Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

MARCUS WEBER

Conformation-Based Transition State Theory

ZIB-Report 07-18 (July 2007)

Conformation-Based Transition State Theory*

Marcus Weber

July 27, 2007

Abstract

For the treatment of equilibrated molecular systems in a heat bath we propose a transition state theory that is based on conformation dynamics. In general, a set-based discretization of a Markov operator $\mathcal{P}^\tau$ does not preserve the Markov property. In this article, we propose a discretization method which is based on a Galerkin approach. This discretization method preserves the Markov property of the operator and can be interpreted as a decomposition of the state space into (fuzzy) sets. The conformation-based transition state theory presented here can be seen as a first step in conformation dynamics towards the computation of essential dynamical properties of molecular systems without time-consuming molecular dynamics simulations.

AMS MSC 2000: 62H30, 82B80, 65C40,

Keywords: dynamical systems, transition state theory, rare events, metastability, cluster analysis

*Supported by the DFG Research Center MATHEON "Mathematik für Schlüsseltechnologien"

Contents

Introduction

In this article, we focus on the dynamics of a molecule in an equilibrated heat bath. The dynamics is assumed to be a time-harmonic Markov process in position space $\overline{\Omega}$[1]. For the theoretical background of such processes[2] see [5]. The equilibrium is assumed to be in detailed balance with a Boltzmann distribution of states. For Fig. 1 a corresponding simulation in the high-dimensional space $\overline{\Omega}$

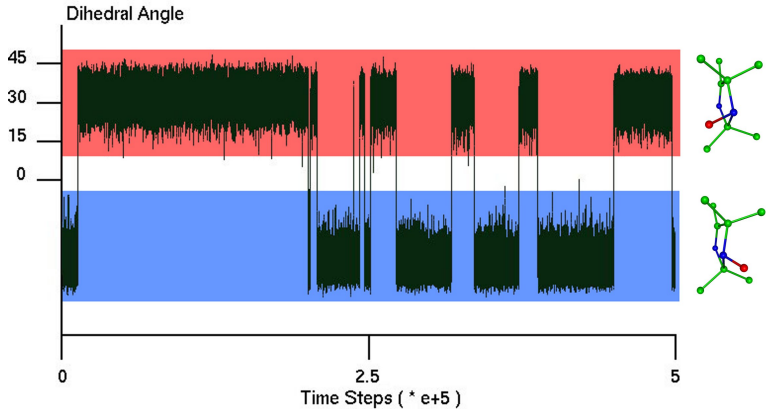

Figure 1: A typical time series in conformation dynamics. The position space $\overline{\Omega}$ of the investigated molecule can be decomposed into $n = 2$ dynamically metastable regions.

has been performed for a small molecule using a hybrid Monte Carlo method [3]. Subsequently, an internal coordinate (a certain dihedral angle) of the molecule has been selected that indicates conformational changes [1]. For the theory of conformations defined as metastable states see [12]. The time series analysis of the selected coordinate in Fig. 1 shows the following: The molecule has $n = 2$ dynamically metastable states. On the basis of this insight, we are interested in a finite state approximation of the system, similar to [9]. In this example, we are interested in a matrix $P_c \in \mathbb{R}^{n\times n}$ which captures the correct transition behaviour between the $n = 2$ metastable states, see also [4, 11]. This article pursues the following aims:

1. The reduced matrix P_c should be constructed in such a way, that it describes the full-dimensional dynamics correctly.

2. The matrix P_c should be computed without using a full-dimensional simulation. This aim is motivated by the fact that in these simulations, transitions are rare events, which normally leads to insufficient statistics.

The problems related to the first aim will be discussed: A set-based reduction (indicated in Fig. 1) does not lead to the correct matrix P_c. The second aim

[1] $\overline{\Omega}$ includes all possible position states of the molecule neglecting translation and rotation. If the molecule has k atoms, this is a $(3k-6)$-dimensional space. $\overline{\Omega}$ is also referred to as *state space*. Because from the viewpoint of the Markov process, it is the space of all possible states of the system.

[2] In the continuous case, a first-order stochastic differential equation leads to a dynamical process without memory. It is Markovian.

needs some approximations. These lead to *conformation-based transition state theory*. In contrast to *standard transition state theory*, the new approach can be applied to molecular systems with arbitrarily complex potential energy surfaces. It takes into account entropical effects and is not restricted to a two-state system. The accuracy of the new theory is related to the approximation properties of a Galerkin discretization method, i.e. the better the approximation, the better the results from conformation-based transition state theory. For a historical overview of standard transition state theory see [7].

Outline. We first focus on the investigation of finite state spaces $|\Omega| = N$. The derivation of a correct reduced transition matrix P_c is organized as follows: In Eq. 3 on page 6, an intuitive but incorrect set-based reduction of Markov chains is used. In Eq. 5 on page 7, the incorrect term (3) is equivalently transformed into a Galerkin discretization approach, and subsequently, the correct choice of basis functions for discretization is derived using Robust Perron Cluster Analysis PCCA+. With the special choice of basis functions, P_c in Eq. 5 is the correct reduced matrix for a finite state space. Based on transition state theory, the computation of the correct reduced matrix P_c in infinite state spaces $\overline{\Omega}$ is derived in Eq. 12 on page 10. This computation can be based on a spectral analysis of a rate matrix Q (18) that can be approximated numerically without molecular dynamics simulations via Monte Carlo quadrature and equations 23 and 26 below.

1 Dimension Reduction of Reversible Markov Chains

1.1 Two Views on Markov Chains

Transition matrix. Given a sequence of random variables

$$X_0 \to X_1 \to X_2 \to \dots,$$

this sequence is called a *Markov chain*, if the probability for the next step only depends on the preceding step, i.e. if it meets the *Markov property*:

$$\begin{aligned} &P(X_i = q^{(i)} | X_{i-1} = q^{(i-1)} \wedge X_{i-2} = q^{(i-2)} \wedge \dots \wedge X_0 = q^{(0)}) \\ &= P(X_i = q^{(i)} | X_{i-1} = q^{(i-1)}). \end{aligned} \tag{1}$$

If the set Ω of states $q \in \Omega$ is finite $|\Omega| = N$, then for every pair of states, there is one conditional probability in Eq. 1. By enumerating the states, these probabilities are specified in a transition matrix $P \in \mathbb{R}^{N \times N}$. P is a *stochastic matrix*, i.e. the elements of P are non-negative and the row sums are 1.

Markov chain as a time series realization. Given a transition matrix P and the Markov property, a time series realization of the Markov chain is possible. In this time series, state $q^{(i)} = a$ follows a given state $q^{(i+1)} = b$ with the corresponding probability $P(a, b)$. At each step i of the time series, the system has one defined state $q^{(i)} \in \{1, \dots, N\}$.

Markov chain as a density propagation. A density propagation is based on an ensemble of states. Given an initial distribution of states $v^{(0)} \in \mathbb{R}^N$, $\|v^{(0)}\|_1 = 1$, the probabilites to reach any of the N states at time-step i are given by a density vector $v^{(i)} \in \mathbb{R}^N, i \in \mathbb{N}$, and evolve in time with

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

The two views are based on Marcov chains' realizations as time series. In the first case, it is the evolution of a single state in time. In the second case, it is the evolution of a distribution of states in time.

Reversibility. In the following, we only treat ergodic dynamical systems in detailed balanced equilibrium. The stationary density $\pi \in \mathbb{R}^N$ of such a process is uniquely defined by the equation

$$\pi^\top P = \pi^\top.$$

Uniqueness can be proven by the theorem of Frobenius-Perron, see [10]. Equilibrated systems obeying a detailed balance condition are characterized by a reversible transition matrix P. The unconditional probability for a transition $i \to j$ is equal to the probability for the backward step $j \to i$. Thus, a reversible Markov chain is defined by the equality

$$DP = P^\top D, \tag{2}$$

where $D = \mathrm{diag}(\pi_1, \ldots, \pi_N)$ is a diagonal matrix of the stationary density vector π.

