Electronic Transactions on Numerical Analysis.
Volume 38, pp. 17-33, 2011.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ON AN SVD-BASED ALGORITHM FOR IDENTIFYING META-STABLE STATES OF MARKOV CHAINS*

RYAN M. TIFENBACH†

Abstract. A Markov chain is a sequence of random variables $X = \{x_t\}$ that take on values in a state space $\mathcal{S}$. A meta-stable state with respect to X is a collection of states $\mathcal{E} \subseteq \mathcal{S}$ such that transitions of the form $x_t \in \mathcal{E}$ and $x_{t+1} \notin \mathcal{E}$ are exceedingly rare. In Fritzsche et al. [Electron. Trans. Numer. Anal., 29 (2008), pp. 46–69], an algorithm is presented that attempts to construct the meta-stable states of a given Markov chain. We supplement the discussion contained therein concerning the two main results.

Key words. Markov chains; conformation dynamics; singular value decomposition

AMS subject classifications. 15A18, 15A51, 60J10, 60J20, 65F15

1. Preliminaries. There is a great deal of interest in the problem of identifying so-called meta-stable states of Markov chains. Let $\{x_t\}$ be a discrete time, time-homogeneous Markov chain with finite state space $\mathcal{S}$; a *meta-stable state* is a proper subcollection of states $\mathcal{E} \subset \mathcal{S}$ such that transitions from $\mathcal{E}$ to $\mathcal{S} \setminus \mathcal{E}$ are exceedingly rare. Different researchers have used different measures to define such collections; see, for example, the *uncoupling measure* defined in [7] or the *coupling matrix* defined in [3]. For the purposes of this note, we will define meta-stable states in the same way as in [6]. For any states $i, j \in \mathcal{S}$, let

$$p_{ij} = P\left[x_{t+1} = j : x_t = i\right]$$

be the probability of transitioning from state i to state j. We say that the collection $\mathcal{E} \subsetneq \mathcal{S}$ is a meta-stable state if there is some "small" number $\epsilon > 0$ such that

$$\frac{1}{|\mathcal{E}|} \sum_{i \in \mathcal{E}} \sum_{j \notin \mathcal{E}} p_{ij} < \epsilon.$$

That is, $\mathcal{E}$ is a meta-stable state if the mean probability of transitioning from a state $i \in \mathcal{E}$ to a state $j \notin \mathcal{E}$ is small.

When the Markov chain is modelled via a stochastic matrix, we can redefine this in terms of row sums of principal submatrices. Let the Markov chain $\{x_t\}$ have associated stochastic matrix M. Let $\mathcal{E}$ be a proper subset of the state space and let $M(\mathcal{E})$ be the corresponding principal submatrix of M. Then, $\mathcal{E}$ is a meta-stable state if for some small $\epsilon > 0$, the average of the row sums of $M(\mathcal{E})$ is greater than or equal to $1 - \epsilon$.

When a Markov chain has two or more disjoint meta-stable states we refer to it and its associated stochastic matrix as *nearly uncoupled*. Meta-stable states are sometimes referred to as *almost invariant aggregates*.

The identification of the meta-stable states of a nearly uncoupled Markov chain is of great importance in biomolecular research and pharmaceutical drug design [2, 4, 9]. In [3, 5], an approach to this problem using the Perron-Frobenius Theorem [1, 8], known as *Perron cluster analysis*, is detailed.

In [6], the authors present an algorithm for uncoupling a stochastic matrix that relies on the singular value decomposition of that matrix, rather than the spectral decomposition. However, we have found a pair of counterexamples which illuminate the fact that additional hypotheses concerning the matrices involved are required.

*Received July 6, 2010. Accepted for publication November 25, 2010. Published online February 14, 2011. Recommended by D. Szyld.

†National University of Ireland Maynooth, Hamilton Institute, Maynooth, Co. Kildare, Ireland (Ryan.Tifenbach@nuim.ie).

ETNA
Kent State University
http://etna.math.kent.edu

2. The singular value decomposition approach.

2.1. Summary. Let M be a square matrix. A singular value decomposition of M is an expression

$$M = U\Sigma V^*,$$

where U and V are unitary matrices, and Σ is a diagonal matrix where the diagonal entries are real and nonnegative and satisfy $\sigma_{ii} \geq \sigma_{jj}$ for all $i < j$. When M is real then U and V can be taken to be real matrices as well, in which case we have

$$M = U\Sigma V^T.$$

The ith columns of U and V are referred to as left and right singular vectors, respectively, of M associated with the singular value σ_{ii}. If M is real and we let the ith columns of U and V be u_i and v_i, respectively, we then have

$$Mv_i = \sigma_{ii}u_i \text{ and } M^T u_i = \sigma_{ii}v_i.$$

We label the singular values of M as $\sigma_i(M) = \sigma_{ii}$. The number $\sigma_1(M)$ is, in fact, equal to the 2-norm of M; that is, $\sigma_1(M) = \|M\|$. See [8] for a thorough exposition of the singular value decomposition.

Below is a brief summary of the algorithm described in [6]. It receives as inputs a matrix M and a threshold value δ.

ALGORITHM 2.1.

1. *We identify the left singular vector u_2 of M associated with the second largest singular value of M. We let*

$$\mathcal{E}_- = \{i \in \mathcal{S}\,|u_2(i) < 0\} \text{ and } \mathcal{E}_+ = \{i \in \mathcal{S}\,|u_2(i) > 0\}.$$

2. *If the average row sum of either of the principal submatrices matrices $M(\mathcal{E}_-)$ and $M(\mathcal{E}_+)$ is less than or equal to $1-\delta$, then we cannot partition the state space any further and we terminate the algorithm.*
3. *Otherwise, if the average row sum of each of the principal submatrices matrices $M(\mathcal{E}_-)$ and $M(\mathcal{E}_+)$ is greater than $1-\delta$, then $\mathcal{E}_-$ and $\mathcal{E}_+$ are each meta-stable states. We then attempt to further partition $\mathcal{E}_-$ and $\mathcal{E}_+$ into even smaller meta-stable states by applying the algorithm to each of $M(\mathcal{E}_-)$ and $M(\mathcal{E}_+)$.*

In the worked examples in [6], the threshold value $\delta = 1/2$ is used.

