Electronic Transactions on Numerical Analysis.
Volume 37, pp. 296-306, 2010.

ISSN 1068-9613.

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

A ROBUST SPECTRAL METHOD FOR FINDING LUMPINGS AND META STABLE STATES OF NON-REVERSIBLE MARKOV CHAINS*

MARTIN NILSSON JACOBI†

Abstract. A spectral method for identifying lumping in large Markov chains is presented. The identification of meta stable states is treated as a special case. The method is based on the spectral analysis of a self-adjoint matrix that is a function of the original transition matrix. It is demonstrated that the technique is more robust than existing methods when applied to noisy non-reversible Markov chains.

Key words. Markov chain, stochastic matrix, metastable states, lumping, aggregation, modularity, block diagonal dominance, block stochastic

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

1. Introduction. The structural dynamics of large biomolecules can often be accurately described as a Markov transition process. Frequently, the dynamics display a separation of time scales where aggregated conformational states are evolving at much slower rate than the detailed molecular dynamics does. The problem of identifying the conformational states from the detailed Markov transition matrix has received recent interest [3, 4, 7, 9]. The technically similar problem of identifying modularity and community structure on complex networks has also been discussed extensively; see, e.g., [6, 13, 15].

Identification of meta stable states is a special case of a more general reduction called (approximate) lumping. Lumping of a Markov chain means that the state space is partitioned into equivalence classes of states called macro states. A coarse grained process is defined by the transitions between the macro states. If the coarse grained process is Markovian, i.e., exhibits no memory, we call the reduction a lumping. A partition into meta stable states is an example of an approximate lumping in the following sense: in the limit of complete stability, i.e., when there are no transitions between the macro states, the macro states define a degenerate case of exact lumping. More generally, the Markov property is fulfilled on the aggregated level if the relaxation process within a meta stable state is fast and mixing so that the memory of exactly how the meta stable state was entered is lost before the transition to a new meta stable state occurs. In this sense, aggregation into meta stable states can be viewed as an approximate lumping. Aside from separation of time scales, there are other generic situations when a Markov process is expected to be lumpable. For example, when a particle interacts with many other particles in a "heat bath" the dynamics of the single particle can be described as a Brownian motion. Technically, the transition matrix of a lumpable Markov chain can be rearranged into a block-stochastic structure; see Figure 5.1 and Definition (5.1). Markov chains with metastable states can be permuted into a block-diagonal structure (see Figure 4.1), which is a special case of a block-stochastic matrix.

The most successful methods for identifying meta stable states and modules in networks are based on the level structure of the eigenvectors whose corresponding eigenvalues are clustered close to the Perron-Frobenius eigenvalue. The technique introduced in this paper is closely related to these spectral method first introduced by Fidler in the 70's at that point as a method for graph partitioning [6]. Fidler noted that the second eigenvector of the graph Laplacian shows tightly connected communities of nodes that are connected to the other

*Received November 11, 2008. Accepted for publication April 30, 2010. Published online October 8, 2010. Recommended by D. Szyld.
†Complex Systems Group, Department of Energy and Environment, Chalmers University of Technology, Gothenburg, Sweden (mjacobi@chalmers.se).

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

communities by relatively few edges or low algebraic connectivity. Later the method was used in connection with the classic graph coloring problem [1]. In the same paper, the idea of using the sign structure of the k first eigenvectors to partition a graph into k aggregates of nodes was introduced. The same idea was later applied to identify meta stable states in Markov chains [3]. For these spectral methods to be stable, the eigenvalue problem must be symmetric with respect to some scalar product. This means that the Markov process must be reversible, or that the network is assumed to be effectively non-directional. A notable exception based on symmetrization using the stationary distribution was presented in [8]. Another exception is a recent method for Markov chains based on singular value decomposition of the Markov transition matrix [7]. However, the SVD-based method is not appropriate for identifying lumpings of Markov chains since the singular vectors typically do not have a level structure (or relevant sign structure) in the theoretical limit of exact lumpability; see, for example, the transition matrix defined in (2.2).

In this paper we present a new robust spectral method for identifying possible lumpings of non-reversible Markov chains. Instead of using the spectrum of the transition matrix directly, we define a self-adjoint "invariance matrix" whose kernel relates to the eigenvectors that define the meta stable states, or, more generally, the lumps of the Markov chain. Since the invariance matrix is self-adjoint by construction, the usual assumption of reversibility can be lifted. We demonstrate the method on both Markov chains with meta stable states and more general block-stochastic structures and compare the performance to other methods reported in the literature, e.g., the methods presented in [8] and [7].

