

Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARCUS WEBER AND KONSTANTIN FACKELDEY

G-PCCA: Spectral Clustering for Non-reversible Markov Chains

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

G-PCCA: Spectral Clustering for Non-reversible Markov Chains

Marcus Weber*and Konstantin Fackeldey†,

July 2, 2015

Abstract Spectral clustering methods are based on solving eigenvalue problems for the identification of clusters, e.g., the identification of metastable subsets of a Markov chain. Usually, real-valued eigenvectors are mandatory for this type of algorithms. The Perron Cluster Analysis (PCCA+) is a well-known spectral clustering method of Markov chains. It is applicable for reversible Markov chains, because reversibility implies a real-valued spectrum. We extend this spectral clustering method also to non-reversible Markov chains and give some illustrative examples. The main idea is to replace the eigenvalue problem by a real-valued Schur decomposition. By this extension, non-reversible Markov chains can be analyzed. Furthermore, the chains need not have a positive stationary distribution. And additionally to metastabilities, dominant cycles and sinks can be identified, too.

*weber@zib.de

†fackeldey@zib.de

1 Introduction

The analysis of Markov chains is used to figure out the transition behavior in many fields of applications, ranging from the analysis of disease evolution in clinical data to molecular simulation. These data can be interpreted as elements of a typically large space.

More precisely, $\{X_i, i \in \mathbb{N}\}$ is a sequence of random variables with values x_i in the finite state space $\Gamma = \{1, \dots, N\}$. A large number N of states makes it difficult to reveal the general transition behavior of the chain. Clustering aims at a reduction of the number of states $n \ll N$ by still describing the underlying stochastic process correctly. Under the assumption that the stochastic process possesses metastable set, which are subsets in the state space, where the system or stochastic process spends much time before switching to a different metastability, the clustering is used to identify the rapidly mixing parts and separate them from each other. In this context spectral clustering methods have celebrated quite great success [MLKCW03, Shu03, WL07, VLKG03, LBDD01, DHFS00, DW05], however besides the assumption of metastability, it has to be assumed that the underlying Markov chain is reversible, i.e., the transition matrix has a real-valued spectrum. Reversibility means that reversing the stochastic process leads to the same transition behavior. This property allows for a clustering in terms of metastable sets.

The authors of [FMSV07] proposed - and further developed by [Tif11] - to replace the eigenvalue problem by a singular value decomposition. Moreover, in [Jac10] it is claimed that the singular vectors do not have the relevant sign structure to identify the metastable states, thus it is not preserving the dynamical structure of the Markov chain. Nevertheless this method has been applied in [TBHY13] however in the context of identifying the collective variables.

In [Fil91] the eigenvalue bounds of the mixing rates for reversible Markov chains have been extended to non-reversible chains by reversibilizing the non-reversible matrix. Based on this clustering methods for non-reversible processes [RM07, HMS04] but also other approaches [Jac10, SS14] have been developed.

In this article we propose a novel clustering method (G-PCCA) aiming at grouping states of a Markov chain by their transition behavior by replacing the eigenvalue decomposition with a Schur decomposition. It turns out that this novel method offers a powerful analysis of the Markov chain which also includes the identification of coherent subsets and the freedom of regarding an arbitrary initial distribution of states. Thus this novel method covers a broader class of applications by including non reversible Markov chains. Since this method is a generalization of PCCA+ towards non-reversible processes it is named G-PCCA (Generalized-Perron Cluster Cluster Analysis).

2 Markov chains and clustering

A finite autonomous Markov chain is given by a sequence X_0, X_1, \dots of random variables $X_k, k = 0, 1, 2, \dots$. Since the set of all states is finite, a transition probability matrix $(P_{ij})_{i,j=1,\dots,N}$ can be given by

$$P_{ij} = \mathbb{P}(X_{k+1} = j; X_k = i) \quad i, j \in \{1, \dots, N\}, k \in \mathbb{N}$$

where \mathbb{P} denotes the conditional probabilities for reaching state j in 1 step of the Markov chain, if the process has started in i . Obviously this matrix is non-negative and stochastic, i.e.

$$P_{ij} \geq 0 \quad \forall i, j \quad \sum_{j=1}^N P_{ij} = 1 \quad \forall i.$$

Let us furthermore denote the initial probabilities by $\eta_i = \mathbb{P}(X_0 = i)$, such that the vector $\eta = (\eta_1, \dots, \eta_N)$ is the initial distribution.