1.2 Non-Markovian Set-Based Reduction

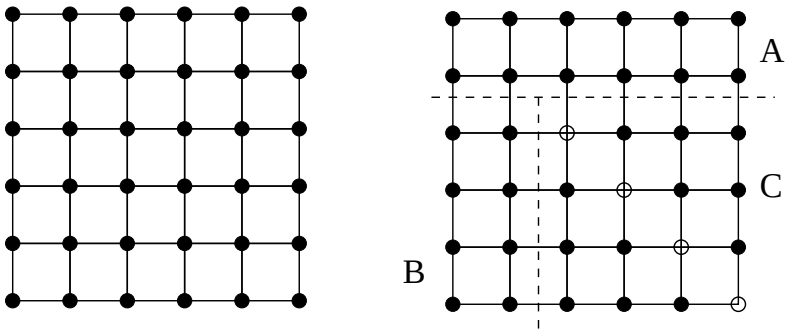

Figure 2: A Markov property which holds for single states (left) can not be transferred to sets of states (right). The transition probability $C \to A$ depends on whether the system has been in B or in A before it entered C.

Characteristic functions. The basis for the following considerations is a Markov chain given by a stochastic matrix $P \in \mathbb{R}^{N\times N}$. We are interested in a dimension reduction of this matrix into a smaller transition matrix $P_c \in \mathbb{R}^{n\times n}$, $n \ll N$. This reduced transition matrix should capture the transition probabilities between subsets of Ω. The n subsets of Ω are represented by a matrix $\chi \in \{0,1\}^{N\times n}$. The element $\chi(i,j)$ is 1 if and only if state i belongs to subset j. Since we are interested in a *decomposition* of Ω into subsets, χ has row sum 1. Although χ is a matrix, its columns are referred to as *characteristic functions* of the subsets. This name becomes clear as soon as we switch from discrete to continuous problems in Section 2.1.

Set-based reduction. Intuitively, in order to compute the matrix $P_c \in \mathbb{R}^{n\times n}$ of transition probabilities between the subsets defined by χ, we have to restrict the unconditional transition probabilities DP to the corresponding sets via $\chi^\top DP\chi$. The matrix elements of $\chi^\top DP\chi$ are the unconditional transition probabilities between the subsets. In order to get the conditional probabilities, we have to normalize this matrix, i.e.

$$P_c = \widetilde{D}^{-1}\chi^\top DP\chi. \tag{3}$$

In Eq. 3, the matrix $\widetilde{D}$ is a diagonal matrix $\widetilde{D} = \mathrm{diag}(\pi_c)$, where $\pi_c = \chi^\top\pi \in \mathbb{R}^n$ is the reduced stationary density of P_c. For the construction (3) of P_c see also [13].

Does the matrix P_c represent a Markov chain between the subsets of Ω? That is, does $P_c(i,j)$ represent the transition probability from set i to set j for each step of a time series realization of P? The answer is no.

Illustrative counter-example. For an illustrative example of how the Markov property is violated in this kind of reduction technique see Fig. 2. Assume there is a set Ω of states (in this example $N = 36$) and certain transition probabilities between them which meet the Markov property and reversibility. In Fig. 2, all non-zero transition probabilities are indicated by lines. Now assume a decomposition of the states into $n = 3$ sets A, B, and C (Fig. 2, right). Furthermore, some of the states (indicated by white circles) inside set C form a barrier, i.e. there is only a low transition probability from the neighboring states to these barrier states. Clearly, in a realization of the reversible Markov chain, the transition probability $C \to A$ is not independ 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 $A \to C$ than if it has been $B \to C$.

Role of P_c. P_c is not a Markov chain, it does not represent a successive transition probability between subsets of Ω in a time series realization because it has lost the Markov property (1). Nevertheless, P_c is a conditional transition matrix: In an equilibrated reversible system, a constant fraction of the ensemble of states undergoes a transition $i \to j$ in τ time units. If we relate this fraction to the fraction of states that started in i, as in Eq. 3, a conditional probability is defined. P_c is an *ensemble-based transition matrix* and cannot be used for a time series realization with a single starting point $i \in \{1,\ldots,n\}$. In the next section, it will be shown how a correct reduced propagator $P_c^\top$ for a time series realization can be defined.

1.3 Subspace-Based Galerkin Discretization

$P_c^\top$ as a propagator. In this section, we will focus on the second view on Markov chains described in Section 1.1. The transition probability matrix $P \in \mathbb{R}^{N\times N}$ leads to a propagation matrix $P^\top$ for density vectors $v \in \mathbb{R}^N$. In an iterative propagation of densities $v^{(0)}, v^{(1)}, v^{(2)} \ldots$ via

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

$v^{(i)}$ is a non-negative vector with $\|v^{(i)}\|_1 = 1$ for all $i \in \mathbb{N}$. $P_c \in \mathbb{R}^{n\times n}$ gives rise to a propagator $P_c^\top$ as well. $P_c^\top$ propagates reduced densities $v_c^{(i)} \in \mathbb{R}^n, i \in \mathbb{N}$. $P^\top$ and $P_c^\top$ lead to the same dynamical behaviour if v_c is the restriction of v for every iteration step, i.e. $v_c^{(i)} = \chi^\top v^{(i)}$ for all $i \in \mathbb{N}$. This is not the case for $P_c^\top$ in (3). A correct reduced propagator $P_c^\top$ meets the following condition: **The propagation of the densities via $P^\top$ reduced to the subsets defined by χ should be equal to the propagation of the reduced densities via $P_c^\top$.** We will see that $P_c^\top$ meets this condition with the correct choice of $\chi \in \mathbb{R}^{N\times n}$ and a slight change of (3) into (5) below.

Galerkin discretization. In order to explain the above condition, we clarify the mathematical meaning of (3) and introduce the following expression:

$$\langle u, v\rangle_\pi := u^\top D v. \tag{4}$$

Equation 4 defines a weighted scalar product. Reversibility of P in Eq. 2 in terms of this scalar product is equivalent to π-self-adjointness of P, see [11, 13]. A short calculation for the expressions in Eq. 3 yields

$$\widetilde{D} = \langle \chi, \chi\rangle_\pi, \quad \chi^\top D P \chi = \langle \chi, P\chi\rangle_\pi,$$

where we used the fact that $\chi(i,j) \in \{0,1\}$ for all elements of χ. Thus, the set-based reduction (3) in terms of the scalar product (4) becomes

$$P_c^\top = \langle P\chi, \chi\rangle_\pi \left(\langle \chi, \chi\rangle_\pi\right)^{-1}. \tag{5}$$

Equation 5 and Eq. 3 only differ if there exist elements $\chi(i,j) \notin \{0,1\}$. Equation 5 is known as the Galerkin discretization of a π-self-adjoint operator $P^\top$ for a non-π-orthonormal set of basis vectors given by χ. This discretization method defines two matrices [6], one matrix $I \in \mathbb{R}^{N\times n}$ is known as the *interpolation operator*

$$I := \langle \cdot, \chi\rangle_\pi \left(\langle \chi, \chi\rangle_\pi\right)^{-1} = D\chi \left(\chi^\top D \chi\right)^{-1}$$

and another matrix $R \in \mathbb{R}^{n\times N}$ with $RI = id$ is known as the *restriction operator*

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

Correct propagator $P_c^\top$. The condition introduced at the beginning of this section for a correct reduced propagator $P_c^\top$ is equivalent to the condition that for each step of the propagation of a density $v \in \mathbb{R}^N$, restriction and propagation commutes:

$$RP^\top v = P_c^\top R v. \tag{6}$$

Using (5), the condition (6) is satisfied for special densities v given by

$$v = D\chi a, \tag{7}$$

with any vector $a \in \mathbb{R}^n$. In order to map v to a reduced density $v_c \in \mathbb{R}^n$ correctly, one has to choose $a = (\chi^\top D\chi)^{-1} v_c$. In this case, v is the interpolation $v = Iv_c$ of v_c. Equation 7 is a subspace condition for v. $P_c^\top$ is the correct reduced operator if v stays in the subspace defined by (7) during the propagation, i.e. if $P^\top v \in \text{span}\{D\chi\}$. This is the case if the rows of χ span an invariant subspace of P:

$$P^\top v = P^\top D\chi a = DP\chi a = D\chi Aa \in \text{span}\{D\chi\},$$

with $A \in \mathbb{R}^{n\times n}$ and $P\chi = \chi A$.

1.4 Interpretation of Subspaces as Subsets

Problem. In Section 1.3, it has been shown that $P_c^\top$ is the correct reduced propagator of $P^\top$ if it is a Galerkin discretization (5) of $P^\top$ with a set of basis vectors χ from an invariant subspace of P. Starting with a density $v = D\chi a$, the propagators $P^\top$ and $P_c^\top$ lead to the same dynamics. Although P is a stochastic matrix and can be interpreted as a Markov chain, the entries of the correct reduced propagator P_c may be negative. Furthermore, the set-based definition of χ as a decomposition of Ω into subsets in Section 1.2 was very intuitive, whereas in Section 1.3, the interpretation of an invariant subspace χ of P as a decomposition of Ω into parts is not immediately clear. We now explain how Robust Perron Cluster Analysis (PCCA+) interprets subspaces as subsets [2, 13].