2. Lumping of Markov chains. Consider a Markov process $x_{t+1} = x_t P$. The $N \times N$ transition matrix P is a row stochastic matrix, i.e., $\sum_j P_{ij} = 1, \forall i$. Consider a partition of the states space Σ into K equivalence classes of states L_k such that $L_k \cap L_l = \emptyset$ and $\bigcup_k L_k = \Sigma$; see [16]. A necessary and sufficient condition for a partition to be a lumping is that [10]

$$\sum_{j \in L_l} P_{ij} \quad \text{is constant for all } i \text{ in an aggregate, } i \in L_k. \tag{2.1}$$

If a Markov chain allows for a non-trivial lumping we call it lumpable. A simple example of a lumpable transition matrix is

$$P = \frac{1}{4}\begin{bmatrix} 3 & 0 & 1 \\ 1 & 2 & 1 \\ 0 & 2 & 2 \end{bmatrix}, \tag{2.2}$$

which, aside from the trivial lumping defined by all states aggregated into one macro-state, allows the non-trivial lumping $\{\{1,2\},3\}$, i.e., state 1 and 2 are lumped into one macro-state.

The condition in (2.1) also immediately defines the transition matrix with dimensions $K \times K$ for the aggregated dynamics

$$\widetilde{P}_{kl} = \sum_{j \in L_l} P_{ij} \quad i \in L_k \tag{2.3}$$

since all states $i \in L_k$ give the same result:

$$\widetilde{P}_{kl} = \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}.$$

In practice, Equation (2.1) is usually not fulfilled exactly. For example, if a transition matrix can be written as

$$P = (1-\epsilon)A + \epsilon B,$$

where A is a transition matrix that fulfills the lumpability condition (2.1) and B is some arbitrary transition matrix. Then, if ϵ is small, we say that P is approximately lumpable. Note that the aggregated transition probabilities in (2.1) are in this case approximately constant with deviations of $\mathcal{O}(\epsilon)$. The reduced dynamics must in this case be approximated, e.g., using a weighted average for the transitions between the aggregated states

$$\widetilde{P}_{kl} = \frac{1}{\sum_{j \in L_l} v_j} \sum_{i \in L_k} \sum_{j \in L_l} v_j P_{ij}, \tag{2.4}$$

where v_j is the stationary distribution. Using the weighted average is natural since it gives the same reduced transition matrix as we find if we estimate the aggregated transition probabilities directly from a stationary time series.

A partition can be represented by a matrix Π defined as $\Pi_{ik} = 1$ if $i \in L_k$ and $\Pi_{ik} = 0$ otherwise. Equation (2.1) can be reformulated as

$$P\Pi = \Pi\widetilde{P}, \tag{2.5}$$

which, if written out explicitly in terms of the elements, implies that the column space of Π spans a right-invariant subspace of P. Assuming that P is diagonalizable, the invariant subspace is spanned by a set of right eigenvectors of P, and due to the 0 or 1 structure of Π the elements in these eigenvectors must be constant over the aggregates. To be more precise, a lumping with K aggregates exists if and only if there are exactly K right eigenvectors of P with elements that are constant over the aggregates; see [14, 12] for details. As an example, the transition matrix defined in (2.2) allows for the lumping $\{\{1, 2\}, 3\}$ as indicated by the two first elements in the right eigenvectors $(1, 1, 1)^T$ and $(-1, -1, 2)^T$ being constant.

It should be noted that there exist other types of aggregation of states where the aim is to preserve (for example) the structure of the equilibrium distribution. A prominent example of this is the renormalization of lattice spin systems. However, in this paper we focus on lumping that respect the *dynamics* of the process. In this case the Markov property is the central constraint, i.e., the mutual information between the past and the future given the present should be zero on both the micro (a prerequisite for the procedure) and macro level (the lumping condition). This leads to the strong conditions on the aggregation seen in Equations (2.1) and (2.5). For a more detailed discussion on how memory appears on the coarse grained level if the lumping criterion is not fulfilled, see [14].

The principle idea behind spectral methods for identifying lumping or meta stable states, as well as modularity in networks, is to search for (right) eigenvectors whose elements are constant over the aggregates, i.e., eigenvectors with a level structure; see, e.g., [14] for details. If the transition matrix is symmetric under some scalar product, the eigenvectors are orthogonal, and it is easy to show that the constant level sets must have different sign structure [1] (the sign structure of a vector is defined by mapping negative elements to -1 and positive elements to $+1$). The sign structure is often used as a lumping criterion rather than the constant levels since this is expected to be a more numerically stable [1, 3].