The transition matrix $P(k)$ of the k th step then meets the semi group property given by

$$P(k) = (P(1))^k = P^k. \quad (1)$$

Equation (1) is also named “markovianity”¹. For the Markov chain represented by P , we aim at a *clustering*, i.e. to find a projection of a Markov chain from a high number of states N to a low number of clusters n .

The Markov chain between the N represented by an $(N \times N)$ -transition matrix P is, thus, replaced by a $(n \times n)$ -matrix P_C providing the transition behavior between the n clusters.

A clustering can also be interpreted as a projection $G(P)$ of the matrix P onto n clusters. However, in order to guarantee that this projection is suitable, it should also meet (1), i.e.

$$(G(P))^k = G(P^k). \quad (2)$$

In general, the projection of a Markov chain is not markovian (cp Fig 1) and thus the stochastic process induced by the $n \times n$ transition matrix between the clusters is in general not a Markov process. Markovianity of the projection

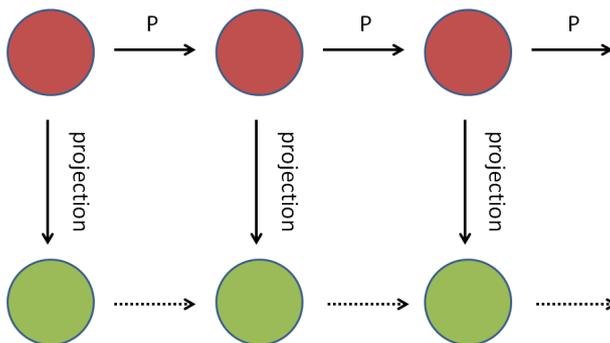


Figure 1: A Markov chain (red) with N states is projected to a chain (green) with n clusters. While there exists a transition matrix for the Markov chain, the projected chain is not markovian.

¹This is a consequence of the Chapman Kolmogorov equation.

$G(P)$ can be guaranteed, if the projection meets

- *invariant subspace condition*: there exists a matrix $X \in \mathbb{R}^{N \times n}$ (for a suitable choice of n) which meets

$$PX = X\Lambda \quad (3)$$

for $\Lambda \in \mathbb{R}^{n \times n}$

- *orthogonality relation*

$$X^T D_\eta X = I_{n \times n}, \quad (4)$$

where $D_\eta = \text{diag}(\eta_1, \dots, \eta_N)$ and $\Lambda \in \mathbb{R}^{n \times n}$, i.e. the X are spanning an n dimensional invariant subspace of P .

In other words: Conditions (3) and (4) of a projection G are sufficient for (2).

We remark, that a singular valued decomposition of P does not meet (3) and consequently a Galerkin projection leads to a projection error [SS14, Chapter 5.2]. In the next section we show how the orthogonality relation and the invariant subspace condition is realized for reversible Markov chains.

2.1 Reversible Markov Chains

If we assume an irreducible, aperiodic and reversible Markov chain then it has an unique vector $\pi = (\pi_1, \dots, \pi_N)^T$ such that

$$\pi^T P = \pi^T \text{ and } \sum_{i=1}^N \pi_i = 1,$$

where π is an invariant distribution or equilibrium distribution. If one takes π as initial distribution, then the chain is stationary. We denote by $D_\pi = \text{diag}(\pi_1, \dots, \pi_N)$ the matrix with the invariant distribution on its diagonal. For stationary Markov chains the detailed balance condition given by

$$\pi_i P_{ij} = \pi_j P_{ji} \quad (5)$$

is a necessary and sufficient characterization of reversibility in terms of transition probabilities and equilibrium distribution. In this special case, (3) is the eigenvalue equation of P , where Λ is the diagonal matrix of the *real* eigenvalues near the Perron root $\lambda_1 = 1$ and X are the corresponding *real* eigenvectors. Since the eigenvalues are real, they can be arranged in an descending order ($1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \dots$). The orthogonality relation is only assured, if the initial distribution equals the equilibrium distribution, i.e., $\eta = \pi$ in (4). The relation between eigenvalues of P close to 1 and metastable states of the Markov chain has been used by several authors in the past [Sch99, DHFS00, DW05, Web06, SS14].

