Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

CHRISTOF SCHÜTTE AND MARCO SARICH

Department of Mathematics and Computer Science, Freie Universität Berlin, Germany

Zuse Institute Berlin, Germany

A Critical Appraisal of Markov State Models

ZIB-Report 15–18 (March 2015)

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

A Critical Appraisal of Markov State Models

Christof Schütte and Marco Sarich

Institut für Mathematik, Freie Universität Berlin, Germany
and
Zuse Institute Berlin, Germany

Abstract

Markov State Modelling as a concept for a coarse grained description of the essential kinetics of a molecular system in equilibrium has gained a lot of attention recently. The last 10 years have seen an ever increasing publication activity on how to construct Markov State Models (MSMs) for very different molecular systems ranging from peptides to proteins, from RNA to DNA, and via molecular sensors to molecular aggregation. Simultaneously the accompanying theory behind MSM building and approximation quality has been developed well beyond the concepts and ideas used in practical applications. This article reviews the main theoretical results, provides links to crucial new developments, outlines the full power of MSM building today, and discusses the essential limitations still to overcome.

Keywords: Markov State Models, transfer operator, molecular dynamics, approximation quality, sparse tensor approximation

AMS classification: 60J20, 92C05

1 Introduction

Applications in modern biotechnology and molecular medicine require simulation of biomolecular systems in atomic representation with immense length and timescales that are *far* beyond the capacity of computer power currently available. The processes that constitute molecular function are *rare event* processes appearing on timescales that are many orders of magnitude, say 10-15 orders of magnitude, longer than the typical time steps of the numerical simulation. As a consequence, there is an increasing need for reduced models that reproduce the correct rare event statistics.

In most molecular systems the biologically interesting and computationally problematic rare events belong to so-called *conformation changes*. Conformations are *metastable sets* of the dynamical behavior of the molecule, that is, regions of the molecule's state space that are attractive for the dynamics in the sense that typical trajectories remain within such regions for long periods of time before exiting towards other metastable sets.

Markov State Modelling is about how to exploit the existence of metastable sets for constructing a reduced molecular dynamics model with good approximation properties on the long timescales. In the standard setting a Markov State Model (MSM)

is a Markov chain whose transition matrix is given by the transition probabilities $\mathbb{P}_\mu(X_\tau \in A_k | X_0 \in A_j)$ of the original molecular dynamics process (X_t) between some subsets $A_1, \ldots, A_m$ of the molecular state space that form the (macro-)states of the MSM. The timescale τ for which the transition probabilities are computed is called the lagtime and typically is much shorter than the timescales of transitions between the metastable sets of the process. Mathematically, the process of reducing the original molecular dynamics process to the MSM process is a discretization of the so-called *transfer operator* of the molecular dynamics process [1].

The main advantage of MSMs is that we know how to coarse grain them optimally. Based on the dominant eigenvectors of a fine-scale MSM one can find aggregated (macro-)states that correspond to the dominant metastable sets of the original molecular dynamics [2, 3]. It has been shown that for molecular systems exhibiting such metastable sets, the Markovian dynamics given by an MSM allows very close approximation of the longest relaxation processes of the underlying molecular system, at least under equilibrium conditions [4, 5, 6]. In fact, whenever we assume sufficient sampling, the error E between dynamic long-term behavior of the original MD process and the MSM process on timescales $t > \tau$ is bounded by

$$E(t) \leq C(\delta + \eta)^2 \exp(-\frac{t}{t_2}), \tag{1.1}$$

where δ can be made small by choosing the discretization fine enough, η is decreasing exponentially with growing lagtime τ, t_2 is the slowest timescale of the original MD process, and C is a constant that depends mildly on the timescale and the number of states of the MSM. We observe that, *in principle*, for long enough lagtime and appropriate discretization, the error on long timescales can be made arbitrarily small. For details of how to compute C, δ, and η given the discretization and the lagtime please see Sec. 4.2. It has been demonstrated that, in many cases of practical relevance, MSM building requires short lagtime, i.e., short MD trajectories only, much shorter than the timescales of interest, compare [7, 8, 9, 10, 11], for example. Thus, MSM building often allows the study of dynamic behavior on long timescales without requiring MD trajectories of comparable length. However, the problem of *how* to optimally choose the appropriate lagtime and discretization in general has *not* been solved in general, in particular for *very high dimensional systems*, and is still a topic of ongoing research [12].