For the detection of metastable states or modularity, the eigenvectors of interest are those corresponding to eigenvalues close to the Perron-Frobenius eigenvalue since these eigenvalues are related to the slow dynamics. In the case of general lumping, the eigenvectors involved are not necessarily distinguished by their appearance in the spectrum. However, as we discuss in Section 5, for large transition matrices the eigenvectors involved in lumping tend to be separated from the rest of the spectrum by being located further away from the origin in the complex plane than the rest of the spectrum, but not necessarily by being closer to 1.

As a complement to the spectral methods, the commutation relation (2.5) can be used directly to identify lumping of Markov chains. Start by making a random assignment of the

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

N states to K aggregates, and construct the corresponding Π matrix. Given the Π matrix, the reduced transition matrix $\widetilde{P}$, defined in Equation (2.3), can be derived by simply ignoring that the row elements are not constant within the aggregates and by using the average defined in (2.4). The left hand side in (2.5), $P\Pi$, defines a K dimensional vector for each of the N states. If the lumping is correct then all states in an aggregate k should have identical K dimensional vectors, and they should be equal to the k-th row of $\widetilde{P}$. If Π is not a lumping, we can try to improve it by assigning state i to aggregate k where $k = \text{argmin}_l \|(P\Pi)_i - \widetilde{P}_l\|$ and $(P\Pi)_i$ is the i-th row of the matrix $P\Pi$. The result is a new aggregation with a new Π matrix. The process can be iterated until convergence. A similar method was introduced by Lafon and Lee [11]. As pointed out in [5], it is similar in structure to the K-means clustering algorithm. It should be noted that this direct clustering technique only works if the aggregated dynamics has long relaxation time, i.e., there is a spectral gap supporting the lumping; see [11] for details. The performance of the algorithm is shown in comparison with the method introduced in this paper in Figures 4.4 and 5.3.

3. A robust method for identifying lumping. We now present the main idea of this paper. We would like to find invariant vectors containing invariant level sets. For moderately sized (unperturbed) transition matrices or for time-reversible Markov chains, the eigenvectors can be used to detect lumping. If the Markov chain is not reversible, a calculation of both eigenvalues and eigenvectors is numerically unstable since, for example, the transition matrix may contain non-trivial Jordan blocks [14]. This is the motivation for the new method. We start by noting that if a normalized vector u is approximately right invariant under P then there must exist a λ such that

$$\|(P - \lambda I)u\|_2 \ll 1, \tag{3.1}$$

where I denotes the identity matrix. The general idea behind Equation (3.1) is similar to the definition of pseudospectra of nonnormal matrices [17]. The square of the 2-norm on the left hand side in (3.1) is not sensitive to small changes in the elements of P, whereas for non-symmetric P the eigenvalue and eigenvector problem can be ill conditioned. Obviously, if u is an eigenvector and λ the corresponding eigenvalue, then (3.1) is zero reflecting the fact that the eigenvector is exactly invariant. The 2-norm of (3.1) is given by

$$u^\dagger Q(\lambda) u,$$

where $u^\dagger$ denotes the conjugated transpose of the vector u. The "invariance matrix" Q is defined as

$$Q(\lambda) = P^\dagger P - \lambda^* P - \lambda P^\dagger + |\lambda|^2 I$$

(note that Q is typically not a stochastic matrix). Regardless of the properties of P, $Q(\lambda)$ is by construction a self-adjoint matrix and its diagonalization is numerically stable. If λ is an eigenvalue of P, then $Q(\lambda)$ is positive semi-definite with a zero eigenvalue corresponding to the eigenvector of P with eigenvalue λ. If λ is not an eigenvalue, then $Q(\lambda)$ is positive definite. For a given λ, (3.1) is minimized by u being the eigenvector of $Q(\lambda)$ corresponding to the smallest eigenvalue of $Q(\lambda)$, or, in the case of degeneracy, by a linear combination of the eigenvectors of the smallest eigenvalue. It should be noted that the eigenvectors of the matrix Q are equivalent to the singular vectors of the shifted transition matrix $P - \lambda I$, which can be found using a singular value decomposition.

