Preserving the Markov Property of Reduced Reversible Markov Chains

Marcus Weber and Susanna Kube

Zuse Institute Berlin, Takustraße 7, D-14195 Berlin, Germany

Abstract. The computation of essential dynamics of molecular systems by conformation dynamics turned out to be very successful. This approach is based on Markov chain Monte Carlo simulations. Conformation dynamics aims at decomposing the state space of the system into metastable subsets. The set-based reduction of a Markov chain, however, destroys the Markov property. We will present an alternative reduction method that is not based on sets but on membership vectors, which are computed by the Robust Perron Cluster Analysis (PCCA+). This approach preserves the Markov property.

Keywords: Perron cluster analysis, dimension reduction, conformation dynamics
PACS: 87.15Ak, 02.10Ud

INTRODUCTION

Molecular motion takes place on different time-scales. The binding process of a ligand to a protein is in the range of milliseconds up to seconds, whereas thermal vibrations of atoms have a time-scale of femtoseconds. Mostly, we are interested in the essential behaviour of the system, i.e. its dynamics on larger time-scales. Since thermal vibration is important for the long-time behaviour of the system, we need a multiscale simulation method. For a computation of the essential dynamics of molecular systems, *conformation dynamics* turned out to be a successful approach [1, 2, 3, 4]. In this method, the thermal vibration of the system is simulated by hybrid Monte Carlo (HMC) [5]. HMC combines short-time molecular dynamics simulations with a Metropolis-Hastings type acceptance step. The corresponding HMC-trajectory is a Markov chain in position space Ω. The Markov chain is almost uncoupled, i.e. there are parts of position space in which the trajectory tends to get trapped. These parts are denoted as metastable subsets of Ω. Conformation dynamics computes the stationary density of the system. It identifies the metastable subsets and determines the transition probabilities between the subsets. These transition probabilities characterize the essential dynamics of the system on larger time-scales.

Throughout the paper, we assume that the continuous state space has already been discretized appropriately such that the dynamics can be modeled as Markov chain on a finite state space Ω [6]. The Markov chain on $|\Omega| = N$ can be represented by a $N \times N$ transition probability matrix P. The number n of nearly decoupled sub-chains and thus the number of metastable conformations equals the number of eigenvalues of P close to the Perron root $\lambda = 1$ [1]. These eigenvalues, also denoted as Perron cluster eigenvalues, are well separated from the rest of the spectrum.

For the computation of the essential dynamics, the number of states is reduced from all N states to the n metastable subsets with $n \ll N$. By a decomposition of the state space into these n metastable subsets, the transition matrix P can be reduced to a $n \times n$ matrix P_c, that contains the transition probabilities between the subsets. This matrix, however, does not represent a Markov chain between subsets of Ω, because a set-based reduction of Markov chains does not preserve the Markov property [7].

An illustrative counter example is shown in Figure 1. Assume, there is a Markov chain on a set of states ($N = 36$) with certain transition probabilities, see Fig. 1. All non-zero transition probabilities are indicated by lines. Consider a decomposition of the states into $n = 3$ subsets A, B, and C (Fig. 1, right). Some of the states inside C (indicated by white circles) form a barrier, i.e. there is only a low transition probability from the neighbouring states to these barrier states. Thus, in a realization of the reversible Markov chain, the transition probability $C \to A$ is not independent of the previous time steps. For instance, the probability for a transition $C \to A$ is higher, if the last transition between different sets has been $A \to C$ instead of $B \to C$.

In the following, we will show how a reduction of P can be done such that the Markov property still holds.

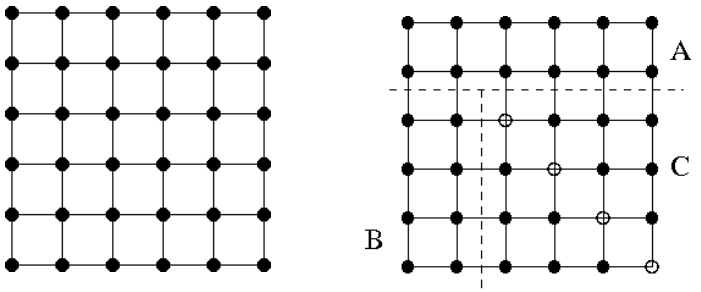

FIGURE 1. The Markov property holds for single states (left) but can not be transferred to subsets of states (right). The transition probability $C \to A$ depends on whether the system has been in B or in A before it enters C.