MSM building in molecular dynamics started with a series of papers more than 15 years ago [13, 14, 1, 2]. Recent years have seen an ever increasing publication activity on how to construct MSMs for very different molecular systems ranging from peptides to proteins, from RNA to DNA, and via molecular sensors to molecular aggregation. Moreover, MSMs have been used to construct kinetic fingerprints from MD simulations which facilitates understanding of essential dynamics and comparison with experimental data [15]. Several recent books review these practical approaches in a lot of algorithmic detail, see [16, 12] for an overview. Also, several MSM software environments are available [17, 18, 19].

By far most of the literature on pratical applications uses standard MSM constructions where the MSM consists of transition probabilities between sets in state space as outlined above. With standard MSMs, despite the theoretical knowledge contained in the error estimate (1.1), a reliable practical estimation of the deviation between the original MD and the MSM process on long timescales is only possible for molecular systems of moderate size [12]. However, more general MSM schemes have been developed that exhibit improved approximation quality especially in high

dimensions [20, 21, 22, 23, 24, 6]. That is, the full power of the idea behind MSMs has *not* been utilized yet. Therefore, the present article concentrates on demonstrating the full breadth of possibilities of MSM building and the resulting opportunities and limitations. We will *not* go into details of how to realize MSMs for specific molecular systems but will instead review how the idea of standard MSMs has been generalized in recent years, why this leads to improved approximation quality and how the error of MSMs in comparison to the original MD process can be controlled including the error resulting from incomplete sampling.

2 The Transfer Operator

In order to allow for the discussion of MSM building beyond standard MSMs we have to introduce the transfer operator of the original MD process. To this end we assume that the MD process is a Markov process $(X_t)_{t\in T}$ in state space $\mathbb{X}$, discrete in time ($T = \mathbb{N}$; as resulting from MD simulations) or time-continuous ($T = \mathbb{R}$), with non-negative transition kernel $p(t, x, y)$ with $\int_X p(t, x, y)dy = 1$ for all $x \in S$. The kernel $p(t, x, y)$ tells us the probability of the MD process to go from x to y in time t, or, more precisely, the density associated with it. We assume that the process has a positive invariant (or stationary) density $\mu(x)$, typically of the form $\mu(x) \propto \exp(-\beta H(x))$ for some energy function H, such that

$$\mu(y) = \int_{\mathbb{X}} p(t, x, y)\mu(x)dx.$$

This setting is rather general and includes most (stochastic as well as thermostatted) cases of MD; an extension also covers pure micro-canonical MD, see [6] for details. We introduce the Hilbert space $L^2_\mu = \{u : \mathbb{X} \to \mathbb{R};\ \int_{\mathbb{X}} u(x)^2\mu(x)dx < \infty\}$ with scalar product

$$\langle u, v\rangle_\mu = \int_{\mathbb{X}} u(x)v(x)\mu(x)dx.$$

The transfer operator is defined as [6]

$$T_t u(y)\mu(y) = \int_{\mathbb{X}} p(t, x, y)u(x)\mu(x)dx.$$

It transports function in state space according to the underlying dynamics and relative to the invariant measure, and transports probability densities into probability densities. T_t plays the role of the propagator in quantum mechanics (QM). However, opposed to QM propagators, it is a bounded operator on the Hilbert space L^2_μ; its eigenvalues all are smaller or equal to 1 in modulus. $\lambda = 1$ always is an eigenvalue with eigenvector $\mathbb{1}$ (the function being constant 1 on all of $\mathbb{X}$). Under additional assumptions on the ergodicity of the Markov process (geometric ergodicity), $\lambda = 1$ is the largest eigenvalue and the rest of the spectrum $\sigma(T_t)$ satisfies $\lambda \in \sigma(T_t) \Rightarrow |\lambda| < 1$. The adjoint operator of T_t wrt the scalar product $\langle\cdot,\cdot\rangle_\mu$ is given by [6]

$$T_t^* u(x) = \mathbb{E}_x\Big(u(X_t)\Big).$$

If the Markov process is reversible the detailed balance condition

$$\mu(x)p(t, x, y) = \mu(y)p(t, y, x)$$

is satisfied and the transfer operator is self-adjoint in L^2_μ [6] such that

$$T_t = T_t^*.$$

This is the situation we will consider in the following: reversible dynamics and self-adjoint transfer operator. However, it should be emphasized that the assumption of reversibility is made for the sake of simplicity of explanations and is not required for doing Markov State Modelling. Generalization of the results presented here to non-reversible Markov processes can be found in [6, 25] (theory) and [26] (practical construction).