Robust Perron Cluster Analysis. For a reversible matrix P, the n-dimensional invariant subspaces χ always have a basis representation in terms of eigenvectors of P, i.e. $\chi = X\mathcal{A}$, with a matrix $X \in \mathbb{R}^{N\times n}$ of eigenvectors of P and $\mathcal{A} \in \mathbb{R}^{n\times n}$. Given the set of dominant eigenvectors X, PCCA+ constructs a regular matrix $\mathcal{A}$ in such a way, that χ can be interpreted as *almost* characteristic functions of a decomposition of Ω. The entries of χ are the corresponding membership values, i.e. $\chi(i,j) \in [0,1]$ denotes the degree of membership of state i w.r.t. (fuzzy) set j, see also [14]. χ can be seen as a decomposition of Ω because via PCCA+, χ is a partition-of-unity

$$\sum_{j=1}^{n} \chi(i,j) = 1, \quad \forall i = 1,\ldots,N.$$

By perturbation theory, it has been shown that if the Markov chain P has n metastable parts, the entries of χ are close to 0 or 1, see [2].

Conclusion. In summary, a correct reduced propagator $P_c^\top$ can be constructed from $P^\top$ via a Galerkin discretization method (5). The corresponding set of basis vectors χ stems from a Robust Perron Cluster Analysis applied to the dominant eigenvectors X of P. Using this analysis, χ can be seen as a (fuzzy) decomposition of Ω into metastable regions.

2 Conformation-Based Transition State Theory (TST)

2.1 Continuous State Spaces

Construction of P. Section 1 focused on transition matrices P in finite state spaces Ω. In molecular dynamics, however, the state space $\overline{\Omega}$ is continuous. This section investigates simulations of molecular motion in a heat bath. In this case, $\overline{\Omega}$ is the position space of the molecule (relative positions of its atoms, neglecting rotation and translation of the molecule in space), and the dynamics is a reversible time-continuous Markov process given by an *infinitesimal generator* $\mathcal{Q}$ (stochastic motion in equilibrium due to heat bath contact) [4]. Thus, a finite transition matrix P is not given but has to be constructed as a discretization of the continuous process. The first step is a time discretization. We compute transition probabilities for certain time intervals $\tau > 0$. From an abstract point of view: If we assume the Markov process to be time-harmonic, then this kind of discretization is, in fact, a computation of a Markov operator $\mathcal{P}^\tau : L_2(\overline{\Omega}) \to L_2(\overline{\Omega})$ from the infinitesimal generator $\mathcal{Q} : L_2(\overline{\Omega}) \to L_2(\overline{\Omega})$ of the process via

$$\mathcal{P}^\tau = \exp(\tau\, \mathcal{Q}). \tag{8}$$

The corresponding theory is given in [5]. The second step is a discretization of $\overline{\Omega}$ into N subsets given by characteristic functions $\Phi_i : \overline{\Omega} \to \{0, 1\}$. This space discretization leads to a Galerkin discretization of $\mathcal{P}^\tau$ similar to Eq. 5. The result is the "transition matrix" $P \in \mathbb{R}^{N\times N}$:

$$P = (\langle \Phi, \Phi \rangle_{\overline{\pi}})^{-1}\ \langle \Phi, \mathcal{P}^\tau \Phi \rangle_{\overline{\pi}}, \tag{9}$$

where $\overline{\pi} : \overline{\Omega} \to \mathbb{R}$ is the continuous stationary distribution in $\overline{\Omega}$.

Role of P. Although P in (9) may be a good discretization of $\mathcal{P}^\tau$, it is not the correct Markov chain (cf. the argumentation of Sections 1.1-1.2 for a set-based dimension reduction of reversible Markov chains again). Therefore, in this section, we do not hope to get a correct propagator on the basis of P. Nevertheless, P plays an improtant role in transition state theory. It is used to approximate eigenfunctions of $\mathcal{P}^\tau$ in the following way. If $X \in \mathbb{R}^{N\times n}$ is the matrix of the n dominant eigenvectors of P, then the dominant eigenfunctions $\xi_i, \ldots, \xi_n : \overline{\Omega} \to \mathbb{R}$ of $\mathcal{P}^\tau$ are approximated by

$$\xi = \Phi X, \text{ i.e.} \quad \xi_j = \sum_{i=1}^{N} X(i,j)\, \Phi_i, \quad j = 1, \ldots, n. \tag{10}$$

correct reduced propagator $P_c^\top$. The functions ξ in (10) can not be interpreted as a decomposition of $\overline{\Omega}$ into parts, but if we apply PCCA+ for the computation of $\chi = X\mathcal{A} \in \mathbb{R}^{N\times n}$, then the eigenfunctions can be transformed into membership functions $\overline{\chi}_1, \ldots, \overline{\chi}_n : \overline{\Omega} \to [0, 1]$ via

$$\overline{\chi} = \xi\mathcal{A}, \text{ i.e.} \quad \overline{\chi}_j = \sum_{k=1}^{n} \mathcal{A}(k,j)\, \xi_k, \quad j = 1, \ldots, n. \tag{11}$$

A Galerkin dicretization of $\mathcal{P}^\tau$ with this non-$\overline{\pi}$-orthonormal set of basis functions $\overline{\chi}_1, \ldots, \overline{\chi}_n$ analogously to Eq. 5 leads to an interpretable and correct reduced propagator $P_c^\top$ of $\mathcal{P}^\tau$. The corresponding scalar product is

$$\langle u, v \rangle_{\overline{\pi}} = \int_{\overline{\Omega}} u(q)\, v(q)\, \overline{\pi}(q)\, dq.$$

2.2 Approximation of Rate Matrices

In the previous section, it has been shown how the correct reduced propagator P_c^τ can be formulated for a continuous dynamical process: It is the Galerkin discretization of $\mathcal{P}^\tau$ with basis functions $\overline{\chi}$. If we assume the dynamics of a molecule in a heat bath as a reversible time-harmonic Markov process, then we can approximate P_c^τ via transition state theory without computing terms like $\mathcal{P}^\tau \overline{\chi}_i$ in the following way.

Infinitesimal generator of $P_c^\top$. TST is a heuristical approach to the computation of transition rates between molecular conformations without dynamics simulations. In this section, we make an approximative assumption. We assume that ξ_i in Eq. 10 is an eigenfunction of $\mathcal{P}^\tau$ with eigenvalue $\lambda_i \approx 1$. Via (8), ξ_i is an eigenfunction of $\mathcal{Q}$ with eigenvalue θ_i and $\lambda_i = \exp(\tau\, \theta_i)$. Using this assumption, Eq. 11, and $\overline{\pi}$-orthonormality of the set of eigenfunctions ξ, we get

$$\begin{aligned}
P_c^\top &= \langle \mathcal{P}^\tau \overline{\chi}, \overline{\chi} \rangle_{\overline{\pi}} \left(\langle \overline{\chi}, \overline{\chi} \rangle_{\overline{\pi}} \right)^{-1} \\
&= \langle \mathcal{P}^\tau \xi \mathcal{A}, \xi \mathcal{A} \rangle_{\overline{\pi}} \left(\langle \xi \mathcal{A}, \xi \mathcal{A} \rangle_{\overline{\pi}} \right)^{-1} \\
&= \langle \xi \Lambda \mathcal{A}, \xi \mathcal{A} \rangle_{\overline{\pi}} \left(\langle \xi \mathcal{A}, \xi \mathcal{A} \rangle_{\overline{\pi}} \right)^{-1} \\
&= \mathcal{A}^\top \Lambda \langle \xi, \xi \rangle_{\overline{\pi}} \mathcal{A} \left(\mathcal{A}^\top \langle \xi, \xi \rangle_{\overline{\pi}} \mathcal{A} \right)^{-1} \\
&= \mathcal{A}^\top \Lambda \mathcal{A} \left(\mathcal{A}^\top \mathcal{A} \right)^{-1} = \mathcal{A}^\top \Lambda \mathcal{A}^{-\top}, \qquad (12)
\end{aligned}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of dominant eigenvalues of $\mathcal{P}^\tau$. Equation 12 proves the existence of an infinitesimal generator $Q_c^\top$ of $P_c^\top$, because $P_c^\top = \exp(\tau Q_c^\top)$ with

$$Q_c^\top = \mathcal{A}^\top \Theta \mathcal{A}^{-\top},$$

and $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_n)$. In other words, if we know the infinitesimal generator $\mathcal{Q}$, its dominant eigenspectrum, and the transformation matrix $\mathcal{A}$ from PCCA+, we can compute the reduced propagator $P_c^\top$ (for arbitrary time intervals τ) and its infinitesimal generator $Q_c^\top$. Note that for this approach, it is not important that Φ_i is a characteristic function. It is important that Φ is a sufficient basis for the approximation of the dominant eigenfunctions ξ of $\mathcal{P}^\tau$. This fact will be discussed in Section 2.3 below.