2.2. Counterexamples. In [6], Theorems 2.2 and 2.3, given below, are used to support Algorithm 2.1. We show that these theorems require further assumptions, as counterexamples that meet their respective conditions are constructible.

A digraph is *simply connected* if for all distinct vertices i and j, there is either a directed path from i to j or a directed path from j to i.

THEOREM 2.2. [6, Theorem 3.2] *Let A be a block-stochastic matrix of the form*

$$A = \mathit{diag}\,(A_1, \ldots, A_m)$$

with m simply connected diagonal blocks of order $n_1, \ldots, n_m$, denoted by

$$A_1, \ldots, A_m.$$

Let S_i be the set of n_i indices corresponding to the block A_i, $i = 1, \ldots, m$. Let

ETNA
Kent State University
http://etna.math.kent.edu

$$A = \tilde{U}\Sigma\tilde{V}^T$$

be a singular value decomposition of A and let $\tilde{u}_1, \ldots, \tilde{u}_m$ be the left singular vectors corresponding to the largest singular value of each of the blocks $A_1, \ldots, A_m$, respectively. Associate with every state s_i its sign structure

$$\textit{sign}\,(s_i) = \left[sgn\,(\tilde{u}_1)_i\,, \ldots, sgn\,(\tilde{u}_m)_i\right].$$

Then,

1. *states that belong to the same block of A exhibit the same sign structure, i.e., for any A_j and all $k, l \in S_j$, we have*

$$\textit{sign}\,(s_k) = \textit{sign}\,(s_l)\,;$$

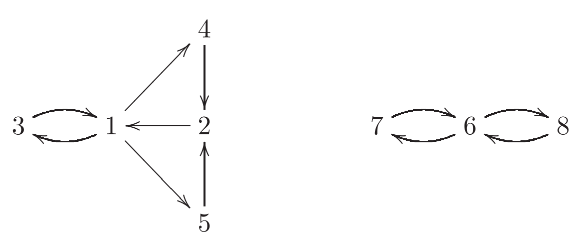
2. *states that belong to different blocks of A exhibit different sign structures, i.e., for any A_i, A_j with $i \neq j$ and all $k \in S_i$, $l \in S_j$ we have*

$$\textit{sign}\,(s_k) \neq \textit{sign}\,(s_l)\,.$$

We provide the following counterexample to Theorem 2.2. Let $A = \mathrm{diag}(A_1, A_2)$, where

$$A_1 = \begin{bmatrix} 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

The diagonal blocks of A are simply connected; the associated digraph is:



A singular value decomposition of A is $A = \tilde{U}\Sigma\tilde{V}^T$, where

$$\tilde{U} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{12} & 1/2 & 0 & 0 & 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{12} & 1/2 & 0 & 0 & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & -1/\sqrt{3} & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{3} & 0 & 0 & 0 & 0 & -1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{12} & -1/2 & 0 & 0 & 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{6} & 1/\sqrt{12} & -1/2 & 0 & 0 & 0 & 0 & -1/\sqrt{2} \end{bmatrix},$$

ETNA
Kent State University
http://etna.math.kent.edu

$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$\tilde{V} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{3} & 2/\sqrt{6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 1/\sqrt{2} \\ 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & -1/\sqrt{2} \end{bmatrix}.$$

The largest singular values of the two blocks are both equal to $\sqrt{2}$. Therefore, the sign structure of s_i can be defined, as in Theorem 2.2, to be the ith row of

$$\begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},$$

since the first two columns of $\tilde{U}$,

$$\begin{bmatrix} 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{12} \\ 1/\sqrt{6} & 1/\sqrt{12} \\ 1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{3} \\ 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{12} \\ 1/\sqrt{6} & 1/\sqrt{12} \end{bmatrix},$$

are both left singular vectors corresponding to $\sqrt{2}$. Note that the states s_2 and s_4 are in the same block but have different sign structures,

$$\text{sign}\,(s_2) = \begin{bmatrix} 1 & 1 \end{bmatrix} \text{ and } \text{sign}\,(s_4) = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

Further, the states s_2 and s_7 are in different blocks but have the same sign structure,

$$\text{sign}\,(s_2) = \text{sign}\,(s_7) = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

ETNA
Kent State University
http://etna.math.kent.edu

If we apply the algorithm to a small perturbation of A, the vector u_2 obtained could be a small perturbation of a linear combination of the first two columns of V. This could lead the algorithm to make two different types of error: it might fail to associate s_2 and s_4 with each other or it might mistakenly associate s_2 and s_7.

Theorem 2.3 (in a very similar way to 2.2) requires further assumptions on A (or possibly $T(\epsilon)$).

THEOREM 2.3. [6, Theorem 4.7] *Let $B = \hat{A} + \epsilon R$ for $\epsilon > 0$ where $\hat{A}$ and B are stochastic. Let $T(\epsilon) = BB^T$ have two largest eigenvalues $\lambda_1(\epsilon) \geq \lambda_2(\epsilon)$. Suppose that the unperturbed matrix $T = T(0)$ can be permuted to $\tilde{T} = PTP^T$ such that $\tilde{T}$ has m uncoupled irreducible blocks. Let $\lambda_1 > \lambda_2 > \ldots > \lambda_m$ be the largest eigenvalues corresponding to each of the blocks. Then, the perturbed orthonormal eigenvector φ_2 corresponding to the perturbed singular value $\lambda_2(\epsilon)$ is of the form*

$$\varphi_2(\epsilon) = \sum_{j=1}^{m} (\alpha_j + \epsilon\beta_j)\,\tilde{u}_j + \epsilon \sum_{j=m+1}^{n} \left\langle \varphi_j, \varphi_2^{(1)} \right\rangle \varphi_j + \mathcal{O}(\epsilon^2),$$

where $\tilde{u}_j$ are the eigenvectors in Theorem 2.2 and α_j, β_j are suitable coefficients.