4. Meta stable states. Detecting meta stable states is an especially simple, but also especially interesting case of general (approximate) lumping. The meta stable states are characterized by their long relaxation time, and hence their dynamics is associated with eigenvalues

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

close to 1. The right eigenvectors involved in the aggregation have corresponding eigenvalues closer to 1 than the rest of the spectrum. As a consequence, meta stable states can be identified by the approximate constant level structure of the eigenvectors of

$$Q(1) = P^{\dagger}P - P - P^{\dagger} + I \tag{4.1}$$

with eigenvalues close to 0. As a consequence, the small eigenvalues of (4.1) include the eigenvectors needed in the aggregation. It should be noted that the actual eigenvalues of P associated with the meta stable states need not be close to 1 for the sub-dominant eigenvectors of $Q(1)$ to reveal the meta stable states; see Figures 4.2 and 4.3. Since Q is self-adjoint, the eigenvectors are orthogonal. The eigenvectors are approximately constant over the aggregates, and orthogonality can then only be achieved if each aggregate has a unique sign structure in the eigenvectors. This observation was used by Aspvall and Gilbert [1] and Deuflhard and coworkers [3, 4], but in these cases under the condition of symmetry of the adjacency matrix or reversibility of the transition matrix, respectively. Using the Q matrix, there is no need to make assumptions on P. It is straightforward to apply the same sign structure criterion to the eigenvectors of Q, but empirical tests have shown that in our case the following simple approach is very robust (see Figure 4.4):

1. Find the eigenvectors $\{u_i\}_{i=1}^{K}$ corresponding to the small eigenvalues of $Q(1)$ in (4.1).
2. For each state $j = 1, \ldots, N$ form a K dimensional vector $u_{\bullet j} = (u_{1j}, u_{2j}, \ldots, u_{Kj})$ of its corresponding elements in the K eigenvectors.
3. Use a standard clustering algorithm (e.g., K-mean) to cluster the states with respects to the $u_{\bullet j}$ vectors.

To test the method we generate two classes of matrices. The first class is of the form

$$P = (1 - \epsilon)B + \epsilon A, \tag{4.2}$$

where B is a block diagonal transition matrix with $3 - 5$ blocks and transition probabilities within the blocks chosen uniformly in the interval $[0, 1]$ and then normalized. The matrix A is a transition matrix with no block diagonal structure generated in the same way as the blocks in B. The parameter ϵ sets the level of perturbation of P from being block diagonal. Figures 4.1–4.3 show the corresponding spectrum and clustering of the elements in the dominant eigenvectors of P and the sub-dominant eigenvectors of Q for an example of a transition matrix of this type with $\epsilon = 0.7$.

It can be argued that matrices of the type in (4.2) are unlikely to appear in practical applications. Instead of the smooth average modulation that decrease the transition probabilities between blocks in (4.2), a more binary modulation often occurs in practice, i.e., many transition probabilities are zero. In this situation, meta stable states occur as a consequence of a higher probability of having transitions within (rather than between) meta stable states. Contrasting the construction in (4.2) this produces a sparse transition matrix. We construct the second class of matrices according to

$$P^{*}_{ij}(\epsilon, \delta) = \chi_{ij}(\epsilon, \delta)B_{ij}, \tag{4.3}$$

where B, as before, is a matrix with random entries chosen uniformly in the interval $[0, 1]$. In the matrix $\chi(\epsilon, \delta)$, the entries are chosen as binary values so that $\chi_{ij} = 1$ with probability δ if i and j are in the same block, $\chi_{ij} = 1$ with probability ϵ if i and j are *not* in the same block, and $\chi_{ij} = 0$ otherwise. Thus, δ controls the overall probability of transitions within the blocks, and ϵ controls the transitions between the blocks with $\delta \geq \epsilon$. The two extreme

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

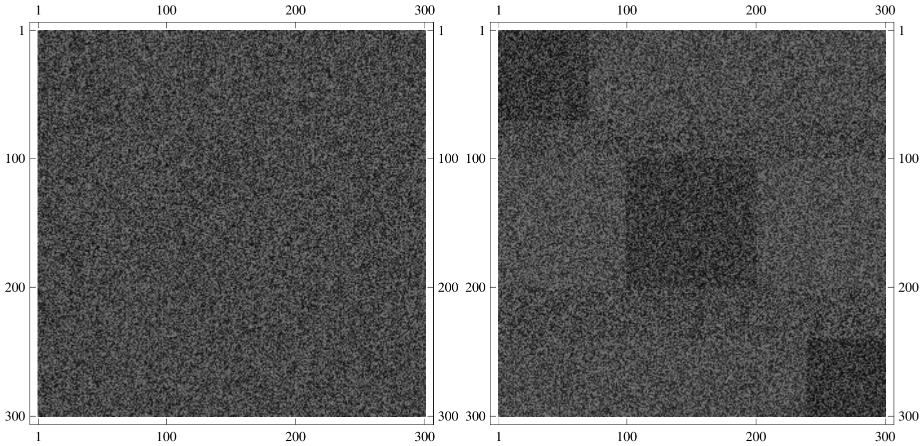

FIGURE 4.1. *A weakly block dominant transition matrix, constructed by (4.2) with $\epsilon = 0.7$, is shown. The time scale separation is not very pronounced; see the spectrum in Figure 4.3. To the left with random permutation and to the right after sorting the matrix according to the aggregation revealed in the clusters of the eigenvectors of the Q matrix shown in Figure 4.2.*