Rate matrix. Instead of computing the matrix P, we will construct the discretization $Q \in \mathbb{R}^{N \times N}$ of $\mathcal{Q}$ with the set of basis functions Φ. Via $QX = X\Theta$, we can derive the important terms (eigenvectors X, $\Lambda = \exp(\tau\Theta)$, and $\mathcal{A}$) for

(12) from this matrix, as well. It is known from theory [5] that Q is a *rate matrix* with the following structure:

$$Q = R\,(K - id). \tag{13}$$

In Eq. 13, $R \in \mathbb{R}_+^{N\times N}$ is a diagonal matrix of *rate factors* $r = (r_1, \ldots, r_N)^\top$ and $K \in \mathbb{R}_+^{N\times N}$ is a stochastic matrix, denoted as the *embedded Markov chain*. The diagonal elements of K are 0. The non-diagonal element $K(i,j)$ denotes the probability that a process starting in subset Φ_i in the moment it switches to another subset it enters Φ_j and not any other neigboring subset. If one analyzes molecular dynamics simulations (similar to Fig. 1) for the computation of the discretized operator Q using a set-based approach, then the probability $K(i,j)$ is not independent of the past steps because the set-based approach does not preserve the Markov property. Similar to P_c in Section 1.2, Q is an *ensemble-based rate matrix*: If the initial states are distributed according to $\overline{\pi}$, then Q represents a one-step realization of the whole system. This ensemble-based viewpoint is the basis for TST.

Embedded Markov chain. The transition probabilities in K are given by TST in the following way. Since we discuss dynamical systems in equilibrium, the probability to observe a state on the boundary between Φ_i and Φ_j is given by the weight of this boundary. Formally, the overlap of Φ_i and Φ_j has the characteristic function $(\Phi_i \cdot \Phi_j)$. This function has measure 0 w.r.t. $\overline{\pi}$. For a non-zero-measurable relaxation of this expression see Section 2.3 below. Having this relaxation in mind, the probability to observe a state at the boundary $(\Phi_i \cdot \Phi_j)$ is proportional to $\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}$. Since the dynamical process is reversible, it will cross the boundary from Φ_i to Φ_j with the same probability as vice versa, i.e. the non-diagonal elements $K(i,j)$ are proportional to $\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}$. If we use the fact that the embedded Markov chain K has row-sum 1, we get the corresponding normalized matrix as

$$K(i,j) = \begin{cases} \frac{\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}}{\langle \Phi_i \rangle_{\overline{\pi}} - \langle \Phi_i, \Phi_i \rangle_{\overline{\pi}}} & \text{, for } i \neq j, \\ \\ 0 & \text{, for } i = j, \end{cases} \tag{14}$$

where $\langle \Phi_i \rangle_{\overline{\pi}} = \int_{\overline{\Omega}} \Phi_i(q)\, \overline{\pi}(q)\, dq$. Due to the denominator, K is not defined if $\Phi_1, \ldots, \Phi_N$ are characteristic functions of subsets of $\overline{\Omega}$. This aspect will be discussed in Section 2.3 below.

Rate factors. Via (14) and (13), Q is known except for the rate factor matrix R. Note that π is the stationary density of the process, i.e.

$$0 = \pi^\top Q \quad \Leftrightarrow \quad 0 = \pi^\top R\,(K - id). \tag{15}$$

In Eq. 15, the stationary density π is assumed to be known. We are looking for the unknown rate factors $R = \mathrm{diag}(r_1, \ldots, r_N)$. (15) is equivalent to

$$r^\top D\,(K - id) = 0, \tag{16}$$

where $D(i,i) = \pi_i = \langle \Phi_i \rangle_{\overline{\pi}}$. Equation 16 is an eigenproblem for the vector of rate factors $r \in \mathbb{R}^N$. A short calculation shows that the solution is given by

$$r_i = \mu \frac{\langle \Phi_i \rangle_{\overline{\pi}} - \langle \Phi_i, \Phi_i \rangle_{\overline{\pi}}}{\langle \Phi_i \rangle_{\overline{\pi}}} \tag{17}$$

with an unknown common scalar factor $\mu > 0$. Uniqueness again can be shown by the Frobenius-Perron theorem [10]. Inserting the solution R into (13) and (14) leads to the result:

$$Q(i,j) = \mu \cdot \begin{cases} \frac{\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}}{\langle \Phi_i \rangle_{\overline{\pi}}} & , \text{for } i \neq j, \\ \\ \frac{\langle \Phi_i, \Phi_i \rangle_{\overline{\pi}}}{\langle \Phi_i \rangle_{\overline{\pi}}} - 1 & , \text{for } i = j. \end{cases} \tag{18}$$

The only missing information is the time scale factor μ. In the next section, the corresponding algorithmic treatment of this missing information and of the computation of Q is shown.

2.3 Algorithmic Ideas for Rate Matrix Computation

Computation of Q. The algorithmic details in this section include all expressions in Eq. 18:

1. Computation of integrals of the form $\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}$. From these integrals, we also get $\langle \Phi_i \rangle_{\overline{\pi}} = \sum_{j=1}^{N} \langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}$ because Φ is a partition of unity.

2. Computation of the time scale factor μ.

These computations are possible because for a closed molecular system with constant temperature T and known potential energy function $V : \overline{\Omega} \to \mathbb{R}$, the density function $\overline{\pi}$ is given by the *Boltzmann distribution*

$$\overline{\pi}(q) \propto \exp(-\beta V(q)) \tag{19}$$

with inverse temperature $\beta = \frac{1}{k_B T}$ and Boltzmann constant k_B.

Sets and fuzzy sets. The algorithmic treatment of the results of Section 2.2 is not only based on characteristic functions Φ_i of subsets of Ω, because this kind of decomposition of $\overline{\Omega}$ is not suitable for the computation of integrals. It is also based on a kind of relaxation of Φ_i into differentiable membership functions. Whenever we need sets, this is indicated by a basis function Φ'_i. Whenever we apply membership functions or *fuzzy sets*, this is indicated by Φ_i with $\Phi'_i \approx \Phi_i$. In the following, a Voronoi tessellation Φ' of $\overline{\Omega}$ is intended. For this kind of decomposition, which is based on *nodes* $q_1, \ldots, q_N \in \overline{\Omega}$ and a distance measure $d : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}_+$, a relaxed counterpart Φ is known which depends on a relaxation factor $\alpha > 0$, see [13]:

$$\begin{aligned} \Phi'_i(q) &= \begin{cases} 1, & \text{if } d(q,q_i) = \min\limits_{j=1,\ldots,N} d(q,q_j), \\ 0, & \text{else} \end{cases}, \\ \\ \Phi_i(q) &= \frac{\exp(-\alpha\, d^2(q,q_i))}{\sum\limits_{j=1}^{N} \exp(-\alpha\, d^2(q,q_j))}. \end{aligned} \tag{20}$$

For $\alpha \to \infty$ there is a point-wise convergence $\Phi \to \Phi'$.