We now produce a matrix that contradicts Theorem 2.3. Let

$$\hat{A} = \begin{bmatrix} 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 \\ 0 & 0 & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$R = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then,

$$T(\epsilon) = \left(\hat{A} + \epsilon R\right)\left(\hat{A} + \epsilon R\right)^T = \hat{A}\hat{A}^T + \epsilon^2 RR^T$$

$$= \begin{bmatrix} 1/5 + 2\epsilon^2 & 1/5 + 2\epsilon^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/5 + 2\epsilon^2 & 1/5 + 2\epsilon^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

ETNA
Kent State University
http://etna.math.kent.edu

(We have chosen R so that $AR^T = RA^T = 0$.)

The matrix $B = \hat{A} + \epsilon R$ is stochastic as long as $\epsilon \leq 1/5$. The nonzero eigenvalue of the first block is $\lambda = 2/5 + 4\epsilon^2 \leq 14/25$. So, the two largest eigenvalues of $T(\epsilon)$ are 3 and 2, with the eigenvector

$$\varphi_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

corresponding to the eigenvalue 2. The stochastic matrix A has two irreducible blocks whose largest singular values are $\sqrt{3}$ and 1, respectively. The left singular vectors of A corresponding to these values (described in Theorem 2.2) are

$$u_1 = \begin{bmatrix} 0 \\ 0 \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } u_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Clearly, φ_2 is not a perturbation of a linear combination of u_1 and u_2.

3. Discussion. The counterexamples were constructed by noting that the algorithm relies on two assumptions that do not necessarily hold true, given the assumptions concerning the unperturbed stochastic blocks A_i.

Let

$$A = \text{diag}(A_1, \ldots, A_m),$$

where each A_i is a stochastic matrix and let $B = A + \epsilon R$. In order for Algorithm 2.1 to correctly recover the blocks A_i from B, we need each of the following 4 assumptions to be satisfied:

1. the m largest singular values of A must be $\sigma_1(A_1), \ldots, \sigma_1(A_m)$;
2. any left singular vector of A associated with one of the m largest singular values must have the form

$$\varphi = \begin{bmatrix} \alpha_1 u_1(A_1) \\ \vdots \\ \alpha_m u_1(A_m) \end{bmatrix},$$

   where $u_1(A_i)$ is a left singular vector of A_i associated with $\sigma_1(A_i)$;
3. every left singular vector $u_i(A_i)$, above, associated with $\sigma_1(A_i)$ must have every entry positive, or every entry negative; and
4. the value ϵ must be small enough that the perturbation by ϵR does not alter the sign structure of the singular vectors.

ETNA
Kent State University
http://etna.math.kent.edu

Using real analysis, it is somewhat straightforward to show that if the first three conditions hold, there is $\delta > 0$ such that if $0 < \epsilon < \delta$, the fourth condition holds (as in [6]).

We explore assumptions concerning the unperturbed stochastic blocks A_i that will guarantee the truth of conditions 1, 2 and 3, above.

3.1. Graph theory definitions. We introduce some graph theoretic constructions to aid in our discussion.

Let G be an undirected graph with vertex set W and let $x, y \in W$. We use the notation $x \sim y$ to represent the presence in G of the undirected edge with endpoints x and y. A walk in G of length l is a sequence of $l + 1$ vertices (not necessarily distinct) $x_0, \ldots, x_l$ such that $x_i \sim x_{i+1}$ for $0 \leq i \leq l - 1$. If there is a walk in G of length greater than or equal to 1 with endpoints x and y we use the notation $x \rightsquigarrow y$; we label such a walk as $\omega : x \rightsquigarrow y$.

A *connected component* of G is a collection of vertices $C \subseteq W$ such that

1. for all $x, y \in C$, we have $x \rightsquigarrow y$; and
2. if $x \in C$ and $y \in W$ satisfy $x \rightsquigarrow y$, we then have $y \in C$.

An undirected graph is *connected* if its entire vertex set forms a single connected component.

An isolated vertex $x \in W$ is a vertex that is not incident to any edge (i.e. for all $y \in W$, $x \nsim y$). Some authors consider a single isolated vertex to be a connected component; however, we will not follow this convention. If the vertex x is isolated, we do not have $x \rightsquigarrow x$ and so x cannot be a member of a connected component (under the definition we use here).

The graph G is bipartite if its vertex set can be partitioned into disjoint sets X and Y such that if $x \sim y$ then x and y are not contained in the same set X or Y.

Let M be a symmetric matrix of order n. The *undirected graph* associated with M is the graph $G(M)$ that has vertex set

$$W = \{w_1, \ldots, w_n\},$$

and has $w_i \sim w_j$ if and only if $m_{ij} \neq 0$. A symmetric matrix M is *irreducible* if $G(M)$ is connected.

We define another graph associated with a matrix M. Let M be an $m \times n$ matrix (m and n are possibly distinct). Let

$$X = \{x_1, \ldots, x_m\} \text{ and } Y = \{y_1, \ldots, y_n\}.$$

The *bigraph* associated with M is the undirected bipartite graph on vertices $V = X \cup Y$ that has $x_i \sim y_j$ if and only if $m_{ij} \neq 0$.

We refer to the labellings of the vertex sets of $G(M)$ and $B(M)$ given above as the *canonical* vertex sets of these graphs.

3.2. Irreducibility of symmetric products. We first consider condition 3. Let A be a stochastic matrix. It is a straightforward application of the well-known Perron-Frobenius theorem to show that every left singular vector associated with $\sigma_1(A)$ has constant sign if and only if AA^T is irreducible; moreover, when this holds the multiplicity of $\sigma_1(A)$ as a singular value of A is 1. The matrix AA^T is irreducible exactly when its associated graph $G(AA^T)$ is connected. We produce necessary and sufficient conditions for a stochastic matrix A to have $G(AA^T)$ connected.