The projection problem $(G(P))^k \neq G(P^k)$ has been discussed for the case of reversible Markov chains: In [Web06] this problem has been solved by looking at the Markov chain as a sequence of distributions instead of a sequence of states [Web02, DW05, Web06]. If $\eta \in \mathbb{R}^n$ is a probability distribution at a certain step of the chain, then $\hat{\eta} = P^T \eta$ denotes the probability distribution of states at the next step of the Markov chain. How does projection and propagation of

distributions commute? This problem is solved by a subspace projection, such that the projection error vanishes. The projection from N states to n clusters can be expressed by a membership matrix C . The non-negative entries C_{ij} of this matrix denote, how probable (or how intensive) the state i of the Markov chain belongs to the cluster j of the projection. The row-sum of this matrix is 1. One part of solving the projection problem is: The membership matrix is constructed via PCCA+, i.e., $C = X\mathcal{A}$ is a linear combination of the leading eigenvectors of P , where $X \in \mathbb{R}^{N \times n}$ is the matrix of the n leading eigenvectors and $A \in \mathbb{R}^{n \times n}$ is the non-singular transformation matrix computed (as a solution of an optimization problem) by PCCA+ [Web06, DW05]. PCCA+ is a well-established method for clustering metastable, reversible Markov chains. This method uses the dominant eigenvalues of the corresponding transition matrix. These eigenvalues are real (because of the reversibility of P) and they are close to the Perron root $\lambda = 1$. The Perron root is algebraically and geometrically simple, if the matrix is irreducible and the Markov chain is aperiodic. If η is an initial distribution of states, then $\eta_c = C^T \eta \in \mathbb{R}^n$ is its projection onto the clusters (thus the projection matrix is $\Pi^f = C^T = \mathcal{A}^T X^T$). In the reversible case, the irreducible and aperiodic matrix is projected via the following equation:

$$P_c = (C^T D_\pi C)^{-1} (C^T D_\pi P C), \quad (6)$$

where $D_\pi \in \mathbb{R}^{N \times N}$ denotes the diagonal matrix constructed by the entries of the vector $\pi \in \mathbb{R}^N$, being the invariant density of P . Note that in the case of a reversible Markov chain, the left eigenvectors of P are then given by

$$Y = D_\pi X,$$

such that the projection (6) clearly meets the orthogonality relation (4). Beyond that, orthonormality holds via $Y^T X = X^T D_\pi X = I$, where $I \in \mathbb{R}^{n \times n}$ is the unit matrix. *We emphasize, that in this setting, the reversibility of the Markov chain implies the orthogonality relation.* By assuming that the starting distribution is a linear combination of the left eigenvectors of P , i.e., $\eta = Y\alpha$, where $\alpha \in \mathbb{R}^n$ is the vector of linear coefficients of this combination the projection also meets the invariance condition (3). By the previous equations we get

$$A^T \alpha = A^T X^T D_\pi X \alpha = C^T Y \alpha = C^T \eta = \eta_c.$$

We can thus define $\Pi^b = D_\pi X A^{-T}$ as the back projection, such that $\eta = \Pi^b \eta_c = D_\pi X A^{-T} \eta_c$ and $\alpha = X^T \eta$. We are then in a position to prove the following

Lemma 1. *The propagation of the projected distributions commutes with the projection of the propagated distributions, i.e.*

$$\Pi^b (P_c^T)^k \Pi^f \eta = (P^T)^k \eta.$$

Proof: Let the number of steps be given by $k \in N$, then

$$\begin{aligned} \Pi^b (P_c^T)^k \Pi^f \eta &= D_\pi X A^{-T} [(C^T P^T D_\pi C) (C^T D_\pi C)^{-1}]^k A^T X^T \eta \\ &= D_\pi X A^{-T} [(A^T X^T P^T D_\pi X A) (A^T X^T D_\pi X A)^{-1}]^k A^T X^T \eta \\ &= D_\pi X A^{-T} [(A^T \Lambda A)^{-T}]^k A^T X^T \eta \\ &= D_\pi X \Lambda^k X^T \eta = D_\pi X \Lambda^k \alpha = (P^T)^k D_\pi X \alpha \\ &= (P^T)^k \eta, \end{aligned}$$

where in $\Lambda \in \mathbb{R}^{n \times n}$ is the diagonal matrix of the the dominant real eigenvalues.
 \square