Transition probabilities. Let us denote with $\mathbb{P}_\mu$ the probability measure given by the stationary density μ, i.e., $\mathbb{P}_\mu(X_0 \in A) = \mu(A) = \int_A \mu(x)dx$. Then all transition probabilities of the MD process (X_t) started from the equilibrium distribution can be computed by means of the transfer operator according to

$$\begin{aligned} \mathbb{P}_\mu(X_t \in B | X_0 \in A) &= \frac{1}{\mu(A)} \int_B \int_A p(t,x,y)\mu(x)dx\,dy \\ &= \frac{1}{\mu(A)} \int p(t,x,y)\mathbb{1}_A(x)\mathbb{1}_B(y)\mu(x)dx\,dy \\ &= \frac{\langle T_t \mathbb{1}_A, \mathbb{1}_B\rangle_\mu}{\langle \mathbb{1}_A, \mathbb{1}_A\rangle_\mu}, \end{aligned} \tag{2.1}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A, i.e., $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise. This formula shows that any information on the long-term of $T_t u$ for $u = 1_A$ (or more general functions u) will allow us to understand the long-term transition behavior of the underlying MD process.

Analogies: QM and Rouse model. In Quantum Mechanics (QM) the propagator P_t describes the evolution of wavefunctions. It is given by $P_t = \exp(-itH/\hbar)$ where H denotes the Hamiltonian of the quantum system. Spectral decomposition into the eigenenergies (eigenfunctions of H) allow to describe the dynamics in terms of eigenenergies and associated eigenphases. Often only the lowest eigenenergies are populated and the dynamics is essentially given by the lowest eigenvalues and eigenvectors of H or P_t, respectively. In close analogy this is what we are going to do with the transfer operator T_t: we will describe the long-term dynamics by the dominant eigenvalues and eigenvectors of T_t. However, the QM propagator is different from the transfer operator since its eigenvalues all have modulus 1 (that is, there is no dissipation). In contrast the transfer operator in MD just has one eigenvalue of modulus 1; all other have modulus strictly smaller than one (thus, there is dissipation). In this respect the Rouse model of polymer physics can be regarded as a closer analogy to MD transfer operators. For the Rouse model the largest eigenvalues dominate the long-term dynamics [27] in the same way as we will see in the following for the MD transfer operator.

Long-term dynamics and dominant timescales t_i. Let us now consider the transfer operator T_τ associated with a certain lagtime τ. Since T_τ is assumed to be self-adjoint it has only real-valued eigenvalues and -vectors. We assume that it only has isolated, positive eigenvalues, ordered according to $1 = \lambda_1 > \lambda_2 \geq ... \geq \lambda_m \geq ...$

(infinitely many ones with possible repetitions due to multiplicity). We associate a timescales with each eigenvalue by setting

$$t_i = \frac{\tau}{|\log \lambda_i|}, \tag{2.2}$$

such that we get a monotonically decreasing sequence of characteristic timescales $t_1 = \infty > t_2 \geq t_3 \geq \ldots$.

Let $u_0, u_1, ...$ be the corresponding normalized eigenvectors. Since T_τ is self-adjoint, the eigenvectors are orthogonal and

$$T_\tau u = \sum_{j=1}^{\infty} \lambda_j \langle u_j, u \rangle_\mu u_j = \sum_{j=1}^{\infty} \langle u_j, u \rangle_\mu u_j \, \exp\Big(-\frac{\tau}{t_j}\Big).$$

The long-time transport properties of the Markov process, we call it "kinetics" in the following, are given by

$$T_{k\tau} u = (T_\tau)^k u = \sum_{j=1}^{\infty} \langle u_j, u \rangle_\mu u_j \, \exp\Big(-\frac{k\tau}{t_j}\Big), \tag{2.3}$$

which can also be written as

$$T_{k\tau} u = (T_\tau)^k u = \sum_{j=1}^{m} \langle u_j, u \rangle_\mu u_j \exp\Big(-\frac{k\tau}{t_j}\Big) + R^k u,$$

with $\|R^k\|_\mu^2 < \exp(-k\tau/t_m) \ll 1$ for large k. Thus, the dominant eigenvalues/timescales govern the kinetics induced by the Markov process. For the more general case that a part of the spectrum of T_τ may be continuous (the "unbounded" states in QM) or some eigenvalues may be negative, visit [6].

