|, |\lambda_n|\}$. Closely related

J. Phys. A: Math. Theor. **41** (2008) 212002

IOP FTC ▶▶▶

Fast Track Communication

to this is the spectral gap $1 - \xi$ of the stochastic matrix. The interest derives from the fact that such bounds characterize the asymptotic rate of convergence of the chain to its stationary distribution. There is also a large literature concerning bounds on non-asymptotic convergence measures characterizing various degrees of mixing in the chain [16, 21–23]. The statement and proof of the fact that the lumped chain's eigenvalues interlace the eigenvalues of the full chain is presented here for the first time.

2. Reversible Markov chains

Let M be an $n \times n$ column stochastic matrix of a regular Markov chain, i.e. all entries non-negative, all columns equal to one and there exists a positive integer k such that M^k has all positive entries. The Markov chain with transition matrix M is said to be *reversible* [12, 13] iff there exists a non-zero vector v such that

$$M_{ij}v_j = M_{ji}v_i, \qquad \forall i, j. \tag{3}$$

If such a vector exists, it follows by summing (3) over j that v must be the 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. Note that the regularity condition assures that $\lambda = 1$ is a simple eigenvalue of M. As discussed above, condition (3) is also called detailed balance or microscopic reversibility in the physical literature [18]. 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 $V = diag\big(v_1^{1/2}, \ldots, v_n^{1/2}\big)$. Then M is reversible iff $V^{-1}MV$ is symmetric.*

Proposition 2. *If M is reversible, then M has all real eigenvalues, say*

$$1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_n > -1. \tag{4}$$

3. Lumping states

Our interest is in the lumped matrix corresponding to an arbitrary partition of the states of the chain. To keep the exposition as transparent as possible, we begin with a partition whose subsets include only one doubleton and the rest all singletons. By iteratively lumping two states at a time, we can reach any desired partition and thus our procedure is perfectly general besides being more transparent.

Now consider lumping the two states in the doubleton (say states 1 and 2). If we start with the Markov chain with transition matrix M, the corresponding column stochastic matrix $\hat{M}$ of the lumped chain that results 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 $\tilde{M}$.
(ii) replace column 2 of $\tilde{M}$ by

$$d_1 \mathrm{col}_1(\tilde{M}) + d_2 \mathrm{col}_2(\tilde{M}),$$

where

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

and then delete column 1 from $\tilde{M}$ to form $\hat{M}$. Physically, this corresponds to summing transition probabilities into the lumped state and averaging transition probabilities of leaving the lumped state.

J. Phys. A: Math. Theor. **41** (2008) 212002   Fast Track Communication

IOP FTC ▶▶▶

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

$$\hat{M} = CMD, \tag{5}$$

where C is the $(n-1) \times n$ *collecting matrix*

$$C = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \tag{6}$$

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

$$D = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ d_2 & 0 & \ldots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix}, \tag{7}$$

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* $\hat{v} = Cv = (v_1 + v_2, v_3, \ldots, v_n)^T$.

Proof. It is clear that for $i, j > 1$ condition (3) is satisfied for $\hat{M}$ and Cv. Let $j = 1$ and $i > 1$. Then from the reversibility of M and the definition of $\hat{M}$

$$\hat{M}_{i,1}(Cv)_1 = (d_1 M_{i,1} + d_2 M_{i,2})(v_1 + v_2) \tag{8}$$

$$= v_1 M_{i,1} + v_2 M_{i,2} \tag{9}$$

$$= M_{1,i} v_i + M_{2,i} v_i \tag{10}$$

$$= \hat{M}_{1,i}(Cv)_i. \tag{11}$$

$\square$

4. The interlacing theorem

Let M be an $n \times n$ 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 $\hat{v} = Cv$. It follows that $\hat{M}$ has all real eigenvalues, say

$$1 = \beta_1 > \beta_2 \geqslant \cdots \geqslant \beta_{n-1} > -1. \tag{12}$$

Our main result is the following.