Lemma 1 shows that for the distributions propagated via P^T , the projected distributions are propagated via P_c^T (without error). Since the projection $\eta_c = C^T \eta$ is a non-negative vector with entry sum 1, it can be interpreted as a distribution on the n clusters. Calculations [Web11] also show, that P_c has the row-sum 1 and, thus, can be understood as the transition matrix of the corresponding projected Markov chain.

Summing up, Lemma 1 shows, that the choice a projection which meets the invariance condition (3) and orthogonality relation (4) leads to a commuting diagram (Fig. 1). In this subspace projection, the initial distribution of the system is given by a linear combination of the left eigenvectors of P . The projected distribution $\eta_c = C^T \eta$ is propagated by a matrix P_c^T , which can be computed according to (6). The diagram commutes, if the membership matrix $C = X\mathcal{A}$ is a linear combination of the right eigenvectors of P .

2.2 Non-reversible Markov chains

In the foregoing section the orthogonality relation in the context of eigenvectors was realized by assuming that the underlying process is reversible. In fact Lemma 1 is only true if the underlying process is reversible. By resigning the reversibility of the underlying Markov chain, an interpretation of a transition matrix in terms of unconditional transition probabilities is not possible since then the eigenvectors do not meet the invariance condition (3) and the subspace condition (4) in general.

Moreover for non reversible processes, the spectrum of its corresponding transition matrix is in general not real but complex.

We thus take advantage of a Schur decomposition. Let therefore \tilde{X} be n Schur vectors of $\tilde{P} = D_\eta^{0.5} P D_\eta^{-0.5}$, then we have

$$\begin{aligned} \tilde{P} \tilde{X} &= \tilde{X} \Lambda \\ \iff D_\eta^{0.5} P D_\eta^{-0.5} \tilde{X} &= \tilde{X} \Lambda \\ \iff P D_\eta^{-0.5} \tilde{X} &= D_\eta^{-0.5} \tilde{X} \Lambda \\ \iff P X &= X \Lambda, \quad X = D_\eta^{-0.5} \tilde{X}. \end{aligned}$$

We have thus shown, that a Schur decomposition meets the invariant subspace condition (3) and the orthogonality condition (4). As a consequence, the projection

$$G(P) = (C^T D_\eta C)^{-1} (C^T D_\eta P C)$$

with Schur vectors X meets (2), to show this, we have the following

Theorem 1. *Let $G(P) = (C^T D_\eta C)^{-1} (C^T D_\eta P C)$, where X are the Schur vectors according to (7) and $C = X\mathcal{A}$ and D_η be some initial distribution of the Markov chain, then*

$$(G(P))^k = G(P^k).$$

Proof:

$$\begin{aligned}
G(P) &= (C^T D_\eta C)^{-1} (C^T D_\eta P C) \\
&= (\mathcal{A}^T X^T D_\eta X \mathcal{A})^{-1} (\mathcal{A}^T X^T D_\eta P X \mathcal{A}) \\
&= (\mathcal{A}^T X^T D_\eta X \mathcal{A})^{-1} (\mathcal{A}^T X^T D_\eta X \Lambda \mathcal{A}) \\
&= (\mathcal{A}^T \mathcal{A})^{-1} (\mathcal{A}^T \Lambda \mathcal{A}) \\
&= \mathcal{A}^{-1} \Lambda \mathcal{A},
\end{aligned}$$

such that $G(P)$ meets the desired criterion:

$$(G(P))^k = (\mathcal{A}^{-1} \Lambda \mathcal{A})^k = \mathcal{A}^{-1} \Lambda^k \mathcal{A} = G(P^k).$$