Metropolis Monte Carlo quadrature. Metropolis Monte Carlo methods can be used to generate a set of points $q \in \overline{\Omega}$ distributed according to a density function which is known except for a normalization constant. In this case, we use Metropolis Monte Carlo in order to compute the integrals $\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}, i, j = 1, \ldots, N$. For each of these integrals, we generate a set $\{q_{ij}^{(1)}, \ldots, q_{ij}^{M(i,j)}\} \subset \overline{\Omega}$ of $M(i,j) \gg 0$ points distributed according to

$$\exp(-\beta V(q))\, \Phi_i(q)\, \Phi_j(q) > 0. \tag{21}$$

For Metropolis Monte Carlo we need not perform molecular dynamics simulations. There are more sophisticated sampling strategies in literature. From each sampling we pick the sampling point $\bar{q}_{ij}$ which maximizes (21) and apply reweighting strategies similar to that in [15]. This is done in the following way. With the Dirac delta-function $\delta : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$, the following equation holds:

$$\frac{\exp(-\beta\, V(\bar{q}_{ij}))\, \Phi_i(\bar{q}_{ij})\, \Phi_j(\bar{q}_{ij})}{\int_{\overline{\Omega}} \exp(-\beta V(q))\, \Phi_i(q)\, \Phi_j(q)\, dq} = \int_{\overline{\Omega}} \delta(\bar{q}_{ij}, q) \frac{\exp(-\beta\, V(q))\, \Phi_i(q)\, \Phi_j(q)}{\int_{\overline{\Omega}} \exp(-\beta V(q))\, \Phi_i(q)\, \Phi_j(q)\, dq}\, dq. \tag{22}$$

The denominator of the left hand side is proportional to the desired integral, the right hand side is the integral of the δ-function for a distribution of data according to (21). The right hand side can be approximated via Monte Carlo quadrature.

$$\frac{1}{M(i,j)} \sum_{k=1}^{M(i,j)} \delta(\bar{q}_{ij}, q_{ij}^{(k)}) \approx \int_{\overline{\Omega}} \delta(\bar{q}_{ij}, q) \frac{\exp(-\beta\, V(q))\, \Phi_i(q)\, \Phi_j(q)}{\int_{\overline{\Omega}} \exp(-\beta V(q))\, \Phi_i(q)\, \Phi_j(q)\, dq}\, dq. \tag{23}$$

On the left side of this expression, we may approximate δ up to an unknown normalization constant with a characteristic funtion of a small ν-environment of $\bar{q}_{ij}$. With this approximation, (23), and (22) we can compute the integrals via

$$\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}} \propto I(i,j) = \frac{\exp(-\beta\, V(\bar{q}_{ij}))\, \Phi_i(\bar{q}_{ij})\, \Phi_j(\bar{q}_{ij})\, M(i,j)}{U_\nu(\bar{q}_{ij})}, \tag{24}$$

where $U_\nu(\bar{q}_{ij})$ is the number of sampling points inside a small ν-environment of $\bar{q}_{ij}$. In order to get the correct scaling for the integral $\langle \Phi_i, \Phi_j \rangle_{\overline{\pi}}$, the matrix $I = (I(i,j))_{i,j}$ in equation (24) is normalized, such that the sum of its elements is 1. Note, that (24) converges against the correct integral if $\nu \to 0$ and $M(i,j) \to \infty$. In practise we need more and more sampling points if ν decreases in order to achieve a statistically relevant ratio $M(i,j)/U_\nu(\bar{q}_{ij})$.

Time scale factor. If we are not interested in the correct scaling of Q, we do not need any dynamics simulations at this point. Note that it is not necessary to compute μ for the identification of invariant parts of $\overline{\Omega}$, because the eigenvector data X does not depend on μ. In order to compute the correct time scale factor μ, if it is needed, we have to include dynamics simulations, but only for a very restricted part of $\overline{\Omega}$ given by a basis function Φ'_k. Assume an initial molecular state in set Φ'_k. The time T_k that is spent before the state leaves Φ'_k via heat bath dynamics is called *exit time* [5]. The probability $P[T_k > t]$ that T_k is greater than a certain time $t > 0$ is given by

$$P[T_k > t] = \exp(-Q(k,k)\, t). \tag{25}$$

From this information, we can compute the unknown scalar factor via heat bath simulation starting in an arbitrary initial set Φ'_k of states. The time $t_{1/2}$, for which the probability is $P[T_k > t_{1/2}] = 0.5$, is called *half-value time*. Note that Q is an ensemble-based rate matrix, i.e. we start a large number $L \gg 0$ of simulations with initial-state distribution of $\overline{\pi}$ restricted to Φ'_k. We stop each single simulation when it leaves Φ'_k. In this case, $t_{1/2}$ is the time when the $L/2$-th simulation stops. Using (25) and (18), μ can be approximated as

$$\mu = \frac{\ln(2)}{t_{1/2}} \cdot \frac{\langle \Phi_k \rangle_{\overline{\pi}}}{\langle \Phi_k \rangle_{\overline{\pi}} - \langle \Phi_k, \Phi_k \rangle_{\overline{\pi}}}. \tag{26}$$

Correct reduced propagator via TST. After computation of the rate matrix Q for a given set (20) of basis functions Φ via (18) and Monte Carlo quadrature (24), the eigenvalue problem $QX = \mu X\Theta$ is solved for the n dominant eigenvalues ($\theta_i \approx 0$) of Q. For the corresponding eigenvectors X, we apply PCCA+ in order to get the transformation matrix $\mathcal{A}$. The correct reduced propagator has the form (12) with $\Lambda = \exp(\tau\mu\Theta)$. The corresponding domain decomposition $\overline{\chi}$ is given by $\overline{\chi} = \Phi X \mathcal{A}$, see (10) and (11). If it is important to know the correct scaling factor μ, one can apply (26) after local simulation of the half-value time $t_{1/2}$ of one given set Φ'_k.

Parameter selection. For the choice of the nodes $q_1, \ldots, q_n$ of the basis functions and for the choice of $\alpha > 0$, note that the space discretization should be fine enough, so that ξ approximates eigenfunctions of $\mathcal{P}^\tau$ well. $\alpha > 0$ has to be large enough, so that $\Phi_i \approx \Phi'_i$, but also small enough to sample (21) via Metropolis Monte Carlo correctly.

Computational cost. Equation 24 means that N^2 samplings have to be performed if N basis functions are selected. But for non-neighboring sets Φ'_i and Φ'_j and $\alpha \gg 0$, the nominator of (24) is negligibly small. The elements of I need only be computed for neighboring sets in $\overline{\Omega}$.

3 Illustrating Examples

TST is based on two important assumptions: First, Φ' is assumed to be a sufficient set of basis functions for the approximation of eigenfunctions of $\mathcal{P}$ resp. $\mathcal{Q}$. Second, instead of sets Φ', a relaxation Φ is used for the computation of overlap integrals in (14). Within these limits, TST is correct. In Section 3.1, it is shown that standard TST is an extreme simplification of conformation-based TST, including the simplifications of Section 2.3. In Section 3.2 and Section 3.3, examples are given which cannot be treated with standard TST, but these examples are accessible to the conformation-based theory. In all of the examples, we compute the infinitesmal generator Q_c instead of the propagator $P_c^\top$, but via $P_c^\top = \exp(\tau\, Q_c^\top)$, the propagator can be derived from the infinitesimal generator.

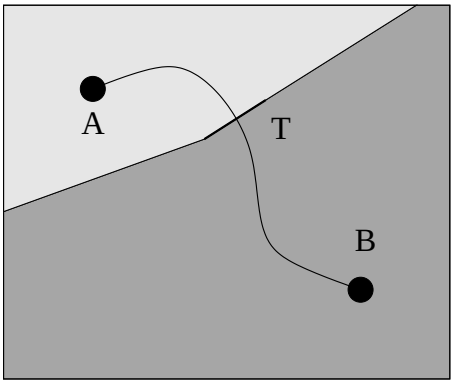

Figure 3: State space $\overline{\Omega}$ with two metastable regions A and B and a corresponding transition path. This path crosses the boundary between A and B at a transition state T.

3.1 Standard TST as a simplification of the conformation-based approach

Setting. A very simple example for TST is illustrated in Fig. 3. In this example, the state space $\overline{\Omega}$ consists of two metastable regions A and B, similar to the situation of the time series in Fig. 1. Once the system is in a state within one of these regions, it stays inside that region with a high probability. Transitions between regions are rare events. The two regions of $\overline{\Omega}$ are given in terms of membership functions Φ'_A and Φ'_B. In Fig. 3, it is also assumed that transitions only take place by crossing a certain transition state T. The transition state is a saddle point of the potential energy surface V: Along the indicated reaction path, T has the *maximal* potential energy. Inside the boundary between A and B, the transition state T has the *minimal* potential engergy value.