points are $\epsilon = 0$ which produce a completely block diagonal matrix, and $\epsilon = \delta$ which gives a matrix without any block diagonal structure. The rows in the matrix P^* are normalized to produce a stochastic matrix. The procedure described above can produce states with no outgoing transitions, i.e., that are ill defined as transition matrices. If this happens we generate a new matrix.

We tested the performance of the Q method and compared it to the following existing techniques: results from the eigenvectors of P, the right and left singular vectors from an SVD as suggested in [7], the clustering method presented in [11], and the spectral method described in [8]. As test cases we used the two classes of matrices described above and measured how stable the meta stable states produced by the different methods were by measuring the average waiting time between jumps between meta stable states. The result is shown in Figure 4.4. For each value of ϵ shown, the average switching time is measured for 100 matrices of size 200×200 in the respective class. The time τ is scaled so that the "correct partitioning" used to generate the matrices has waiting time 1. The results indicate that the Q method is more robust against perturbations than previously reported methods. It is especially interesting to note that for high ϵ values, i.e., when the original block diagonal structure is almost lost, the Q method still produce aggregations that are more stable than random partitions. None of the other methods are capable of finding these very weak meta stable states.

5. Block stochastic matrix. As discussed earlier, matrices with dominant block diagonal structure are special cases of lumpable Markov processes. The more general structure of lumpable transition matrices is shown in Figure 5.1. A block-stochastic matrix is a matrix of the form

$$P = \begin{bmatrix} \widetilde{P}_{11}a_{11} & \widetilde{P}_{12}a_{12} & \cdots & \widetilde{P}_{1m}a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{P}_{k1}a_{k1} & \widetilde{P}_{k2}a_{k2} & \cdots & \widetilde{P}_{mm}a_{mm} \end{bmatrix}, \tag{5.1}$$

where $\widetilde{P}$ is the $m \times m$ transition matrix of the reduced dynamics and each of the a_{ij} is a transition matrix in itself. Naturally, a_{ij} and a_{ji} must have the same dimensions for a fixed i and for $j = 1, \ldots, m$.

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

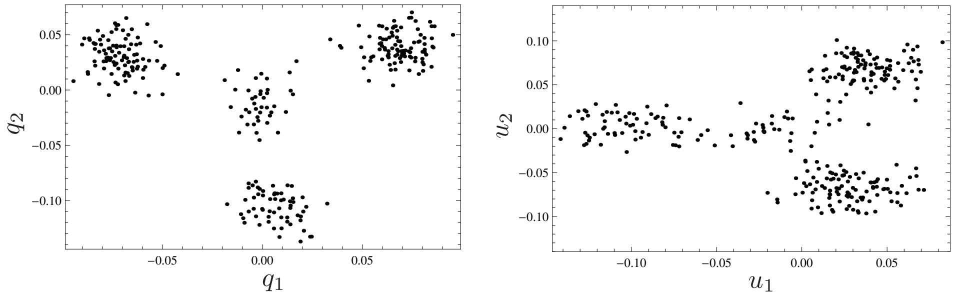

FIGURE 4.2. *The figure shows the clustering of the elements in the second and third smallest, respectively largest, eigenvectors of $Q(1)$ (to the left) and of P (to the right) for the matrix shown in Figure 4.1. Note that the clusters are more distinct in the eigenvectors of $Q(1)$ shown on the left.*