Because of this insight, we are interested in the dominant eigenvalues and eigenvectors of T_t, and thus in the eigenvalue problem $T_t u = \lambda u$ in L^2_μ, or, respectively, in the variational formulation of the eigenvalue problem

$$u \in L^2_\mu, \ \lambda \in \mathbb{R}: \quad \langle T_\tau u, v \rangle_\mu = \lambda \langle u, v \rangle_\mu, \quad \forall v \in L^2_\mu. \tag{2.4}$$

In general, Markov State Models are appropriate discretizations of the eigenvalue problem that allow to approximate the dominant eigenvalues and eigenvectors of T_t well, and thus encode the kinetics of the underlying Markov process.

3 Galerkin Discretization of Transfer Operators

In order to computationally treat the eigenvalue problem, we have to define a finite dimensional ansatz space spanned by N linearly independent basis functions $\phi_1, \ldots, \phi_N$

$$S = \text{span}\{\phi_1, \ldots, \phi_N\} = \{u \in L^2_\mu | \quad u = \sum_{j=1}^{N} a_j \phi_j, \quad a_j \in \mathbb{R}\}.$$

We will discuss two approaches that share two essential characteristics. First, they make use of the ansatz space S to derive a finite linear equation that approximates

the original eigenvalue problem. Second, the entries of the matrices representing the linear equation can be estimated from observations of the underlying Markov process.

The Galerkin approach is based on the restriction of the variational eigenvalue problem to the subspace S:

$$u_S \in S,\ \lambda_S \in \mathbb{R}: \quad \langle T_\tau u_S, v\rangle_\mu = \lambda_S \langle u_S, v\rangle_\mu, \quad \forall v \in S \tag{3.1}$$

3.1 Standard MSMs

The standard form of MSMs, as utilized by most articles on the topic [12, 16, 28], is derived from the Galerkin ansatz as follows [13]: Let $\{A_j\}_{j=1,\dots,N}$ be a complete partition of the state space $\mathbb{X}$ into non-overlapping sets, i.e.,

$$\cup_{j=1}^N A_j = \mathbb{X}, \qquad \mu(A_j \cap A_k) = 0, \quad j \neq k,$$

and choose the ansatz functions as the indicator functions of the sets A_j,

$$\phi_j = \mathbb{1}_{A_j}.$$

When inserting this into (3.1) with $u_S = \sum_{j=1}^N c_j \mathbb{1}_{A_j}$ and varying v over the basis $\mathbb{1}_{A_k}$, $k = 1, \dots, k$ of S we get a system of N equations

$$\sum_{j=1}^N \langle T_\tau \mathbb{1}_{A_j}, \mathbb{1}_{A_k}\rangle_\mu \, c_j = \lambda_S \langle \mathbb{1}_{A_k}, \mathbb{1}_{A_k}\rangle_\mu \, c_k, \qquad k = 1, \dots, N$$

for the coefficients c_k of the eigenvector u_S. By using the form of the transition probabilities (2.1) we immediately see that these equations are equivalent to the finite-dimensional eigenvalue problem

$$\hat{T} c = \lambda_S c \tag{3.2}$$

with the coefficient vector $c = (c_k)_{k=1,\dots,N}$ and the transition matrix

$$\hat{T}_{jk} = \mathbb{P}_\mu\Big(X_t \in A_k | X_0 \in A_j\Big). \tag{3.3}$$

Thus, discretization of the transfer operator eigenproblem using a complete partition of state space results in a eigenproblem of the transition matrix of the MD process. This is the standard setting used in by far most publications regarding pratical use of MSMs, see [12]. However, the general Galerkin discretization approach (3.1) is far more general and allows MSM building schemes with much improved approximation properties.