Approximation of Q_c. Metastability of A and B can be expressed by the following approximations:

$$\mathcal{P}^\tau \Phi'_A \approx \Phi'_A, \qquad \mathcal{P}^\tau \Phi'_B \approx \Phi'_B.$$

Thus, Φ'_A and Φ'_B are approximations of two dominant eigenfunctions of $\mathcal{P}^\tau$ and also of $\mathcal{Q}$. In this case, where $N = n = 2$, the correct reduced operator $Q_c = Q$ is given by Eq. 18 using these two basis functions:

$$Q_c = \mu \begin{pmatrix} -\frac{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}}{\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}} + \langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}} & \frac{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}}{\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}} + \langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}} \\ \\ \frac{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}}{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}} + \langle \Phi_B, \Phi_B \rangle_{\overline{\pi}}} & -\frac{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}}{\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}} + \langle \Phi_B, \Phi_B \rangle_{\overline{\pi}}} \end{pmatrix}. \tag{27}$$

Now, we will only focus on the computation of $Q_c(1,2)$. The other elements can be computed analogously. For the approximation of $\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}}$ and $\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}$, we use Eq. 23:

- For $\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}}$, it is assumed that there is one deep local minumum $q_A \in \overline{\Omega}$ of V inside region A of the state space, see black circle in Fig. 3. Further,

assume $\Phi_A(q_A) = 1$, and nearly all sampling points[3] are located in a small environment of q_A. Then $\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}} \propto \exp(-\beta\, V(q_A))$.

- For $\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}$, it is assumed that there is one transition state $q_T \in \overline{\Omega}$ of V inside the boundary AB, see Fig. 3. Furthermore, $\Phi_A(q_T) = \Phi_B(q_T) = 0.5$, and nearly all transitions cross a small environment of q_T. Then $\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}} \propto \frac{1}{4} \exp(-\beta\, V(q_T))$.

- The summand $\langle \Phi_A, \Phi_B \rangle_{\overline{\pi}}$ in the denominator of (27) can be neglected, because $\frac{1}{4} \exp(-\beta\, V(q_T)) \ll \exp(-\beta\, V(q_A))$.

With these assumptions and simplifications, the transition rate $A \to B$ is proportional to the Boltzmann expression $\exp(-\beta\,(V(q_T) - V(q_A)))$, where the difference $V(q_T) - V(q_A)$ is the potential energy barrier between A and B. This is equivalent to the result of standard TST. The above simplifications can also be applied for systems with more than 2 conformations.

Limits of standard TST. In order to derive standard TST from conformation-based TST, we introduced some assumptions and simplifications which limit the accuracy of standard TST. It is easy to construct counter-examples for standard TST: There might be more than one transition state on the boundary between A and B. A and B might be metastable subsets of $\overline{\Omega}$ but consist of many local minima and plateaus. In this case, entropical effects play an important role in the evaluation of $\langle \Phi_A, \Phi_A \rangle_{\overline{\pi}}$ and standard TST does not lead to correct results.

3.2 Entropical Transitions

The next example can not be treated with standard TST. In Fig. 4, the existence of barriers lead to entropical effects. The transition states are not saddle-points of the potential energy surface beacuse the potential energy function V is constant. In this artificial example, we investigate the dynamics of a point in a 2-dimensional box $\overline{\Omega}$. This box has two chambers A and B, which are connected by a small passage. Transitions between A and B only take place through this passage. The state space $\overline{\Omega}$ is discretized into 9×5 smaller boxes given by their characteristic functions $\{\Phi'_1, \ldots, \Phi'_{45}\}$. Following the ideas of Section 2.2, the elements of K in (13) can easily be determined. All common edges of the discretization boxes (exept for the wall between A and B) have the same statistical weight, because in $\overline{\Omega}$, the stationary distribution is uniform. The edges of the discretization boxes located on the wall between A and B have no weight. Since K is a stochastic matrix, the entries of K are uniquely determined by these considerations. For the computation of the rate factors R in (13), we apply Eq. 15. The stationary distribution is uniform. All discretization boxes Φ'_i have the same weight $\pi_i = 1/45$. Solving the eigenvalue problem (16) with these constant weights yields R up to an unkown scaling factor μ. After computation of Q, we are now interested in a 2-dimensional reduction of Q, i.e. $n = 2$. First, we will investigate the set-based reduction $\hat{Q}_c$. Then we will compute the subspace-based reduction Q_c.

[3] Distributed according to (21).

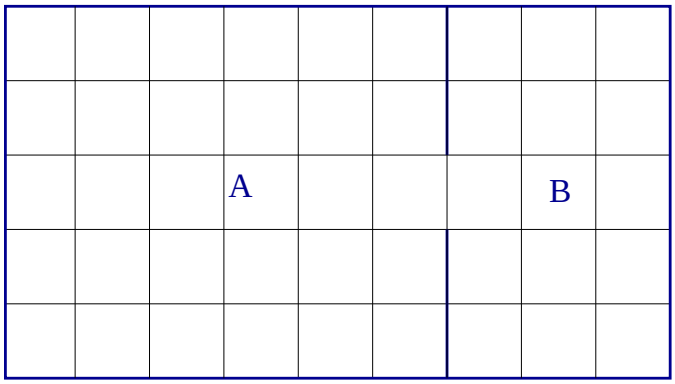

Figure 4: The state space $\overline{\Omega}$ is a box with two "chambers" A and B (thick lines). These chambers are connected by a small passage. For the computations, the space $\overline{\Omega}$ is discretized into 45 small congruent boxes (thin lines). Chamber A has twice the volume of chamber B.

Set-based reduction. Analogously to P in (3) and (5), the dimension of Q can be reduced to $n = 2$. For characteristic basis functions $\hat{\overline{\chi}}_A, \hat{\overline{\chi}}_B : \overline{\Omega} \to \{0, 1\}$ representing the two chambers of $\overline{\Omega}$, the Galerkin based approach (5) and the "intuitive" approach (3) are equivalent. Up to an unknown scaling factor μ, the set-based transition rate matrix $\hat{Q}_c$ is

$$\hat{Q}_c = \begin{pmatrix} -0.5 & 0.5 \\ 1.0 & -1.0 \end{pmatrix}. \tag{28}$$

Due to the set-based dimension reduction, (28) is an ensemble-based transition rate matrix. The result (28) can also be justified theoretically. In a detailed balanced equilibrium, always the same fraction of states crosses the passage from A to B as the fraction of states going from B to A. Divided by the volumes of A and B, the reaction rate $A \to B$ is half the reaction rate $B \to A$. However, $\hat{Q}_c$ cannot be used for a time series realization (cf. Section 1.2). For a single state $q \in \overline{\Omega}$, transition rates from A to B are not independ of the position of q inside chamber A.

Subspace-based reduction. In the subspace-based approach, we have membership functions $\overline{\chi}_A, \overline{\chi}_B : \overline{\Omega} \to [0, 1]$ instead of sets. The membership functions can be computed by PCCA+, see Fig. 5. Using the subspace-based approach, the position of $q \in \overline{\Omega}$ inside A determines the degree of membership $\overline{\chi}_A(q) \in [0, 1]$ with respect to A. In this case, the above theory has shown that there exists a transition rate matrix Q_c which is valid for a time series realization. The Galerkin discretization of Q with the membership functions of Fig. 5 is

$$Q_c = \begin{pmatrix} -0.2800 & 0.2800 \\ 0.4481 & -0.4481 \end{pmatrix}. \tag{29}$$

Again, this result has a theoretical justification. The ratio of $Q_c(1, 2)$ and $Q_c(2, 1)$ is equal to the ratio $\pi_c(B)/\pi_c(A)$ of the weights $\pi_c(A) = \langle \overline{\chi}_A \rangle_{\overline{\pi}}$ and $\pi_c(B) = \langle \overline{\chi}_B \rangle_{\overline{\pi}}$. In a two-dimensional example (this is also true for the example

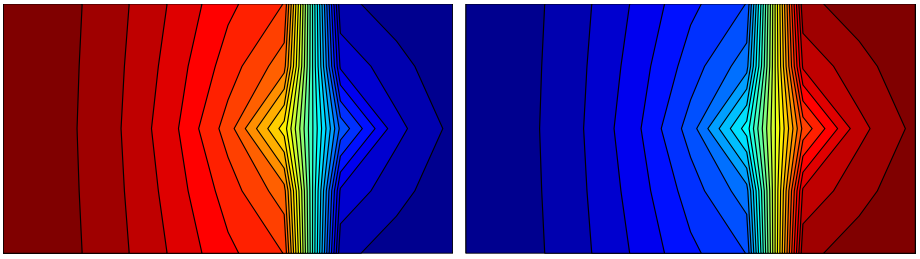

Figure 5: The two membership functions $\overline{\chi}_A$ (left) and $\overline{\chi}_B$ (right) approximated via PCCA+ using the discretization of Fig. 4. High membership values are colored red, low membership values are colored blue. The weights of the two conformations are $\pi_c(A) = 0.6154$ and $\pi_c(B) = 0.3846$.

in Fig. 3), the ratio of $Q_c(1,2)$ and $Q_c(2,1)$ is always determined by the weights of the conformations. Moreover, since $Q_c(1,1) = -Q_c(1,2)$ and $Q_c(2,2) = -Q_c(2,1)$ in the two-dimensional case, Q_c is completely determined up to a scaling factor by the weights of the conformations.