□

Remark 2.1. *Note that the Markov chain in Theorem 1 does neither need to be aperiodic nor irreducible. Moreover - in contrast to the reversible case - the initial distribution η does not have to be the stationary distribution. Theorem 1 may also be interpreted as commutativity between propagation in time (k steps) and discretization G , which is a desired property for long term predictions.*

In the real Schur decomposition the matrix Λ is an upper triangle matrix with possibly 2×2 -blocks on its diagonal. The remaining problem is, that an arrangement of the Schur decomposition in descending order (of eigenvalues) is no longer possible. In [Bra02] it has been proposed to arrange the Schur-values according to a absolute distance to a given target value z . For the reversible case $z = 1$ should be chosen, to guarantee that P_C is close to unit matrix allowing for a clustering into metastable states (the eigenvalues of P_C correspond to these selected values).

For the non reversible case, however, we can apply another method by arranging the Schur-values according to the distance from the unit circle. In this case P_C has eigenvalues close to the unit circle and, thus, similar to a permutation matrix, which can be seen as a clustering of states in the sense of coherent sets [FPG14]. This feature of G-PCCA is shown in the section of illustrative examples below.

2.3 G-PCCA

So far we have not yet explained how to obtain the matrix \mathcal{A} . In the framework of G-PCCA, this step is identical to PCCA+ [Web06, DW05]. The problem of finding the matrix \mathcal{A} can be converted to an optimization problem. More precisely, G-PCCA finds a transformation matrix \mathcal{A} mapping the column vectors of Schur vectors X , spanning the invariant subspace, to the basis $C = X\mathcal{A}$ used for the projection $G(P)$. Finding an optimal $n \times n$ -basis transformation matrix \mathcal{A} is the aim of this algorithm. As input the matrix X of the invariant subspace is needed. The output of G-PCCA is the above mentioned matrix of membership vectors C . The column vectors of both matrices, X and C , span the same subspace. Thus, G-PCCA provides an invariant subspace projection of P , such that the subspace spanning vectors C have an interpretation in terms of membership vectors. To do so, the matrix C has to meet the following properties explaining the simplex structure of C :

- $\sum_{J=1}^{n_C} C_{iJ} = 1$ (partition of unity)
- $C_{iJ} \geq 0 \quad \forall i \in \{1, \dots, n\}$ (positivity)
- $C = X\mathcal{A}$, \mathcal{A} non-singular, (invariance).

These conditions imply the feasible set of transformations \mathcal{A} . The selection of \mathcal{A} is realized by a convex maximization problem [DW05].

3 Examples

In this section we investigate two types of examples. These examples will show, that G-PCCA is indeed a powerful generalization of PCCA+. Instead of computing a projected Markov chain of a reversible metastable processes, it can be used to rigorously analyze non-reversible chains or in order to find transient states which have a common target set of states.

3.1 Illustrative metastability example

In the first example, we analyze the following transition network in Figure 2. The corresponding transition matrix has one real eigenvalue ($\lambda_1 = 1$) and 8

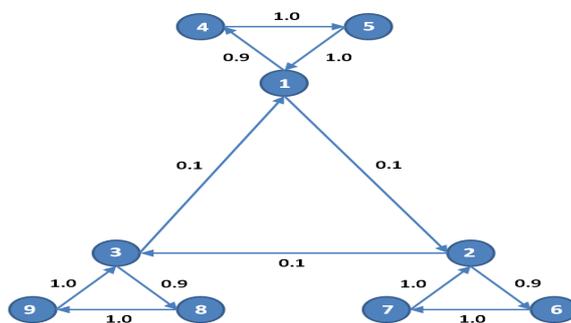


Figure 2: Transition network with three metastable states, but with a directed, non-reversible transition pattern.

complex eigenvalues. Out of these, the two eigenvalues with the highest real absolute value are $\lambda_{2,3} = 0.9483 \pm 0.0279i$. These values are close to $\lambda_1 = 1$ and indicate in total three metastabilities. Analyzing this network via PCCA+ is impossible. If we make it reversible before applying PCCA+, we spoil the directed structure of the network, see Fig. 3.