Perfect approximation of long-term dynamics. In theory we can now ask what happens if we use an arbitrarily fine discretization. The answer is given in [13], Cor. 5.4: If the sets A_k are made arbitrarily fine, i.e., if $\max_k \mu(A_k) \to 0$ and $N \to \infty$, and all dominant eigenvalues are isolated, then the dominant eigenvalues and eigenvectors of the discretized problem (3.2) converge to the ones of the full transfer operator T_τ. This means that the long-term dynamics of the MD process is approximated *perfectly*, at least regarding the propagation of functions (2.3), if the MSM is made finer and finer.

On the first glance this theoretical result seems counter-intuitive: In general, i.e., without assuming that the lagtime τ is very long, the MD process (X_t) will *not* be

Markovian if restricted to some arbitrarily fine discretization sets A_k. Despite this, the Markov model $\hat{T}$ allows for perfect approximation of the long-term dynamics. The resolution of the apparent contradiction lies in the observation that the dynamical behavior as of (2.3) does *not* concern single MD trajectories but the resulting kinetics in an ergodic sense and that the mathematical convergence result just states that on this particular level the memory of the MD process regarding the discretization sets is without importance.

The curse of dimension. It is quite obvious that if we consider systems with growing dimension d, the size N of typical full partitions will explode with d. There are essentially two scenarios in which this fundamental problem can be circumvented, (1) if there is a subspace $\mathbb{X}_r$ of reduced dimension $r \ll d$ spanned by r reaction coordinates of the system to which the box discretization can be restricted (i.e., the sets A_i have the form of cylinders that are based on a full partition of $\mathbb{X}_r$), or (2) if a geometry-based clustering of sampling data allows to identify reliable partition sets, cf. [12]. Both cases do *not* offer general solutions to the problem but depend on the system at hand and expert intervention. Because of this limitation of standard MSM building other ansatz spaces have been considered. We will discuss two alternatives to standard MSMs (core-set and meshless MSMs). To this end we first have to discuss the general form of discretization matrices resulting from the Galerkin ansatz.

3.2 Galerkin discretization: The general case.

In general, the finite-dimensional variational problem (3.1) is equivalent to the finite-dimensional generalized eigenvalue problem [6]

$$Pc = \lambda_S M c \tag{3.4}$$

with $u_S = \sum_{j=1}^{N} c_j \phi_j$ and

$$P_{jk} = \langle T_\tau \phi_j, \phi_k \rangle_\mu, \qquad M_{jk} = \langle \phi_j, \phi_k \rangle_\mu.$$

By defining $\hat{c} = Mc$, we can rewrite (3.4) in form of the equivalent finite eigenvalue problem

$$\hat{T}\hat{c} = \lambda_S \hat{c}, \qquad \hat{T} = PM^{-1}. \tag{3.5}$$

The matrix $\hat{T}$ is a matrix representation of the so called projected transfer operator [6]

$$QT_\tau Q : S \to S,$$

where $Q : L^2_\mu \to S$ is the orthogonal projection onto S with respect to $\langle \cdot, \cdot \rangle_\mu$ that can be expressed as

$$(Qf)(x) = \sum_{i,j=1}^{N} \phi_i(x) M^{-1}_{ij} \langle \phi_j, f \rangle_\mu,$$

where M^{-1} is the inverse of the matrix M defined above and the inverse exists since the ϕ_j are linearly independent. That is, solving the restricted eigenvalue problem is equivalent to computing the eigenvalues of the projected operator QTQ.

The entries of the matrix P can always be rewritten in the following form:

$$P_{jk} = \langle \phi_j, T^* \phi_k \rangle_\mu = \int \mathbb{E}_x\Big(\phi_j(x)\phi_k(X_t)\Big)\mu(x)dx = \mathbb{E}_\mu\Big(\phi_i(X_0)\phi_k(X_t)\Big), \tag{3.6}$$

where $\mathbb{E}_x$ denotes expectation with respect to the paths of $(X_s)_{s\in[0,t]}$ when starting in $X_0 = x$. This shows that the entries of P are correlation functions of the underlying dynamics. Recently, an approach to Markov State Modelling using the Rayleigh Ritz variational formula for eigenvalues of self-adjoint transfer operators has been proposed [29, 30]. The resulting linear variation method again results in the restricted eigenvalue problem (3.4) with P and M in the above correlation matrix form.