PROPAGATION OF DENSITIES

There are two possible views on Markov chains: P represents the evolution of a single state in time, whereas $P^\top$ is the propagator of a distribution of states in time. More precisely, the transition matrix P can be used for a time-series realization of the Markov chain. In this time series, a state $q^{(i+1)} = a \in \Omega$ follows a given state $q^{(i)} = b \in \Omega$ with the corresponding conditional probability $P(a,b)$. In contrast to that, density propagation is based on an ensemble of states. Given an initial distribution of states $v^{(0)} \in R^N_+, \|v^{(0)}\|_1 = 1$, the probabilties to reach any of the N states at time-step i are given by a density vector $v^{(i)} \in R^N_+$ and evolve in time according to

$$v^{(i+1)} = P^\top v^{(i)}.$$

This equation defines the *Markov property*: The distribution of states at a certain time step $(i+1)$ only depends on the previous time step (i). The aim is to find a reduction method for the density vectors v and for the matrix P from a number of N states to n metastable states. The reduced vectors are denoted by v_c. In our approach, the two ways

- propagation of v with $P^\top$ first, then reduction to v_c, or
- reduction of v to v_c first, then propagation of v_c with ${P_c}^\top$

commute, i.e., they lead to the same result. Thus, the Markov property of P is transferred to P_c. The reduction of P is done via a Galerkin discretization, and P is assumed to be a reversible Markov chain, i.e. P is D-self adjoint ($DP = P^\top D$), where D is the diagonal matrix of the stationary distribution of the Markov chain P.

DIMENSION REDUCTION

Mathematically, a set-based reduction $P_c \in R^{n\times n}$ of a reversible Markov chain P can be written as [8]

$$P_c = (\chi^\top De)^{-1}\chi^\top DP\chi. \tag{1}$$

In this equation, $e \in R^{N\times n}$ is the constant 1-matrix, and $\chi \in R^{N\times n}$ is the matrix of the characteristic vectors of the n sets, i.e., $\chi(i,j) = 1$, if state j belongs to set i, and $\chi(i,j) = 0$ otherwise. For this matrix χ of characteristic vectors, equation (1) is equal to

$$P_c = (\chi^\top D\chi)^{-1}\chi^\top DP\chi, \tag{2}$$

which is the D-weighted Galerkin discretization of the Markov chain P. For a general choice of χ, equations (1) and (2) are different. In (2) the expression

$$R := \chi^\top$$

can be seen as the restriction operator of the Galerkin discretization of $P^\top$ and

$$I := D\chi(\chi^\top D\chi)^{-\top}$$

as the corresponding interpolation operator [9]. Via $v := D\chi a$ for some $a \in R^n$ we construct a start density in a subspace of $P^\top$. Let X be a D-orthogonal set of eigenvectors of P, with $PX = X\Lambda$ and $X^\top DX = I$. Furthermore, $\chi = X\mathscr{A}$ is a basis transformation of X with a regular matrix $\mathscr{A} \in R^{n\times n}$. Then one can show that

$$RP^\top v = P_c^\top Rv. \tag{3}$$

The equation can be interpreted as follows. If the column vectors of χ stem from an invariant subspace of P, then propagation and restriction commute for a restricted operator P_c constructed according to (2); see also [7]. Thus, in order to preserve the Markov property, we have to find a basis χ of an invariant subspace X of P. Given this basis, restriction R and interpolation I are defined. Starting with a reduced density $v_c = \chi^\top v$ for a given full density $v := D\chi a$, the Markov property is preserved. The only problem is that, in general, a basis χ cannot be interpreted as a decomposition of the state space into subsets. However, via PCCA+ [10, 11] we can find a matrix $\mathscr{A}$ with $\chi = X\mathscr{A}$ such that the columns of χ provide non-negative membership vectors of (fuzzy) subsets of Ω. If the corresponding eigenvalues Λ of the eigenvectors X are part of the Perron cluster eigenvalues of P, the columns of χ define the metastable conformations (long term dynamics). Thus, P_c includes the essential dynamics of the system. Since $P_c = \mathscr{A}^{-1}\Lambda\mathscr{A}$, the eigenvalues of P_c are equal to the dominant eigenvalues of P. This shows that the correct timescales of motion are preserved.

ILLUSTRATIVE EXAMPLE

We constructed a reversible Markov chain P for the example in Fig. 1, such that a crossing of the barrier between two subsets is a rare event. The three dominant eigenvalues of this matrix P are $\lambda_{1,2,3} = \{1, 0.99, 0.99\}$ followed by a gap $\lambda_4 = 0.93$. Thus, P has three metastable subsets. For a set based reduction according to Fig. 1, we get

$$P_c^{(1)} = \begin{pmatrix} 0.9917 & 0.0025 & 0.0058 \\ 0.0016 & 0.9945 & 0.0039 \\ 0.0025 & 0.0025 & 0.9950 \end{pmatrix}. \tag{4}$$