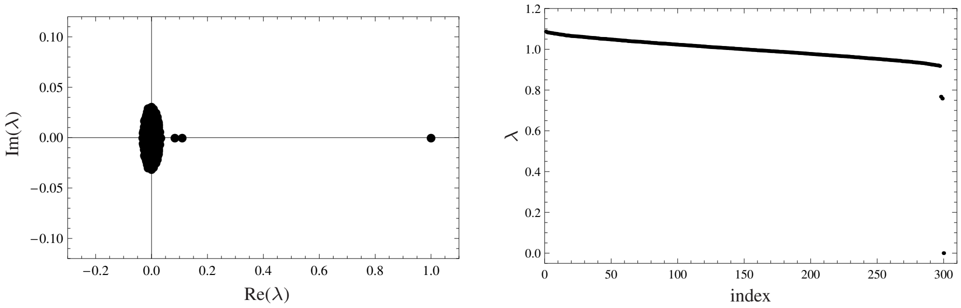

FIGURE 4.3. *The spectrum of the transition matrix P in Figure 4.1 to the left and of the corresponding $Q(1)$ matrix on the right. The aggregation into meta stable states is associated with the dominant eigenvalues of P, i.e., the Perron-Frobenius eigenvalue and the two eigenvalues close to 0.1, or alternatively with the three smallest eigenvalues of $Q(1)$ (to the far right in the figure). Note that even though the dominant eigenvalues of P are not clustered close to 1, the spectrum and eigenvectors of $Q(1)$ show the meta stable states more clearly than the eigenvectors of P; see Figure 4.2.*

The spectra of large block stochastic matrices tend to separate out the eigenvalues associated with the lumping. The separation is, however, different from the one occurring in block diagonally dominant transition matrices. Instead of clustering around the Perron-Frobenius eigenvalue, the reducing eigenvalues of a block stochastic matrix is separated by a large distance to the origin in the complex plane; see Figure 5.2. The reason behind the separation is also different from the block diagonal case, and this only appears as a statistic effect for large random transition matrices, as the following argument shows. When lumping a Markov chain, the spectrum of the reduced dynamics is always a subset of the original spectrum [2]. For a large transition matrix of size $N \times N$ with uncorrelated random transition probabilities, the spectrum is typically concentrated to a disk with radius $\sim 1/(2\sqrt{N})$, except for the Perron-Frobenius eigenvalue. For a block stochastic matrix, the eigenvalues of the lumped Markov chain $\widetilde{P}$ are typically concentrated to a disk with radius $\sim 1/(2\sqrt{K})$ where K is the number of states in the lumped chain. If $K \ll N$, then it should be expected that the eigenvalues associated with the lumped process separate from the rest of the spectrum. An example of this can be seen in Figure 5.2. However, it should be noted that the separation is only a typical behavior, it is not necessary for the Markov chain to be lumpable. (This seems to be incorrectly stated in [5, 12]; see [14] for details.)

If the spectrum does not show any separated eigenvalues that indicate the best choice of λ in $Q(\lambda)$, the implementation of the Q method is less straightforward when searching for

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

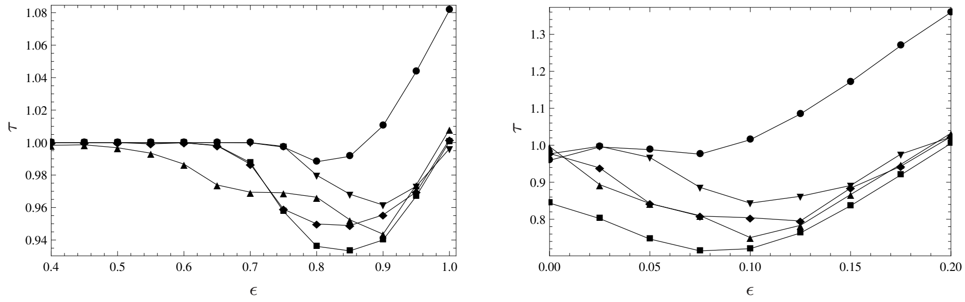

FIGURE 4.4. *The average result of identifying the meta stable states of a dominant block diagonal transition matrices defined in (4.2) is shown on the left, and for matrices defined in (4.3) with $\delta = 0.2$ on the right. The average waiting time for transitions between meta stable states (normalized against the result for the partitioning used to generate the matrices) for 100 test matrices of size 200×200 for each ϵ value is used as a measure of the quality of the results. The result for the Q method is displayed as ●, the result for using eigenvectors of P as ■, the singular vectors from SVD [7] as ◆, results from the clustering method suggested in [11] (see Section 2 for details) as ▼, and the method suggested in [8] as ▲.*