PROPOSITION 3.1. *Let A be a stochastic matrix. Then, AA^T is irreducible if and only if $B(A)$ contains exactly one connected component. Moreover, A^TA is irreducible if and only if $B(A)$ is connected.*

Proof. Let n be the order of A. We first consider the reducibility of the matrix AA^T. Let $G = G(AA^T)$, $B = B(A)$ and let W and $X \cup Y$ be the canonical vertex labellings of G and B, respectively. The ijth entry of AA^T is

$$(AA^T)_{ij} = \sum_{k=1}^{n} (A)_{ik}(A^T)_{kj} = \sum_{k=1}^{n} a_{ik}a_{jk}.$$

It is clear that $(AA^T)_{ij} \neq 0$ if and only if there is some index k such that $a_{ik} \neq 0$ and $a_{jk} \neq 0$. Thus, $w_i \sim w_j$ if and only if, for some k, we have $x_i \sim y_k$ and $x_j \sim y_k$. Any walk in B of length 2 will have one of two forms:

$$x_i \sim y_k \sim x_j \text{ or } y_i \sim x_k \sim y_j.$$

Therefore, the edge $w_i \sim w_j$ is present in G if and only if there is a walk of length 2 from x_i to x_j in B. Clearly, this implies that there is a walk $\omega_G : w_i \rightsquigarrow w_j$ in G if and only if there is a walk $\omega_B : x_i \rightsquigarrow x_j$ in B. Thus, AA^T is irreducible if and only if every member of X is contained in the same connected component of B.

Note that since A is stochastic, for every index i, there is some index k such that $a_{ik} \neq 0$. Thus, for every $x \in X$ there is at least one $y \in Y$ such that $x \sim y$ (in B); so, X contains no isolated vertices. As well, every edge in B has the form $x \sim y$ for some $x \in X$ and some $y \in Y$. Thus, every connected component of B contains at least one member of X and every member of X is contained in a connected component of B. This implies that every member of X is contained in the same connected component if and only if B has only one connected component. Therefore, AA^T is irreducible if and only if $B(A)$ contains exactly one connected component, possibly together with some amount of isolated vertices.

In the exact same manner, A^TA is irreducible if and only if every member of Y is contained in a single connected component of $B = B(A)$. As before, every connected component of B contains at least one member of Y. However, unlike the members of X, some members of Y may be isolated vertices. (It is entirely possible that there is an index j such that for all k, $a_{kj} = 0$.) So, A^TA is irreducible if and only if $B(A)$ contains exactly one connected component and no isolated vertices. □

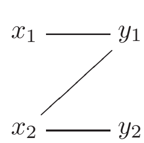
REMARK 3.2. Recall that isolated vertices are not considered to be part of a connected component. Thus, Proposition 3.1 informs us that $G(AA^T)$ is connected exactly when $B(A)$ consists of one connected component together with any amount (0 or more) of isolated vertices and, further, that $G(A^TA)$ is connected when $B(A)$ contains one connected component and no isolated vertices.

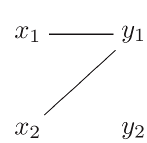
We included the discussion of the reducibility of A^TA in Proposition 3.1 to address the issue of using right singular vectors in Algorithm 2.1, rather than left singular vectors.

We note that a stochastic matrix A may be reducible even though AA^T is irreducible – for example, consider the matrices

$$A_1 = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Both are clearly reducible; but, examination of their bigraphs shows that $A_1A_1^T$ and $A_2A_2^T$ are both irreducible.

$B(A_1) =$ 

and $B(A_2) =$ 

.

The bigraph $B(A_1)$ contains one connected component (and no isolated vertices); the matrices $A_1A_1^T$ and $A_1^TA_1$ are both irreducible. The bigraph $B(A_2)$ contains one connected component and one isolated vertex; the matrix $A_2A_2^T$ is irreducible, but the matrix $A_2^TA_2$ is reducible.

Although the construction of the bigraph $B(M)$ associated with M is somewhat counterintuitive, the graph $B(M)$ is the most straightforward way of clearly visualising the combinatorial properties of MM^T. Consider, for example, the stochastic matrices

$$M_1 = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ and } M_2 = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

These matrices are both irreducible. Using the ideas found in [1], one can show that each of these matrices has 1 as a simple eigenvalue, M_1 has -1 as a simple eigenvalue (M_2 does not) and that the remaining eigenvalues of both matrices satisfy $|\lambda| < 1$. However, the matrices $M_1M_1^T$ and $M_2M_2^T$ are, in fact, identical, are reducible and have 1 as a singular value with multiplicity three.

$$M_1M_1^T = M_2M_2^T = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

These facts become apparent when examining $B(M_1)$ and $B(M_2)$, but are somewhat obfuscated by more traditional graph representations of M_1 and M_2:

$$B(M_1) = \begin{tikzcd} x_1 \arrow[r, dash] \arrow[dr, dash] & y_2 \\ x_2 \arrow[r, dash] \arrow[ur, dash] & y_4 \\ x_3 \arrow[r, dash] & y_1 \\ x_4 \arrow[r, dash] & y_3 \end{tikzcd} \text{ and } B(M_2) = \begin{tikzcd} x_1 \arrow[r, dash] \arrow[dr, dash] & y_3 \\ x_2 \arrow[r, dash] \arrow[ur, dash] & y_4 \\ x_3 \arrow[r, dash] & y_1 \\ x_4 \arrow[r, dash] & y_2 \end{tikzcd}.$$

In [6], the authors propose a preprocessing step that first aggregates members of large cycles, or clusters, within $G(M)$ by considering only the edges $w_i \sim w_j$ such that $m_{ij} \geq \chi$, where χ is some chosen tolerance value. It seems likely that, if χ is well-chosen, this will address most counterexamples involving irreducible blocks A where AA^T is reducible or "nearly" reducible. However, it is unclear whether this preprocessing step will address the second issue, which we discuss in the next section.