Unfortunately, the reduced matrix P_c does not preserve the Markov property, see Fig. 2 right. For the normalized vector $v = D\chi a$, with $a = (0, 0, 1/0.45)$, the difference for the first propagation step is $\|RPv - P_cRv\| \approx 4 \cdot 10^{-4}$. The error increases for further iteration steps, but then decreases again because the reduction method (4) preserves the steady state solution of the system, and the propagation converges against the steady state. To preserve the Markov property, PCCA+ is applied. The corresponding invariant subspace is given by the matrix X of the three dominant eigenvectors of P. Via PCCA+, we compute a set of membership vectors $\chi = X\mathscr{A}$, see Fig. 2 left. This is done as follows. First, the dominant eigenvalues of P and the corresponding eigenvectors X are determined. Afterwards, via solving an optimization problem, a tranformation matrix $\mathscr{A}$ is constructed such that $\chi = X\mathscr{A}$ is a nonnegative matrix with row sum 1. The optimal solution maximizes the trace of (4) with regard to χ. Using the optimal membership matrix χ, we can compute the reduced Markov chain via (2) as

$$P_c^{(2)} = \begin{pmatrix} 0.9924 & 0.0025 & 0.0051 \\ 0.0018 & 0.9948 & 0.0034 \\ 0.0025 & 0.0025 & 0.9950 \end{pmatrix}. \tag{5}$$

The difference between (4) and (5) is small, but whereas $P_c^{(2)}$ represents a Markov chain, $P_c^{(1)}$ is not interpretable as a transition matrix. It is important to mention that the diagonal entries of $P_c^{(2)}$ are higher than the entries of $P_c^{(1)}$. The trace of $P_c^{(2)}$ equals the sum of the corresponding $n = 3$ dominant eigenvalues of P, which is also an upper bound for the trace of $P_c^{(1)}$ in (1). In (5), the state space is decomposed into fuzzy sets represented by membership vectors χ, see Fig. 2 left. The decomposition shows that A, B and C "overlap" in the transition regions. This fact can be used to characterize transition states of molecular systems [12] by evaluating the membership vectors χ.

REFERENCES

1. C. Schütte, *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*, Habilitation thesis, Department of Mathematics and Computer Science, Freie Universität Berlin (1999).

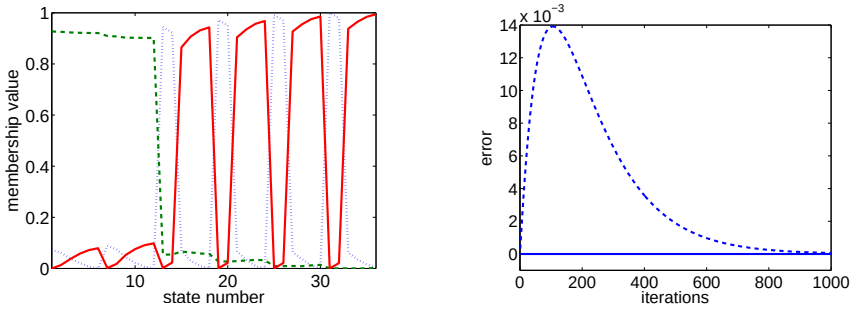

FIGURE 2. Left: Membership values for the 36 states clustered into three (dashed, dotted, and solid line) metastable parts in Fig. 1. Right: Error of set-based reduction and propagation (dashed line). The error of a subspace-based propagation (solid line) is always in the scale of the machine accuracy. The propagated densities converge to the correct reduced steady state solution in the two approaches.

2. P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, *Linear Algebra Appl.* **315**, 39–59 (2000).
3. P. Deuflhard, "From Molecular Dynamics to Conformational Dynamics in Drug Design," in *Trends in Nonlinear Analysis*, edited by M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, Springer, 2003, pp. 269–287.
4. M. Weber, R. Becker, R. Köppen, and V. Durmaz, Classical hybrid Monte-Carlo simulations of the interconversion of hexabromocyclododecane, ZIB-Report 07-31, Zuse Institute Berlin (2007).
5. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216–222 (1987).
6. J. D. Chodera, N. Singhal, W. C. Swope, V. S. Pande, and K. A. Dill, *J. Chem. Phys.* **126**, 155101 (2007).
7. M. Weber, Conformation based Transition State Theory, ZIB-Report 07-18, Zuse Institute Berlin (2007).
8. C. D. Meyer, *SIAM Rev.* **31**, 240–272 (1989).
9. S. Kube, and M. Weber, *J. Chem. Phys.* **126** (2007).
10. P. Deuflhard, and M. Weber, "Robust Perron Cluster Analysis in Conformation Dynamics," in *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, edited by M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte, Elsevier, 2005, vol. 398C, pp. 161–184.
11. M. Weber, *Meshless Methods in Conformation Dynamics*, Doctoral thesis, Department of Mathematics and Computer Science, Freie Universität Berlin (2006), published by Verlag Dr. Hut, München.
12. M. Weber, and T. Galliat, Characterization of Transition States in Conformational Dynamics using Fuzzy Sets, ZIB-Report 02-12, Zuse Institute Berlin (2002).