Core set MSMs. The first alternative to standard (full-partition) MSM building schemes are so-called core set MSMs [6, 22, 21, 31, 26]: The ansatz space S results from the following set-based construction: We start with disjoint sets, the so-called core sets C_j for $j = 1, \ldots, N$ that do *not* partition the whole state space $\mathbb{X}$ but only form the core basins of the metastable sets. The ansatz space then is constructed by choosing the so called committor functions $\{q_j\}$, $j = 1, \ldots, N$, as ansatz functions [22]. The committor $q_j(x)$ is defined as the probability that starting in state $x \in \mathbb{X}$ the next core set that the Markov process will visit is C_j. The committors $\{q_j\}$, $j = 1, \ldots, N$ form a set of linearly independent, non-negative functions that constitute a partition of unity ($\sum_i q_i(x) = 1$ for all $x \in \mathbb{X}$). Setting $\phi_i = q_i$ we end up with discretization matrices

$$P_{jk} = \langle Tq_j, q_k\rangle_\mu, \qquad M_{jk} = \langle q_j, q_k\rangle_\mu.$$

Unfortunately, efficient explicit computation of the committor functions is not feasible in high dimensions. However, in core-set MSM buidling these ansatz functions are never computed explicitly. Instead one defines the so-called milestoning processes $(\hat{X}_t^-)$ and $(\hat{X}_t^+)$ based on the core sets C_j, by setting $\hat{X}_t^- = k$ if the original process at time t came last from C_k, and $\hat{X}_t^+ = k$ if the original process at time t went next to C_k. Utilizing this, the two matrices P and M can be written in the form [6, 22]

$$P_{jk} = \mathbb{P}_\mu\Big(\hat{X}_t^+ = k, \hat{X}_0^- = j\Big), \qquad M_{jk} = \mathbb{P}_\mu\Big(\hat{X}_0^+ = k, \hat{X}_0^- = j\Big), \tag{3.7}$$

and thus can be computed by means of trajectories of the MD process without need to compute the committor functions explicitly.

In comparison to standard full partition MSMs, core-set MSMs show significantly enhanced approximation quality (see next section) and in principle can also be constructed for high dimension systems as long as the system exhibits only a limited number of strongly metastable sets.

Meshless MSMs. In [24, 32, 33] a meshless discretization of the transfer operator has been presented. It uses the ansatz functions

$$\phi_i(x) = \frac{1}{Z_i}\exp(-\alpha\|x - x_i\|_2), \quad Z_i = \sum_j \exp(-\alpha\|x - x_j\|_2),$$

where $\|\cdot\|_2$ denotes the Euclidean norm, and the set of points x_i are free and adaptively chosen during the process of exploring state space by the dynamics. The ϕ_i form a non-negative partition of unity like the committor functions. But instead of being chosen in a problem-adapted way according to some pre-defined core sets like the committor functions, the meshless ansatz functions can be adapted to the exploration of state space by the dynamics by means of moving/choosing the points x_i appropriately.

3.3 Computing the discretized transfer operator

As we have seen, the Galerkin approach for discretizing the transfer operator lead to linear systems, where the entries of the matrices are given in terms of probabilities or expectation values. So computing approximate solutions of the restricted eigenvalue problem (3.1) is based on sampling these stochastic quantities. Constructing a full-partition MSM aims at estimating the matrix $\hat{T}$ from (3.3) given by

$$\hat{T}_{jk} = \mathbb{P}_\mu\Big(X_\tau \in A_k | X_0 \in A_j\Big).$$

Having sampled a long trajectory $(x_i)_{i=0,\dots,K}$ of the original process with stepsize Δt, a maximum likelihood estimator is given by [6]

$$\tilde{T}^{(K)}_{jk} = \frac{\#(x_{i+\Delta i} \in A_k | x_i \in A_j)}{\#(x_i \in A_j)},$$

where $\Delta i = \tau/\Delta t$. The above formula can be easily adapted to the case where a collection of short trajectories is available instead of one long trajectory (cf. [9]). Obviously, for core-set MSM building the construction can be done in close analogy. Alternative approaches utilize the fact that the process has to pass the surface between discretization boxes for efficient computation of the transition matrix $\tilde{T}$ without requiring long trajectories [34].