3.3. Interlacing of singular values. We now consider assumptions 1 and 2: Let $A_1, \ldots, A_m$ be stochastic matrices with largest singular values $\sigma_1(A_i)$ and corresponding left singular vectors $u_1(A_i)$. Let

$$A = \text{diag}(A_1, \ldots, A_m)$$

and let $B = A + \epsilon R$ be stochastic. Let $\varphi_i(\epsilon)$ be a left singular vector associated with one of the m largest singular values of B. Under what conditions must $\varphi_i(\epsilon)$ necessarily be a small perturbation of a vector of the form

$$\varphi = \begin{bmatrix} \alpha_1 u_1(A_1) \\ \vdots \\ \alpha_m u_1(A_m) \end{bmatrix}?$$

We show that if the singular values of the matrices A_i interlace in undesirable ways, this condition is not satisfied.

LEMMA 3.3. *Let $M \neq 0$ be a square, real matrix and let y be a real vector of the same order as M. Then, there is $\delta > 0$ such that if $0 < \epsilon < \delta$, then*

$$\left\|(1-\epsilon)^2 MM^T + \epsilon^2 yy^T\right\| < \|M\|^2 .$$

Proof. Let

$$f(t) = (1-t)^2 \|M\|^2 + t^2 \|y\|^2 .$$

Note that

$$\left.\frac{df}{dt}\right|_{t=0} = -2 \|M\|^2 < 0.$$

Thus, there is $\delta > 0$ such that if $0 < t < \delta$, then

$$f(t) < f(0) = \|M\|^2 .$$

Via the triangle inequality concerning the norm of an operator, if ϵ is real, we have

$$\begin{aligned} \left\|(1-\epsilon)^2 MM^T + \epsilon^2 yy^T\right\| &\leq (1-\epsilon)^2 \left\|MM^T\right\| + \epsilon^2 \left\|yy^T\right\| \\ &= (1-\epsilon)^2 \|M\|^2 + \epsilon^2 \|y\|^2 \\ &= f(\epsilon). \end{aligned}$$

So, if $0 < \epsilon < \delta$, we have $\|(1-\epsilon)^2 MM^T + \epsilon^2 yy^T\| \leq f(\epsilon) < f(0) = \|M\|^2$. ☐

We will make use of the following theorem and lemma, taken from [8]. Lemma 3.4 is a straightforward corollary of the Rayleigh-Ritz Theorem ([8, Theorem 4.2.2]). In our following discussion, we will often make reference to the largest eigenvalue, λ_1, of a matrix; we emphasise that this is in the sense that $\lambda_1 \geq \lambda_2$ for any other eigenvalue λ_2, and not necessarily $|\lambda_1| \geq |\lambda_2|$.

LEMMA 3.4. *Let M be a Hermitian matrix and let λ_1 be the largest eigenvalue of M. Then, for any vector y with $\|y\| = 1$,*

$$y^* M y \leq \lambda_1,$$

with equality if and only if y is an eigenvector associated with λ_1.

THEOREM 3.5. [8, Theorem 4.3.8] *Let M be a symmetric matrix of order n, y be a vector of order n and a be a real number. Let*

$$\hat{M} = \begin{bmatrix} a & y^* \\ y & M \end{bmatrix}.$$

ETNA
Kent State University
http://etna.math.kent.edu

Let the eigenvalues of M and $\hat{M}$ be $\{\lambda_i\}$ and $\{\hat{\lambda}_i\}$, respectively, and arrange them in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \text{ and } \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_{n+1}.$$

Then, the eigenvalues of M and $\hat{M}$ interlace: for all $1 \leq i \leq n$, $\hat{\lambda}_i \geq \lambda_i \geq \hat{\lambda}_{i+1}$.

PROPOSITION 3.6. *Let X and Y be stochastic matrices. Let the two largest singular values of X be $\sigma_1(X) \geq \sigma_2(X)$ and let the largest singular value of Y be $\sigma_1(Y)$. Suppose that $\sigma_2(X) \geq \sigma_1(Y)$. Then, there is a matrix R and a value $\delta > 0$ such that if $0 < \epsilon < \delta$, the matrix*

$$B = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} + \epsilon R$$

satisfies the properties:
1. *B is stochastic;*
2. *the second singular value (in magnitude) of B is $\sigma_2(B) = \sigma_2(X)$; and*
3. *any left singular vector of B associated with $\sigma_2(B)$ is of the form*

$$\varphi_2 = \begin{bmatrix} u_2(X) \\ 0 \end{bmatrix},$$

where $u_2(X)$ is a left singular vector associated with $\sigma_2(X)$.

Proof. Let the orders of X and Y be m and n, respectively.

It is well-known that

$$M = \sum \sigma_i u_i v_i^*$$

is an orthonormal singular value decomposition of M if and only if

$$MM^T = \sum \sigma_i^2 u_i u_i^*$$

is an orthonormal eigenvalue decomposition of MM^T and for all i,

$$v_i = \frac{1}{\|M^T u_i\|} M^T u_i.$$

Thus, we will recast this as an eigenvalue problem. We will produce a matrix R and a value $\delta > 0$ such that if $0 < \epsilon < \delta$, the matrix

$$B = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} + \epsilon R$$

satisfies the properties:
1. B is stochastic;
2. the second eigenvalue value of BB^T is $\lambda_2(BB^T) = \sigma_2(X)^2$; and
3. any eigenvector of BB^T associated with $\lambda_2(BB^T)$ is of the form

$$\varphi_2 = \begin{bmatrix} u_2(X) \\ 0 \end{bmatrix},$$

where $u_2(X)$ is an eigenvector of XX^T associated with $\sigma_2(X)^2$.