Theorem. *The eigenvalues of* $\hat{M}$ *interlace those of* M

$$1 = \lambda_1 = \beta_1 > \lambda_2 \geqslant \beta_2 \geqslant \lambda_3 \geqslant \cdots \geqslant \beta_{n-1} \geqslant \lambda_n. \tag{13}$$

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* α^k *and* $\hat{\alpha}^k$ *converge to zero, where*

$$\alpha = \max\{|\lambda_2|, \ldots, |\lambda_n|\} \tag{14}$$

and

$$\hat{\alpha} = \max\{|\beta_2|, \ldots, |\beta_{n-1}|\}. \tag{15}$$

J. Phys. A: Math. Theor. **41** (2008) 212002   Fast Track Communication

IOP FTC ▶▶▶

Proof. Our proof relies on a variant of the Cauchy interlacing theorem which states that if A is an $n \times n$ symmetric matrix and Q is an $n \times (n-1)$ matrix with orthonormal columns, then the eigenvalues of $Q^T A Q$ interlace the eigenvalues of A. This variant follows by completing Q to an orthogonal matrix $\tilde{Q}$ and noting that $Q^T A Q$ is a principal submatrix of $\tilde{Q}^T A \tilde{Q}$.

Recall that $S = V^{-1} M V$ is symmetric as is $\hat{S} = \hat{V}^{-1} \hat{M} \hat{V}$, where we have set $\hat{V} = \mathrm{diag}(\sqrt{v_1 + v_2}, \sqrt{v_3}, \ldots, \sqrt{v_n})$. Inserting the definition of $\hat{M}$, we have

$$\hat{S} = \hat{V}^{-1} \hat{M} \hat{V} \tag{16}$$

$$= \hat{V}^{-1} C M D \hat{V} \tag{17}$$

$$= (\hat{V}^{-1} C V)(V^{-1} M V)(V^{-1} D \hat{V}) \tag{18}$$

$$= (\hat{V}^{-1} C V) S (V^{-1} D \hat{V}). \tag{19}$$

Defining $Q_1 = \hat{V}^{-1} C V$, we calculate directly that

$$
Q_1 = \begin{pmatrix} \frac{1}{\sqrt{v_1+v_2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{v_3}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{v_n}} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{v_1} & 0 & \cdots & 0 \\ 0 & \sqrt{v_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{v_n} \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{\sqrt{v_1}}{\sqrt{v_1+v_2}} & \frac{\sqrt{v_2}}{\sqrt{v_1+v_2}} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{20}
$$

Similarly, setting $Q_2 = V^{-1} D \hat{V}$, we get

$$
Q_2 = \begin{pmatrix} \frac{1}{\sqrt{v_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{v_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{v_n}} \end{pmatrix} \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ d_2 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{v_1+v_2} & 0 & \cdots & 0 \\ 0 & \sqrt{v_3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{v_n} \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{\sqrt{v_1}}{\sqrt{v_1+v_2}} & 0 & \cdots & 0 \\ \frac{\sqrt{v_2}}{\sqrt{v_1+v_2}} & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{21}
$$

We note that $Q_1 = Q_2^T$ and $Q_1 Q_2 = I_{n-1}$. Thus the variant of the Cauchy interlacing theorem stated above applies and the eigenvalues of $\hat{M}$ interlace those of M. Since M and $\hat{M}$ are both stochastic matrices, the largest eigenvalues are both equal to 1. □

J. Phys. A: Math. Theor. **41** (2008) 212002

IOP FTC ▶▶▶

Fast Track Communication

Corollary. *When we lump a reversible Markov chain to a partition with m subsets, the eigenvalues of the lumped chain interlace the eigenvalues of the original chain according to*

$$
\begin{aligned}
&\lambda_1 = \beta_1 = 1 \\
&\lambda_2 \geqslant \beta_2 \geqslant \lambda_{2+n-m} \\
&\cdots \\
&\lambda_i \geqslant \beta_i \geqslant \lambda_{i+n-m} \\
&\cdots
\end{aligned}
\tag{22}
$$

The proof follows from the fact that lumping $k = n - m + 1$ 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 states of a bigger chain. One important implication of our theorem is that coarse graining can only increase the rate of relaxation toward equilibrium. More generally, our theorem provides a solid tool for the analysis of a large class of Markov 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 toward an equilibrium or stationary state rests on the use of Markov chain descriptions. With typically 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 lumping many states, 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 [24]. The behavior of such systems under lumping is a highly interesting yet largely unexplored field.