In contrast to that, G-PCCA can directly be applied to the Schur vectors of

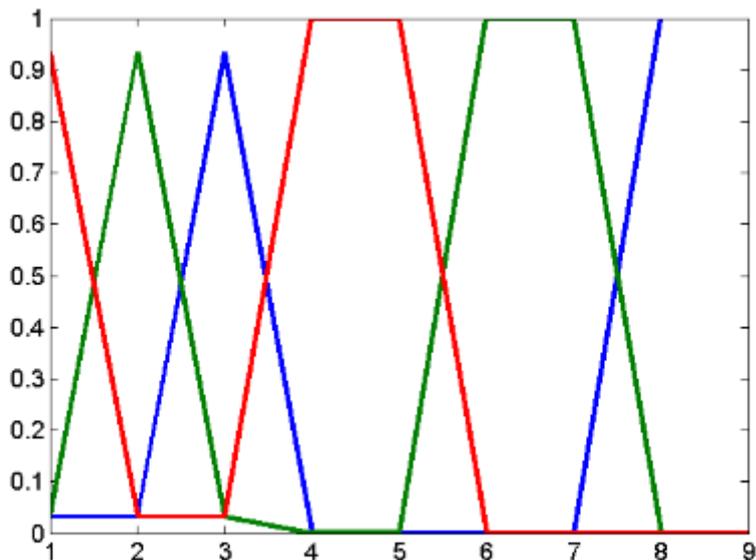


Figure 3: Applying PCCA+ to the network in Fig. 2 after making it reversible. The figure shows the membership χ of the 9 states to the three clusters (colored curves). One can see that state 3 belongs to the “green” and “red” cluster with the same grade of membership, although, the directed graph has only direct transitions from the “blue” to the “red” cluster.

the system. We assume an equal initial distribution. Then, Λ is given by

$$\Lambda = \begin{pmatrix} 1.0000 & 0 & 0 \\ 0 & 0.9483 & 0.0279 \\ 0 & -0.0279 & 0.9483 \end{pmatrix},$$

which corresponds to the eigenvalue analysis. After taking a proper linear combination of the leading Schur vectors, the result of G-PCCA clearly shows the different grades of membership that reflect the directed structure of the network, see Fig. 4.

3.2 Add-on of G-PCCA compared to PCCA+

In order to illustrate which kind of results are to be expected by G-PCCA, we construct a random matrix. b). First, three 10-by-10 random matrices A1, A2, and A3 are constructed using the Matlab-routine [MAT10] “RAND(10,10)”. Another 10-by-10 zero matrix Z is constructed such that “C=[[Z, A1,Z];[Z,Z,A2];[A3,Z,Z]];”. After adding a random matrix with entries between 0 and 0.1 to C, the rows of this matrix are rescaled such that the resulting matrix P is stochastic. In Fig. 5 this transition matrix is depicted with its clearly visible block structure.

According to theory, any positive initial distribution η is possible. We will chose a random initial distribution. The rescaled matrix used for a Schur decomposition is given by $\tilde{P} = D_\eta^{0.5} P D_\eta^{-0.5}$. This matrix has a partial real Schur

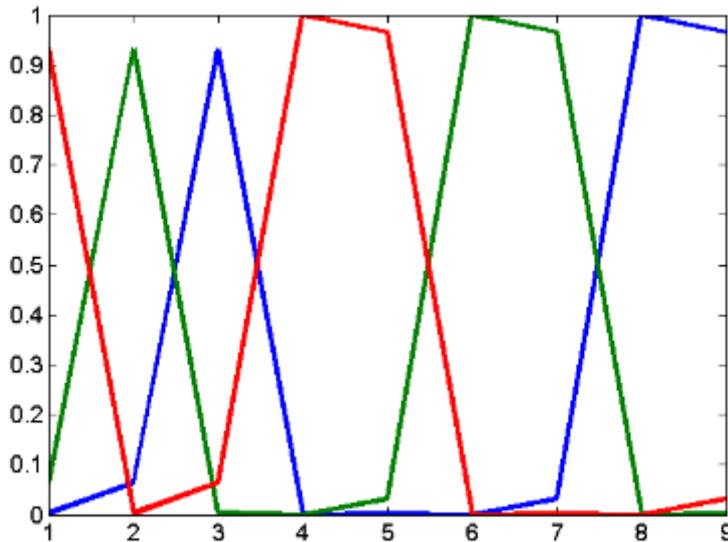


Figure 4: Applying G-PCCA directly to the network in Fig. 2. The figure shows the membership χ of the 9 states to the three cluster. One can see that, e.g., state 3 belongs to the “green” and “red” cluster with a different grade of membership, as we expect it from the directed graph.