3.3 Three conformations

Let us now look at a 2-dimensional potential energy function $V : \mathbb{R}^2 \to \mathbb{R}$ with three local minima:

$$\begin{aligned} V(x,y) \quad = \quad & 3\,\exp\big(-x^2-(y-\frac{1}{3})^2\big) - 3\,\exp\big(-x^2-(y-\frac{5}{3})^2\big) \\ & -5\,\exp(-(x-1)^2-y^2) - 5\,\exp(-(x+1)^2-y^2) \\ & +0.2\,x^4 + 0.2\,(y-\frac{1}{3})^4. \end{aligned} \tag{30}$$

This potential energy function has also been investigated in [8]. The x-symmetric function (30) is shown in Fig. 6. We want to investigate the transitions between the two deeper minima of V. Transitions can either directly cross the barrier between these two minima, or they visit the plateau at the top in Fig. 6.

Choice of parameters. In order to discretize $\overline{\Omega}$ sufficiently, a regular grid of 21×16 nodes is used for the Voronoi tessellation and its relaxation (20) with $\alpha = 10$. Two different temperatures are investigated with $\beta^{(1)} = 1.67$ ("high temperature") and $\beta^{(2)} = 3.34$ ("low temperature"). The Q-matrix is computed for these two temperatures via (24) and (18) using the approach $M(i,j) = U_\nu(\overline{q}_{ij})$ of Section 3.1.

Membership functions. After computation of the Q-matrix, the eigenvalue problem is solved, and PCCA+ is applied in order to extract membership functions $\overline{\chi}$ of the three conformations. The membership functions are different for the two temperatures. In the high-temperature case, the membership functions are smoother than the membership functions of the low-temperature case. The high-temperature membership functions are shown in Fig. 7. On the ba-

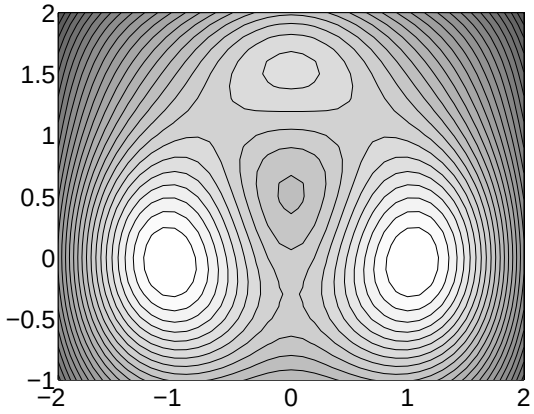

Figure 6: Potential energy function with three local minima, two deeper ones (left and right) and a connecting plateau (top).

sis of the membership functions $\overline{\chi}$ of the conformations, the reduced transition rate matrices $Q_c^{(1)}$ and $Q_c^{(2)}$ can be computed for the two temperatures. The first conformation is the connecting plateau, the second conformation is the left minimum, and the third conformation is the right minimum in Fig. 6.

High temperature. The reduced stationary density for the case $\beta^{(1)} = 1.67$ is

$$\pi_c^{(1)} = \begin{pmatrix} 0.053 \\ 0.478 \\ 0.469 \end{pmatrix}. \tag{31}$$

The population of the first conformation (connecting plateau) is very small compared to the population of the other conformations. The transition rates from the connecting plateau to any of the deeper minima are about ten times higher than the transition rates of the reverse "reactions":

$$Q_c^{(1)} = \begin{pmatrix} -0.008799 & 0.004390 & 0.004409 \\ 0.000491 & -0.000750 & 0.000259 \\ 0.000490 & 0.000271 & -0.000761 \end{pmatrix}. \tag{32}$$

The entries of $Q_c^{(1)}$ reflect the symmetry of the potential energy function. $Q_c^{(1)}$ can be used for a time series simulation. The starting vector for our simulation is $v^{(0)} = (0, 1, 0)^\top$ (the left minimum). In Fig. 8, the results of this simulation are shown. The population of conformation 3 goes up, whereas the population of conformation 2 is reduced. One can also see that there is some occupancy in conformation 1 long before the equilibrium state is reached. This means that a non-negligible fraction of transitions crosses the connecting plateau. Because of the unknown scaling factor μ, the x-axis in Fig. 8 has no unit.

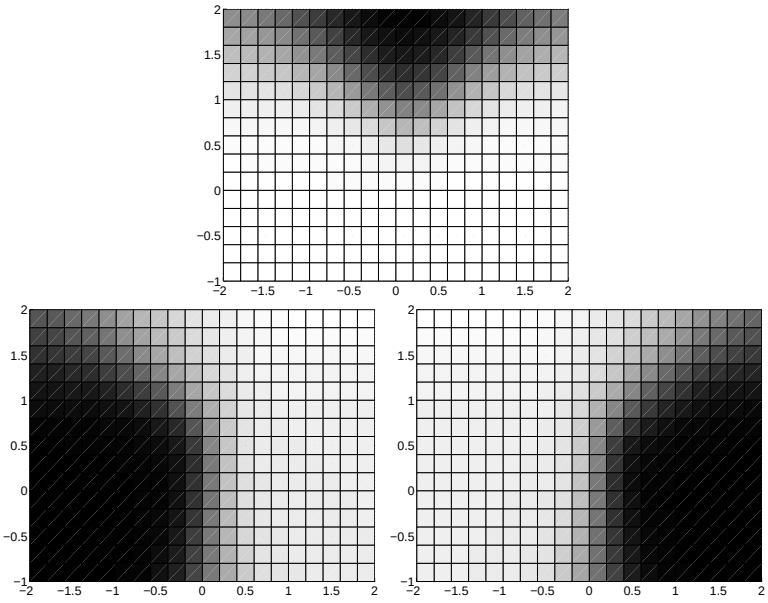

Figure 7: The membership functions $\overline{\chi}$ of the three conformations. Exemplified for $\beta = 1.67$. The conformations are more "crisp" for higher β-values.

Low temperature. In the low temperature case $\beta^{(2)} = 3.34$, the reduced stationary density,

$$\pi_c^{(2)} = \begin{pmatrix} 0.002 \\ 0.490 \\ 0.508 \end{pmatrix}, \tag{33}$$

has a lower entry for the first conformation than (31). Furthermore, the transition rates from the connecting plateau to any of the deeper minima are about 200 times higher than the reverse transition rates:

$$Q_c^{(2)} = \begin{pmatrix} -0.006903 & 0.003402 & 0.003500 \\ 0.000016 & -0.000025 & 0.000009 \\ 0.000016 & 0.000009 & -0.000025 \end{pmatrix}. \tag{34}$$

Since $Q_c^{(1)}$ and $Q_c^{(2)}$ are only known up to a scaling factor μ, which may be different for the two temperatures, the entries of $Q_c^{(1)}$ and $Q_c^{(2)}$ can not be compared directly. It is only ratios of matrix elements that are meaningful. $Q_c^{(2)}$ can be used for a kinetics simulation of the transition between the two deeper minima, see Fig. 9. In the low-temperature case, the connecting plateau also reaches its equilibrium density very fast. In Fig. 9, the population curve of conformation 1 can not be distinguished from the x-axis (the popolutaion is almost zero).

Time series realization. Via $P_c = \exp(\tau\, Q_c)$, a transition matrix can be obtained from the infinitesimal generator. For the high-temperature case and

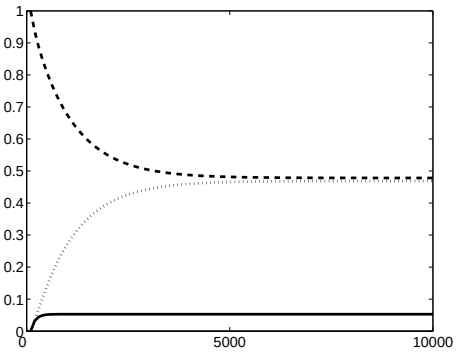

Figure 8: Time-population-plot of a kinetics simulation of the transition between the two deeper minima (dashed and dotted curve). The connecting plateau is populated very early (solid curve).