ETNA
Kent State University
http://etna.math.kent.edu

Let $\alpha_1 = \sigma_1(X)$, $\alpha_2 = \sigma_2(X)$ and $\beta = \sigma_1(Y)$.

The matrix XX^T is symmetric and nonnegative and its two largest eigenvalues are $\alpha_1^2 \geq \alpha_2^2$. By an application of the well-known Perron-Frobenius Theorem (see [1], for example), there is an eigenvector u_1 of XX^T associated with α_1^2 such that the entries of u_1 are nonnegative and $\|u_1\| = 1$. Let

$$v_1 = \frac{1}{\alpha_1} X^T u_1 \text{ and } v = \frac{1}{v_1^T \mathbb{1}} v_1,$$

where $\mathbb{1}$ is the column vector with every entry equal to 1. Since u_1 is entrywise nonnegative, v_1 is entrywise nonnegative. Thus, v is a nonnegative vector and the sum of the entries in v is 1. Note that

$$Xv_1 = \alpha_1 u_1,\ X^T u_1 = \alpha_1 v_1,\ Xv = \frac{\alpha_1}{v_1^T \mathbb{1}} u_1 \text{ and } v_1^T v = \frac{1}{v_1^T \mathbb{1}}.$$

Let

$$R = \begin{bmatrix} 0 & 0 \\ \mathbb{1} v^T & -Y \end{bmatrix}.$$

We have

$$B = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} + \epsilon \begin{bmatrix} 0 & 0 \\ \mathbb{1} v^T & -Y \end{bmatrix} = \begin{bmatrix} X & 0 \\ \epsilon \mathbb{1} v^T & (1-\epsilon) Y \end{bmatrix}.$$

The matrix B is clearly stochastic as long as $0 \leq \epsilon \leq 1$: it is nonnegative, the sum of the entries in any row of the $(2,1)$th block is ϵ and the sum of the entries in any row of the $(2,2)$th block is $1-\epsilon$. We express B as a sum of two matrices, $B = B_1 + B_2$, where

$$B_1 = \begin{bmatrix} \alpha_1 u_1 v_1^T & 0 \\ \epsilon \mathbb{1} v^T & (1-\epsilon) Y \end{bmatrix} \text{ and } B_2 = \begin{bmatrix} X - \alpha_1 u_1 v_1^T & 0 \\ 0 & 0 \end{bmatrix}.$$

We note that

$$\begin{aligned} B_1 B_2^T &= \begin{bmatrix} \alpha_1 u_1 v_1^T X^T - \alpha_1^2 u_1 u_1^T & 0 \\ \epsilon \mathbb{1} v^T X^T - \epsilon \alpha_1 \mathbb{1} v^T v_1 u_1^T & 0 \end{bmatrix} \\ &= \begin{bmatrix} \alpha_1^2 u_1 u_1^T - \alpha_1^2 u_1 u_1^T & 0 \\ \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} u_1^T - \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} u_1^T & 0 \end{bmatrix} \\ &= 0. \end{aligned}$$

Thus, $B_2 B_1^T = (B_1 B_2^T)^T = 0$ and so

$$BB^T = (B_1 + B_2)^T (B_1 + B_2) = B_1 B_1^T + B_2 B_2^T,$$

where $B_1 B_1^T$ and $B_2 B_2^T$ are orthogonal:

$$\left(B_1 B_1^T\right)\left(B_2 B_2^T\right) = \left(B_2 B_2^T\right)\left(B_1 B_1^T\right) = 0.$$

Thus, the nonzero eigenvalues of BB^T are the nonzero eigenvalues of $B_1 B_1^T$ together with the nonzero eigenvalues of $B_2 B_2^T$.

We calculate

ETNA
Kent State University
http://etna.math.kent.edu

$$
\begin{aligned}
B_2 B_2^T &= \begin{bmatrix} XX^T - \alpha_1 X v_1 u_1^T - \alpha_1 u_1 v_1^T X^T + \alpha_1^2 u_1 u_1^T & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} XX^T - \alpha_1^2 u_1 u_1^T - \alpha_1^2 u_1 u_1^T + \alpha_1^2 u_1 u_1^T & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} XX^T - \alpha_1^2 u_1 u_1^T & 0 \\ 0 & 0 \end{bmatrix}.
\end{aligned}
$$

The largest eigenvalue of $B_2 B_2^T$ is α_2^2 and we have $B_2 B_2^T y = \alpha_2^2 y$ if and only if

$$
y = \begin{bmatrix} u_2 \\ 0 \end{bmatrix},
$$

where $XX^T u_2 = \alpha_2^2 u_2$.

Next, we calculate $B_1 B_1^T$:

$$
\begin{aligned}
B_1 B_1^T &= \begin{bmatrix} \alpha_1^2 u_1 u_1^T & \epsilon \alpha_1 u_1 v_1^T v \mathbb{1}^T \\ \epsilon \alpha_1 \mathbb{1} v^T v_1 u_1^T & \epsilon^2 \mathbb{1} v^T v \mathbb{1}^T + (1-\epsilon)^2 YY^T \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1^2 u_1 u_1^T & \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} u_1 \mathbb{1}^T \\ \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} u_1^T & \epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T \end{bmatrix}.
\end{aligned}
$$

Let U be a unitary matrix such that $Uu_1 = e_m$, where e_m is the column vector, of order m, with its mth entry equal to 1 and every other entry equal to 0; since $\|u_1\| = \|e_m\| = 1$, such a matrix exists. Then, we have

$$
\begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix} B_1 B_1^T \begin{bmatrix} U & 0 \\ 0 & I \end{bmatrix}^T = \begin{bmatrix} \alpha_1^2 e_m e_m^T & \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} e_m \mathbb{1}^T \\ \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} e_m^T & \epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T \end{bmatrix}
$$

$$
= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \alpha_1^2 & \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1}^T \\ 0 & \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} & \epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T \end{bmatrix}.
$$