decomposition of the form $\tilde{P}\tilde{X} = \tilde{X}\Lambda$ with a non-diagonal matrix Λ . In our realization:

$$\Lambda = \begin{pmatrix} 1.0000 & -0.0267 & -0.0928 \\ 0 & -0.3884 & -0.6836 \\ 0 & 0.6426 & -0.3884 \end{pmatrix}.$$

Besides the diagonal element “1” (Perron root) there is a 2-by-2-block on the diagonal of Λ , which belongs to a complex eigenvalue pair $-0.3884 \pm 0.6628i$ near the unit circle. The absolute values of these three eigenvalues are well separated from the other absolute values. The matrix of rescaled Schur vectors used for G-PCCA is constructed by $X = D_\eta^{-0.5}\tilde{X}$. Note, that the first column vector of X is constant, each element is “1”. This is a necessary condition for the G-PCCA algorithm. Using this matrix for G-PCCA provides a 30-by-3-membership matrix $C = X\Lambda$. The three columns of this matrix are plotted in Fig. 6. They correspond to the three different clusters of states which have a similar transition pattern. This transition pattern is revealed by computing

$$P_C = \begin{pmatrix} 0.0417 & \mathbf{0.8543} & 0.1041 \\ 0.0993 & 0.0591 & \mathbf{0.8416} \\ \mathbf{0.8416} & 0.0762 & 0.0823 \end{pmatrix},$$

where the highest entries are marked. This matrix can be interpreted as the transition matrix between the three clusters of states. Note, that this matrix is not diagonal dominant.

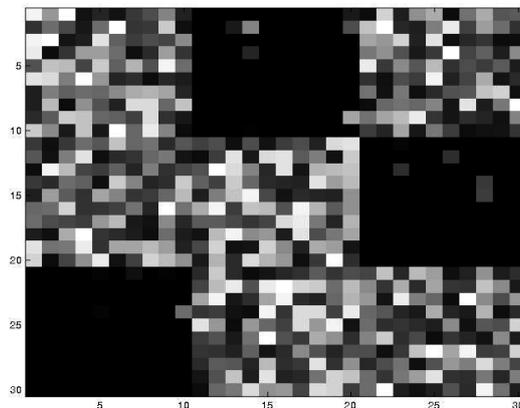


Figure 5: A realization of a 30-by-30-transition matrix b). Gray scale of the entries from white to black.

4 Conclusions

The Galerkin projection of a Markov operator P onto a coarse grained matrix P_C has to be chosen with care, since in general the markovianity is not preserved, which is necessary to map the correct dynamics. We showed, that each projection, which meets the invariant subspace condition (3) and the orthogonality relation (4) preserves markovianity. The novel method G-PCCA is also capable to treat non-reversible Markov chains, by using a Schur decomposition.

Acknowledgment. The work of Marcus Weber has been done for the DFG Collaborative Research Center 765.

References

- [Bra02] J.H. Brandts. Matlab code for sorted real schur forms. *Num Lin Alg App*, 9(3):249–261, 2002.
- [DHFS00] P. Deuffhard, W. Huisinga, A. Fischer, and Ch. Schuette. Identification of almost invariant aggregates in reversible nearly uncoupled markov chains. *Linear Algebra and its Applications*, 315(1-3):39 – 59, 2000.
- [DW05] P. Deuffhard and M. Weber. Robust perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398(0):161 – 184, 2005. Special Issue on Matrices and Mathematical Biology.
- [Fil91] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1):62–87, 1991.