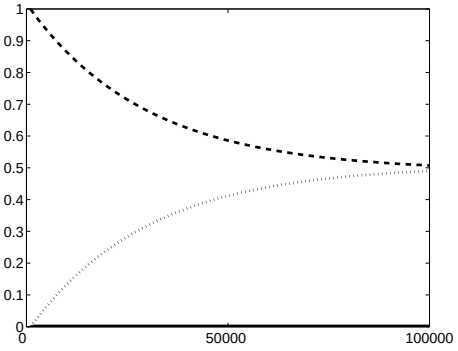

Figure 9: Time-population-plot of a kinetics simulation of the transition between the two deeper minima (dashed and dotted line). The connecting plateau is almost unoccupied (solid curve is almost constant zero).

$\tau = 100$, the transition matrix is

$$P_c^{(1)} = \begin{pmatrix} 0.4269 & 0.2861 & 0.2869 \\ 0.0320 & 0.9360 & 0.0320 \\ 0.0319 & 0.0331 & 0.9350 \end{pmatrix}. \tag{35}$$

For the low-temperature case with $\tau = 1000$, we get

$$P_c^{(2)} = \begin{pmatrix} 0.0033 & 0.4912 & 0.5055 \\ 0.0023 & 0.9818 & 0.0158 \\ 0.0023 & 0.0153 & 0.9824 \end{pmatrix}. \tag{36}$$

The matrices (35) and (36) can be used for a time series realization because they are the correct reduced propagators. In fact, Fig. 8 and Fig. 9 have been created by a plot of one hundred steps of these propagators.

Defuzzification. We have compared $Q_c^{(1)}$ and $Q_c^{(2)}$ in terms of their transition rates between subsets of $\overline{\Omega}$. However, in the above examples, we do not have subsets but membership functions. Membership functions represent *fuzzy sets*. One can *defuzzify* these membership functions by crisp functions $\hat{\chi} \in \{0,1\}^{N\times n}$, if we set $\hat{\overline{\chi}}_i(q) = 1$, by setting $\overline{\chi}_j(q) \leq \overline{\chi}_i(q)$ for all $j = 1, \ldots, n$. (32) and (34) may be good approximations of the correct reduced infinitesimal generator, but if we want to speak about transitions within $\overline{\Omega}$, we should compute Q_c on the basis of subsets as well.

Set-based discretization in the high-temperature case. The membership functions in the high-temperature case in (32) are not crisp, i.e. there is a non-negligible part of $\overline{\Omega}$, where $\overline{\chi}$ attains values far away from 0 or 1. Thus, the reduced stationary density for the defuzzified membership functions,

$$\hat{\pi}_c^{(1)} = \begin{pmatrix} 0.040 \\ 0.485 \\ 0.475 \end{pmatrix}, \tag{37}$$

is different from $\pi_c^{(1)}$ in (31). And also the reduced propagator in this case,

$$\hat{Q}_c^{(1)} = \begin{pmatrix} -0.047117 & 0.023711 & 0.023406 \\ 0.001951 & -0.003660 & 0.001709 \\ 0.001964 & 0.001744 & -0.003708 \end{pmatrix}, \tag{38}$$

is not identical to $Q_c^{(1)}$ in (32). Note that (32) is valid for a time series simulation, whereas (38) is only an ensemble-based transition rate matrix. It is remarkable that the rates in (38) are higher than the rates in (32), although $\hat{Q}_c^{(1)}$ is based on non-overlapping sets and $Q_c^{(1)}$ is a propagator between fuzzy sets. The reason is, that the membership functions $\overline{\chi}$ from PCCA+ are optimal in some sense [13]. PCCA+ aims at maximizing metastability, i.e. minimizing transitions between the fuzzy sets.

Set-based discretization in the low-temperature case. In the low-temperature case, the membership functions $\overline{\chi}$ from PCCA+ are already almost crisp. Therefore, the reduced stationary density $\hat{\pi}_c^{(2)}$ is almost identical to $\pi_c^{(2)}$ in (33):

$$\hat{\pi}_c^{(2)} = \begin{pmatrix} 0.002 \\ 0.490 \\ 0.508 \end{pmatrix}. \tag{39}$$

The transition rate matrix, however, differs from (34):

$$\hat{Q}_c^{(2)} = \begin{pmatrix} -0.026999 & 0.013816 & 0.013183 \\ 0.000057 & -0.000097 & 0.000035 \\ 0.000053 & 0.000034 & -0.000087 \end{pmatrix}. \tag{40}$$

This is due to the fact that the transition rates depend very sensitively on the membership values in the transition region between the conformations. Analogously to the high-temperature case, the transition rates in (40) are higher than in (34). The reason is that PCCA+ aims at maximizing metastability of

the conformations. The ratio of the transition rates from and to the connecting plateau is similar to the subspace-based result (ratios 1/10 and 1/200 for the two temperatures). The most remarkable difference between the set-based and the subspace-based discretization method is the following: In the low-temperature case in (40), the transitions from one of the minima directly to the other minimum have a much lower rate than the transitions to the connecting plateau. In contrast to that, in the high temperature case in (38), the transition rates are almost identical. This statement is true only for the set-based approach. In the subspace-based approach, the transition rates in (34) and (32) are always much higher for transitions into the connecting plateau as for the direct transitions between the minima. Q_c may be a good approximation for the correct infinitesimal generator according to the above theory, but (32) and (34) cannot be seen as transition rate matrices between *sets*. The knowledge of $\overline{\chi}$ is essential for the interpretation and application of the subspace-based Q_c.

Conclusion

It has been shown that the correct reduced transition matrix P_c is a Galerkin discretization of the full-dimenstional Markov operator $\mathcal{P}^\tau$, if and only if the set of discretizing basis functions span an invariant subspace of the operator $\mathcal{P}^\tau$. There is no chance to compute a correct reduced propagator P_c for equilibrated systems with a purely set-based approach. Only by using PCCA+, there is a possibility to compute the correct reduced propagator based on "fuzzy" sets, i.e. on membership functions. If it is possible to compute the integrals in Eq. 18 in an efficient way without time-consuming molecular dynamics simulations, conformation-based TST is a powerful tool for the investigation of conformational changes. Note that in section 3.3, only local maximization methods have been applied.

References

[1] A. Amadei, A. B. M. Linssen, B. L. De Groot, and H. J. C. Berendsen. Essential degrees of freedom of proteins. *Molecular Engineering*, 5(1-3):71–79, 1995.

[2] P. Deuflhard and M. Weber. Robust Perron Cluster Analysis in Conformation Dynamics. In M. Dellnitz, S. Kirkland, M. Neumann, and C. Schütte, editors, *Lin. Alg. App. – Special Issue on Matrices and Mathematical Biology*, volume 398C, pages 161–184. Elsevier, 2005.

[3] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[4] W. Huisinga. *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics*. Doctoral thesis, Freie Universität Berlin, 2001.

[5] M. Kijima. *Markov Processes for Stochastic Modeling*. Stochastic Modeling Series. Chapman and Hall, 1997.

[6] S. Kube and M. Weber. A coarse graining method for the identification of transition rates between molecular conformations. *J. Chem. Phys.*, 126:024103–024113, 2007.

[7] K. J. Laidler and M. C. King. The development of transition-state theory. *J. Phys. Chem.*, 87:2657–2664, 1983.

[8] Ph. Metzner, Ch. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, 125:084110, 2006.

[9] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124:044104, 2006.

[10] C. R. Johnson R. A. Horn. *Matrix Analysis*. Cambridge University Press, 1990. (chapter 8).

[11] Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation thesis, Department of Mathematics and Computer Science, Freie Universität Berlin, 1999.

[12] Ch. Schütte and W. Huisinga. Biomolecular Conformations can be Identified as Metastable Sets of Molecular Dynamics. In P. G. Ciarlet and C. Le Bris, editors, *Handbook of Numerical Analysis*, volume X. Special Volume Computational Chemistry, pages 699–744. North–Holland, 2003.

[13] 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.

[14] M. Weber and T. Galliat. Characterization of Transition States in Conformational Dynamics using Fuzzy Sets. ZIB-Report 02-12, Zuse Institute Berlin, 2002.

[15] M. Weber, S. Kube, L. Walter, and P. Deuflhard. Stable computation of propability densities for metastable dynamical systems. ZIB-Report 06-39, Zuse Institute Berlin, 2006. Accepted for publication in SIAM J. Multisc. Mod.