So, the nonzero eigenvalues of $B_1 B_1^T$ are the nonzero eigenvalues of

$$
M = \begin{bmatrix} \alpha_1^2 & \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1}^T \\ \frac{\epsilon \alpha_1}{v_1^T \mathbb{1}} \mathbb{1} & \epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T \end{bmatrix}.
$$

We will show that if $\epsilon > 0$ is sufficiently small, then the largest eigenvalue of M is $\hat{\lambda}_1 > \alpha_1^2$ and that the second largest is $\hat{\lambda}_2 < \beta^2$. Recall that the largest eigenvalue of $B_2 B_2^T$ is α_2^2 and that

$$
\beta^2 \leq \alpha_2^2 \leq \alpha_1^2.
$$

Thus, we will have shown that the second largest eigenvalue of

$$
BB^T = B_1 B_1^T + B_2 B_2^T
$$

ETNA
Kent State University
http://etna.math.kent.edu

is α_2^2, where α_2^2 is an eigenvalue of

$$B_2 B_2^T = \begin{bmatrix} XX^T - \alpha_1^2 u_1 u_1^T & 0 \\ 0 & 0 \end{bmatrix}$$

and is not an eigenvalue of $B_1 B_1^T$. This will imply that any eigenvector associated with this eigenvalue has the desired form.

Let

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_{n+1}$$

be the eigenvalues of M and let

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

be the eigenvalues of

$$\epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T.$$

We note that Lemma 3.3 implies that if $\epsilon > 0$ is sufficiently small,

$$\lambda_1 = \left\|\epsilon^2 \|v\|^2 \mathbb{1}\mathbb{1}^T + (1-\epsilon)^2 YY^T\right\| < \left\|YY^T\right\| = \beta^2.$$

Now, if $\epsilon \neq 0$, then $e_1^T M e_1 = \alpha_2^2$ but $Me_1 \neq \alpha_2^2 e_1$. Thus, by Lemma 3.4, $\hat{\lambda}_1 > \alpha_1^2$.

Finally, Theorem 3.5 implies that $\hat{\lambda}_1 \geq \lambda_1 \geq \hat{\lambda}_2$. Therefore, $\hat{\lambda}_1 > \alpha_1^2$ and $\hat{\lambda}_2 \leq \lambda_1 < \beta^2$. $\square$

3.4. Subdominant singular values of nearly uncoupled stochastic matrices. We can now see the conditions under which Algorithm 2.1 can reliably be applied. Let $A_1, \ldots, A_m$ be stochastic matrices, let

$$A = \mathrm{diag}(A_1, \ldots, A_m)$$

and let $B = A + \epsilon R$ be stochastic. In addition to assuming that ϵ is small enough that it does not significantly alter the singular vector sign structure among the m largest singular values, we must have

1. the bigraph $B(A_i)$ associated with each A_i contains exactly one connected component; and
2. for all $i \neq j$, $\sigma_2(A_j) < \sigma_1(A_i)$.

If we are using right singular vectors, we must replace condition 1 with the (slightly) stronger condition that each $B(A_i)$ is connected. We will examine, briefly, the problem of satisfying the second condition above. First, we produce a theorem concerning the largest singular value of a stochastic matrix.

A *doubly stochastic matrix* is a nonnegative square matrix that has the sum of the entries in any row or column equal to 1. That is, A is doubly stochastic if both A and A^T are stochastic.

THEOREM 3.7. *Let A be a stochastic matrix and let $\sigma_1(A)$ be the largest singular value of A. Then, $\sigma_1(A) \geq 1$. Further, $\sigma_1(A) = 1$ if and only if A is doubly stochastic.*

Proof. A simple fact concerning the eigenvalues and singular values of a matrix M is that for any eigenvalue λ of M, we have $|\lambda| \leq \sigma_1(M)$. Thus, for a stochastic matrix A, $1 \leq \sigma_1(A)$.

Suppose that A is doubly stochastic. Then, the matrix AA^T is a product of stochastic matrices and so is, itself, stochastic. So, the largest eigenvalue of AA^T is 1 and so we have $\sigma_1(A) = \sqrt{1} = 1$.

Suppose that A is stochastic and that $\sigma_1(A) = 1$. Then, for any vector y,

$$\|Ay\| \leq \sigma_1(A)\|y\| = \|y\|.$$

Let $v^T = \mathbb{1}^T A$, where $\mathbb{1}$ is the column vector with every entry equal to 1. Therefore,

$$v^T \mathbb{1} = \mathbb{1}^T A \mathbb{1} = n,$$

where n is the order of A. Further,

$$\|v\| = \|\mathbb{1}^T A\| \leq \|\mathbb{1}\| = \sqrt{n}.$$

We apply the Cauchy-Schwarz inequality: for real vectors x and y,

$$|x^T y| \leq \|x\|\|y\|,$$

with equality if and only if $y = ax$ for some real number a. We have

$$n = v^T \mathbb{1} \leq \|v\|\|\mathbb{1}\| = \sqrt{n}\|v\| \leq n.$$

We have equality in the Cauchy-Schwarz inequality and so the vector v is a scalar multiple of $\mathbb{1}$. However, $v^T \mathbb{1} = n$ implies that, in fact, we have $v = \mathbb{1}$. Thus, the sum of the entries in each column of A is 1. ☐

This theorem provides a rough estimate as to the number of meta-stable states the state space of a Markov chain may contain. If A is stochastic and has m singular values greater than 1, then we expect to find m or fewer meta-stable states.

Now, the condition that $\sigma_2(A_j) < \sigma_1(A_i)$ is somewhat difficult to work with – it concerns collections of stochastic matrices, rather than individual matrices. We propose the following conditions on the unperturbed blocks of a block diagonal stochastic matrix.