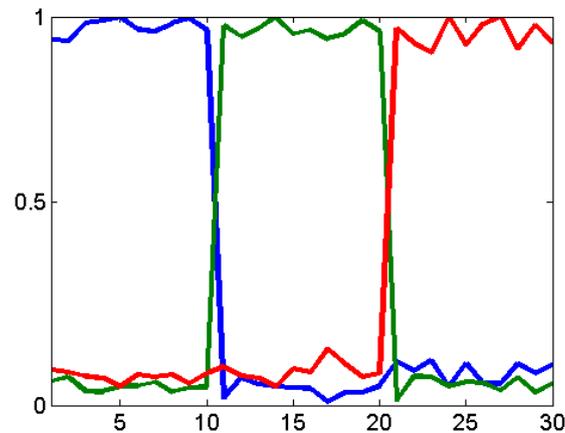


Figure 6: G-PCCA on the basis of Schur vectors can reveal the block-structure of the random matrix in Fig. 5. The three columns of the matrix C are indicating the states which are assigned to each of the clusters. Note, that these are not metastable cluster of states. The states assigned to one of the clusters simply have a common transition pattern.

- [FMSV07] D. Fritzsche, V. Mehrmann, D.B. Szyld, and E. Virnik. An SVD approach to identifying metastable states of Markov chains. *ETNA*, 29:46–69, 2007.
- [FPG14] G. Froyland and K. Padberg-Gehle. Almost-invariant and finite-time coherent sets: directionality, duration, and diffusion. In Gary Froyland Wael Bahsoun, Chris Bose, editor, *Ergodic Theory, Open Dynamics, and Coherent Structures*, volume 70 of *Proceedings in Mathematics and Statistics*, pages 171–216, 2014.
- [HMS04] W. Huisinga, S. Meyn, and C. Schütte. Phase transitions and metastability in markovian and molecular systems. *Ann. Appl. Probab.*, 14(1):419–458, 2004.
- [Jac10] M.N. Jacobi. A robust spectral method for finding lumpings and meta stable states of non-reversible Markov chains. *ETNA*, 37:296–306, 2010.
- [LBDD01] C. Li, G. Biswas, M. Dale, and P. Dale. Building models of ecological dynamics using hmm based temporal data clustering - a preliminary study. In F. Hoffmann et al, editor, *Advances in Intelligent Data Analysis*, volume 2189 of *Lecture Notes in Computer Science*, pages 53–62. Springer Berlin Heidelberg, 2001.
- [MAT10] MATLAB. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts, 2010.
- [MLKCW03] C. S. Möller-Levet, F. Klawonn, K.-H. Cho, and O. Wolkenhauer. Fuzzy clustering of short time-series and unevenly distributed

- sampling points. In *LNCS, Proceedings of the IDA2003*, pages 28–30. Springer Verlag, 2003.
- [RM07] T. Runolfsson and Yong Ma. Model reduction of nonreversible markov chains. In *Decision and Control, 2007 46th IEEE Conference on*, pages 3739–3744, 2007.
- [Sch99] Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm and Application to Biomolecules*. Habilitation thesis, Freie Universität Berlin, 1999.
- [Shu03] R. H. Shumway. Time-frequency clustering and discriminant analysis. *Stat Prob Let*, 63(3):307–314, 2003.
- [SS14] M. Sarich and C. Schütte. Utilizing hitting times for finding metastable sets in non-reversible markov chains. Technical Report 14-32, ZIB, Takustr.7, 14195 Berlin, 2014.
- [TBHY13] J. D. Tjakra, J. Bao, N. Hudon, and R. Yang. Analysis of collective dynamics of particulate systems modeled by markov chains. *Powder Technology*, 235(0):228 – 237, 2013.
- [Tif11] R. Tifenbach. On an SVD-based Algorithm for Identifying Metastable States of Markov Chains. *ETNA*, 38:17–33, 2011.
- [VLKG03] M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos. A wavelet-based anytime algorithm for k-means clustering of time series. In *In Proc. Workshop on Clustering High Dimensionality Data and Its Applications*, pages 23–30, 2003.
- [Web02] M Weber. Clustering by using a simplex structure, 2002.
- [Web06] M. Weber. *Meshless Methods in Conformation Dynamics*. PhD thesis, Freie Universität Berlin, 2006.
- [Web11] M. Weber. A subspace approach to molecular markov state models via a new infinitesimal generator, 2011. habilitation thesis.
- [WL07] T. Warren Liao. A clustering procedure for exploratory mining of vector time series. *Pattern Recogn.*, 40(9):2550–2562, 2007.