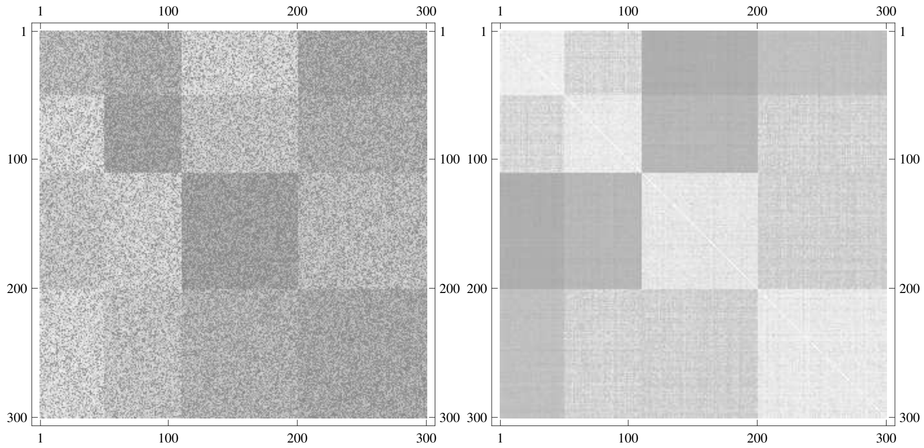

FIGURE 5.1. *A block stochastic transition matrix P, constructed as (5.1) with $\epsilon = 0.5$, is shown to the left and PP^{T} to the right. The SVD method from [7] is based on the right matrix and it is clear that the two matrices share the same block stochastic structure. However, the numerical tests show that Q method based on the left matrix is more stable to perturbations than the SVD method based on the eigenvectors of the right matrix. The spectrum of P is shown in Figure 5.2.*

general lumping. This is of course also the case for other spectral methods if the eigenvalues involved in the lumping do not separate from the rest of the spectrum. Perhaps the easiest way is to choose a set of numbers $\{\lambda_i\}_{i=1}^{K}$ randomly in the disk of radius 1 in the complex plane, use the eigenvectors of $Q(\lambda_i)$, $i = 1, \ldots, K$, corresponding to the smallest eigenvalue, cluster the elements in the same way as for the meta stable states, and check how well the result satisfies the lumpability criterion (2.1). The procedure must be repeated a few times to find the configuration with the most satisfying result. It is possible to design more sophisticated methods by reusing the λ's that seem to produce good results. However, we use the simplest possible approach choosing between 2 and 5 (guided by the separation in the spectrum) λ values randomly in the complex plane and repeating 10 times. The results are shown in Figure 5.3 in comparison with other methods. In these numerical test, we use the deviation

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

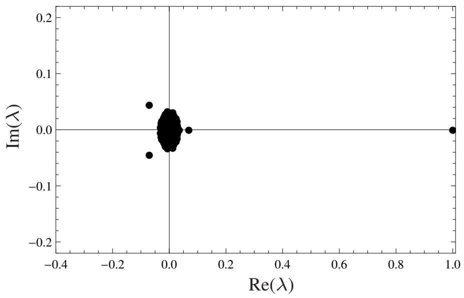

FIGURE 5.2. *The figure shows the spectrum of the block stochastic matrices shown in Figure 5.1. The eigenvalues associated with the eigenvectors that are involved in the lumping process are separated from the rest of the spectrum, but typically not close to the Perron-Frobenius eigenvalue. It should be noted that, though the separation in the spectrum typically appears for large block stochastic matrices, this is a statistical effect and not necessary for the transition matrix to be lumpable; see the main text for further discussion on this.*

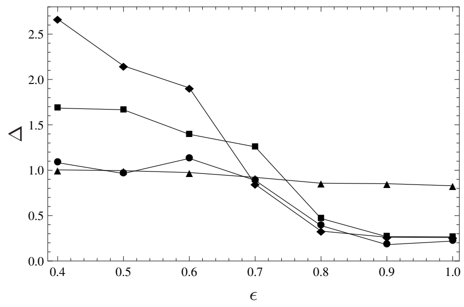

FIGURE 5.3. *The average result of inferring lumping of states of a block stochastic transition matrices defined in (5.1). The deviation from fulfilling the lumpability condition, defined in (5.2), is normalized against the deviation produced by the lumping used when producing the matrix, i.e., smaller values implies better results. For each ϵ value* 100 *independent realizations of* 200×200 *matrices on the form (5.1) were used to calculate the average performance. The result for the Q method is displayed as ●, the result for using eigenvectors of P as ◆, the singular vectors from SVD [7] as ■, and results from the clustering method suggested in [11] (see Section 2 for details) as ▲. For moderate ϵ, the Q method and the clustering performs equally superior to the other methods, while for larger ϵ all methods except the clustering technique show approximately equal performance.*

from the lumpability condition (2.5)

$$\Delta = \|P\Pi - \Pi\widetilde{P}\|_2 \tag{5.2}$$

as a measure of how well the different methods perform. For each value of ϵ, tests were performed with 100 matrices of the form $P = (1-\epsilon)B + \epsilon A$ generated, where B was constructed according to (5.1) and A was a random transition matrix. The numerical tests indicate that the Q method is more stable than using P directly or the SVD method. The clustering method performs almost as well as the Q method.