PROPOSITION 3.8. *Let $A_1, \dots, A_m$ be stochastic matrices and let*

$$A = \mathit{diag}(A_1, \dots, A_m).$$

Under the extra assumptions that for all i, $B(A_i)$ contains only one connected component and $\sigma_2(A_i) < 1$, the conclusions of Theorems 2.2 and 2.3 hold true.

Proof. The extra assumption guarantees that we have $\sigma_2(A_j) < 1 \leq \sigma_1(A_i)$ whenever $i \neq j$. ☐

Let A be a stochastic matrix; we refer to $\sigma_2(A)$ as the *subdominant singular value* of A. We present a result concerning the case that the subdominant singular value is strictly less than 1.

THEOREM 3.9. *Let A be a stochastic matrix. Then, $\sigma_1(A) = 1$ and $\sigma_2(A) < 1$ if and only if A is doubly stochastic and $B(A)$ is connected.*

Proof. First, suppose that A is doubly stochastic and that $B(A)$ is connected. Then, by Theorem 3.7 and Proposition 3.1, we have AA^T irreducible and stochastic, and $\sigma_1(A) = 1$. By the Perron-Frobenius Theorem, the multiplicity of 1 as an eigenvalue of AA^T is 1. So, the multiplicity of 1 as an singular value of A is 1 and we have $\sigma_2(A) < 1$.

So, suppose that A is stochastic and that $\sigma_2(A) < \sigma_1(A) = 1$. By Theorem 3.7, the matrix A is doubly stochastic. Thus, AA^T and $A^T A$ are both symmetric stochastic matrices

ETNA
Kent State University
http://etna.math.kent.edu

that have their second eigenvalues strictly less than 1. This implies that AA^T and A^TA are both irreducible and so we have $B(A)$ connected (again, by Proposition 3.1). □

Identifying exactly when a stochastic matrix has subdominant singular value strictly less than 1 is an interesting and possibly important open problem. Further, identifying exactly what probabilistic qualities the subdominant singular value of a stochastic matrix measures could lead to new insights in Markov chain analysis.

3.5. n-Pentane analysis. In [6], Algorithm 2.1 is applied, successfully, to experimental data concerning the state space of the n-Pentane molecular structure. We examine the singular values of one of the matrices involved, with our previous observations in mind.

One of the data sets analysed, $Ph300$, is a stochastic matrix of order 255. The algorithm successfully constructed seven meta-stable states, of orders 20, 42, 47, 46, 24, 36 and 40. We calculate the seven principal submatrices of $Ph300$ on these meta-stable states and label them $M_1, \ldots, M_7$. We then calculate, using MATLAB, the two largest singular values of each of these principal submatrices,

$$\sigma_1(M_1) = 1.5322,\ \sigma_2(M_1) = 0.6845,\ \sigma_1(M_2) = 1.5915,\ \sigma_2(M_2) = 0.9578,$$

$$\sigma_1(M_3) = 1.6322,\ \sigma_2(M_3) = 0.9756,\ \sigma_1(M_4) = 1.3912,\ \sigma_2(M_4) = 0.7853,$$

$$\sigma_1(M_5) = 1.2959,\ \sigma_2(M_5) = 0.4724,\ \sigma_1(M_6) = 1.5370,\ \sigma_2(M_6) = 0.9025,$$

$$\sigma_1(M_7) = 1.7787 \text{ and } \sigma_2(M_7) = 1.1035.$$

These calculated singular values satisfy $\sigma_2(M_j) < \sigma_1(M_i)$ for all $i \neq j$; moreover, in all cases except $i = 7$, they satisfy the stronger condition that $\sigma_2(M_i) < 1$. These results seem to support our proposed extra assumptions on the unperturbed blocks A_i. An examination of the data set $Ph500$ used in [6] yields similar results.

The SVD algorithm is a useful and robust algorithm. It performs well, for example on the data presented in [6], precisely because, in practical problems, the additional assumptions we present here are usually satisfied. However, we present this supplemental discussion to show that applications of this algorithm require care, as these extra assumptions are not necessarily true of the general stochastic matrix.

4. Acknowledgements. The author would like to thank Steve Kirkland for suggesting the consideration of the bigraph of a matrix. As well, we acknowledge the helpful suggestions of two referees that greatly improved this work.

REFERENCES

[1] R. BRUALDI AND H. RYSER, *Combinatorial Matrix Theory*, Cambridge University Press, New York, 1991.

[2] P. DEUFLHARD, *From molecular dynamics to conformational dynamics in drug design*, in Trends in Nonlinear Analysis, M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, eds., Springer, New York, 2003, pp. 269–287.

[3] P. DEUFLHARD, W. HUISINGA, A. FISCHER, AND C. SCHÜTTE, *Identification of almost invariant aggregates in reversible nearly uncouple Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.

[4] P. DEUFLHARD AND C. SCHÜTTE, *Molecular conformation dynamics and computational drug design*, in Applied Mathematics Entering the 21st Century, J. Hill and R. Moore, eds., Sydney, Australia, 2004, ICIAM, pp. 91–119.

[5] P. DEUFLHARD AND M. WEBER, *Robust Perron cluster analysis in conformation dynamics*, Linear Algebra Appl., 398 (2005), pp. 161–184.

[6] D. FRITZSCHE, V. MEHRMANN, D. SZYLD, AND E. VIRNIK, *An SVD approach to identifying metastable states of Markov chains*, Electron. Trans. Numer. Anal., 29 (2008), pp. 46–69. http://etna.math.kent.edu/vol.29.2007-2008/pp46-69.dir.

[7] D. HARTFIEL AND C. MEYER, *On the structure of stochastic matrices with a subdominant eigenvalue near 1*, Linear Algebra Appl., 272 (1998), pp. 193–203.

[8] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[9] C. SCHÜTTE AND W. HUISINGA, *Biomolecular conformations can be identified as metastable sets of molecular dynamics*, in Computational Chemistry, P. Ciarlet and C. L. Bris, eds., Vol. X of Handbook of Numerical Analysis, Elsevier, Amsterdam, 2003, pp. 699–744.