For physical understanding of the processes studied using the Markov chain, one typically wants to choose the subsets in the partition so as to disrupt the relaxation spectrum as little as possible. How this can be done generally is an open problem; how to carry this out for concrete examples is usually based on physical reasoning. For rapid convergence of MCMC algorithms on the other hand, the aim is the opposite—we want to find a version of the algorithm with fastest convergence. For non-asymptotic convergence, some progress along these lines has been realized [16, 23].

References

[1] Onsager L 1931 Reciprocal relations in irreversible processes *Phys. Rev.* **37** 405–26
[2] Casimir H B G 1945 On Onsager's principle of microscopic reversibility *Rev. Mod. Phys.* **17** 343–50
[3] Metropolis N, Rosenbluth A, Teller A and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92
[4] Landau D P and Binder K 2000 *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge: Cambridge University Press)

J. Phys. A: Math. Theor. **41** (2008) 212002

IOP FTC ▶▶▶

Fast Track Communication

[5] Hoffmann K H, Schubert S and Sibani P 1999 Age reinitialization in hierarchical relaxation models for spin-glass dynamics *Europhys. Lett.* **38** 613–8

[6] Hoffmann K H and Schoen C 2005 Kinetic features of preferential trapping on energy landscapes *Found. Phys. Lett.* **18** 171–82

[7] Geman S and Geman D 1984 Gibbs distributions and the Bayesian restoration of images *IEEE Trans. PAMI* **6** 721–41

[8] Gilks W R, Richardson S and Spiegelhalter D J (ed) 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)

[9] Horn R A and Johnson C R 1985 *Matrix Analysis* (New York: Cambridge University Press)

[10] Courant R and Hilbert D 1924 *Methoden der Mathematischen Physik* (Berlin: Springer)

[11] Haemers W H 1995 Interlacing eigenvalues and graphs *Linear Algebra Appl.* **227–228** 593–616

[12] Burke C K and Rosenblatt M 1958 A Markovian function of a Markov chain *Ann. Math. Stat.* **29** 1112–22

[13] Kemeny J G and Snell J L 1960 *Finite Markov Chains* (Princeton, NJ: Van Nostrand-Reinhold)

[14] Andresen B, Hoffmann K H, Mosegaard K, Nulton J, Pedersen J M and Salamon P 1988 On lumped models for thermodynamic properties of simulated annealing problems *J. Phys. France* **49** 1485–92

[15] Salamon P, Sibani P and Frost R 2002 *Facts, Conjectures, and Improvements for Simulated Annealing* (Philadelphia, PA: SIAM)

[16] Jerum M, Son J-B, Tetali P and Vigoda E 2004 Elementary bounds on Poincaré and log-Sobolev constants for decomposable Marcov chains *Ann. Appl. Probab.* **14** 1741–65

[17] Avrachenko K and Haviv M 2004 The first Laurent series coefficients for singularly perturbed stochastic matrices *Linear Algebra Appl.* **386** 243–59

[18] Van Kampen N G 1992 *Stochastic Processes in Physics and Chemistry* (Amsterdam: North-Holland)

[19] Wales D J 2004 *Energy Landscapes: With Applications to Clusters, Biomolecules and Glasses* (New York: Cambridge University Press)

[20] Rothblum U G and Tan C P 1985 Upper bounds on the maximum modulus of subdominant eigenvalues of nonnegative matrices *Linear Algebra Appl.* **66** 45–86

[21] Diaconis P and Saloff-Coste L 1993 Comparison theorems for reversible Marcov chains *Ann. Appl. Probab.* **3** 696–730

[22] Diaconis P and Saloff-Coste L 1996 Logarithmic Sobolev inequalities for finite markov chains *Ann. Appl. Probab.* **6** 695–750

[23] Madras N and Randall D 2002 Marcov chain decomposition for convergence rate analysis *Ann. Appl. Probab.* **12** 581–606

[24] Sibani P, Schoen J C, Salamon P and Andersson J O 1993 Emergent hierarchical structures in complex system dynamics *Europhys. Lett.* **22** 479–85