From a computational perspective, the Q method is, in the case of general lumping, significantly slower than the other spectral methods. The reason is that several random choices of λ's must be tried, and, in addition, $Q(\lambda)$ is a complex matrix if λ is complex. Neither of these complications occur when searching for meta stable states since then we know beforehand that $\lambda = 1$ is a good choice. In the implementation used to produce the results in Figure 5.3 the Q method is approximately 15 times slower than using P directly or the SVD method. On the other hand the results are also better. The slowdown scales proportional to the number of λ-setups we need to try. A more sophisticated selection procedure for choosing the regions where the sub-dominant eigenvectors of $Q(\lambda)$ show a clear signal would probably increase the efficiency of the algorithm.

6. Conclusions. We have introduced a new spectral method for identifying lumping in large Markov chains with the identification of meta stable states as an important special case. The key element of the method is to define a family of self-adjoint matrices from the transition matrix. The eigenvectors of the self-adjoint matrices are, as opposed to those of the transition matrix itself, stable to perturbations or noisy estimation of the transition probabilities. The robustness of the method is tested and compared to the results from previous methods, including a direct clustering method introduced in [11] and the recently suggested SVD based method introduced in [7]. The Q method is shown to be more robust than previous techniques.

We mentioned in the introduction that the method presented here can be used to reduce networks by aggregating nodes. The examples in this paper are, however, focused completely on lumping of Markov chains. The relation between lumping of Markov chain and reduction of complex networks was discussed recently in [5]. The authors define a diffusion process on the network using the standard method of the graph Laplacian. It should be noted, however, that reduction of networks can be defined with respect to other types of dynamics than diffusion. Straightforward jump processes are, for example, defined directly by multiplication of the adjacency matrix. Since the lumpability condition considered in this paper applies to general linear processes, not only Markov processes with stochastic transition matrix, the methods introduced can be used to reduce networks by aggregation with respect to different criteria depending on which dynamic process we are considering on the network. For the Q method to be different from using the eigenvectors of the transition matrix directly, the graph must be directed.

REFERENCES

[1] B. Aspvall and J. Gilbert, *Graph coloring using eigenvalue decomposition*, SIAM J. Algebra Discr., 5 (1984), pp. 526–538.
[2] D. Barr and M. Thomas, *An eigenvector condition for Markov chain lumpability*, Oper. Res., 25 (1977), pp. 1028–1031.
[3] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.
[4] P. Deuflhard and M. Weber, *Robust Perron cluster analysis in conformational dynamics*, Linear Algebra Appl., 398 (2005), pp. 161–184.
[5] E. Weinan, T. Li, and E. Vanden-Eijnden, *Optimal partition and effective dynamics of complex networks*, P. Natl. Acad. Sci. USA, 105 (2008), pp. 7907–7912.
[6] M. Fiedler, *Algebraic connectivity of graphs*, Czech. Math. J., 23 (1973), pp. 298–305.
[7] D. Fritzsche, V. Mehrmann, B. 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.mcs.kent.edu/vol.29.2007-2008/pp46-69.dir.
[8] G. Froyland, *Statistically optimal almost-invariant sets*, Physica D, 200 (2005), pp. 205–219.
[9] C. Jensen, D. Nerukh, and R. Glen, *Sensitivity of peptide conformational dynamics on clustering of a classical molecular dynamics trajectory*, J. Chem. Phys., 128 (2008), 115107 (7 pages).
[10] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, second ed., Springer, New York, 1976.
[11] S. Lafon and A. Lee, *Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization*, IEEE Trans. Pattern Anal. Mach. Intell., 28 (2006), pp. 1393–1403.
[12] M. Maila and J. Shi, *A random walks view of spectral segmentation*, in Artificial Intelligence and Statistics 2001, T. Jaakkola and T. Richardson eds., Morgan Kaufmann Publishers, San Francisco, CA, 2001, pp. 92–97.
[13] M. Newman, *Modularity and community structure in networks*, P. Natl. Acad. Sci. USA, 103 (2006), pp. 8577–8582.
[14] M. Nilsson Jacobi and O. Görnerup, *A spectral method for aggregating variables in linear dynamical systems with application to cellular automata renormalization*, Adv. Complex. Syst., 12 (2009), pp. 1–25.

[15] A. POTHEN, H. SIMON, AND K.-P. LIOU, *Partitioning sparse matrices with eigenvectors of graphs*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 430–452.
[16] L. C. G. ROGERS AND J. W. PITMAN, *Markov functions*, Ann. Probab., 9 (1981), pp. 573–582.
[17] L. TREFETHEN AND M. EMBREE, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.