4 How accurate can MSMs be?

The computation of the discretization by the described methods is based on an appropriate construction of the ansatz space S on the one hand, and an accurate sampling of the probabilistic entries of the matrices on the other hand. The overall error can be decomposed into these two parts,

$$\|T_\tau - \tilde{T}^{(K)}\| \leq \|T_\tau - QT_\tau Q\| + \|QT_\tau Q - \tilde{T}^{(K)}\|. \tag{4.1}$$

Here, the first part describes the discretization error that is related to the approximation of the restricted eigenvalue problem, and the second part measures the sampling error relative to a pre-defined discretization.

4.1 Discretization error: Eigenvalues

The Galerkin ansatz allows for a quite general estimation of the discretization error: Assume that the ansatz functions ϕ_i that span the finite-dimensional ansatz space S are non-negative and form a partition of unity (which is the case for indicator functions, committor functions as well as meshless ansatz functions). Then $S \subset L^2(\mu)$ is a subspace with

$$\mathbb{1} \in S. \tag{4.2}$$

Furthermore, for an $m \leq \dim(S) =: n$, let $1 = \lambda_1 > \lambda_2 > ... > \lambda_m$ be the m dominant eigenvalues of a self-adjoint operator T_t, i.e. for every other eigenvalue λ it holds $\lambda < \lambda_m$. Let $u_1, u_2, ..., u_m$ be the corresponding normalized eigenvectors, and let Q denote the orthogonal projection onto S with respect to $\langle \cdot, \cdot \rangle_\mu$. Moreover, let $1 = \hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_m$ be the dominating eigenvalues of the projected operator

QTQ, that is, solutions of the corresponding restricted eigenvalue problem. Then [4, 6]

$$\max_{i=1,\dots,m} |\lambda_i - \hat{\lambda}_i| \leq \lambda_1 (m-1)\delta^2, \tag{4.3}$$

where

$$\delta = \max_{i=1,\dots,m} \|(\mathrm{Id} - Q)u_i\| \tag{4.4}$$

is the maximal projection error of the leading m eigenvectors to the space S. That is, the better my ansatz space allows to approximate the leading eigenvectors the smaller the discretization error will be. The generality of this result guarantees that it holds for standard full MSMs as well as core set or meshless MSMs. Surprisingly a similar result can be proved for any non-dominant eigenvalue of the transfer operator [35, 6], that is, the above statement is not limited to the cluster of the m largest eigenvalues. In [35] it is also discussed how to estimate the projection error δ just from trajectories, i.e., without using information on the eigenvectors.

The estimate (4.4) of the discretization error does neither require any assumptions on a spectral gap nor on the Markovianity of the MD process. The reasons for this are as discussed above: the eigenvalues characterize the dynamics in an ergodic (average) sense and not in any sense related to single trajectories.

4.2 Discretization error: Long-term transport

One can not only bound the discretization error regarding the eigenvalues, but also compare the full long-term transport of the MD process with the discretized process. The associated transport error in the L^2_μ-space is

$$E(k) = \|QT_\tau^k Q - (QT_\tau Q)^k\|,$$

where long-term transport belongs to large k. In [5] it has been shown that -under the same assumptions as above- this error depends on the lag time τ and the projection error δ (4.4) for the dominant m eigenvectors:

$$E(k) \leq (m\delta + \eta)\Big[m^{1/2}(k-1)\,\delta + \frac{\eta}{1-\eta}(1-\eta^{k-1})\Big] \cdot \exp\Big(-\frac{k\tau}{t_2}\Big), \tag{4.5}$$

with the spectral gap quantity

$$\eta = \lambda_{m+1}/\lambda_2 = \exp\Big(-\frac{t_2 - t_{m+1}}{t_2 t_{m+1}}\,\tau\Big).$$

In particular, whenever δ decreases faster as $m^{3/2}$ for increasing m, and in addition we are interested in timescales associated with large enough k, the error can be made arbitrarily small by

- increasing the number m, so that δ is small enough (choose discretization appropriately), and
- choosing the lagtime τ large enough so that η is small enough (select appropriate timescale).

In comparison to the result (4.3) on the eigenvalue error, the latter estimate bounds the full long-term transport error. The fact that one needs an additional spectral gap condition (make η small by selecting appropriate τ) shows that spatial relaxation requires some memory loss of the MD process that only comes with long enough timescales.

Remark. The estimate (4.5) reduces the much cruder one (1.1) by setting $t = k\tau$, and $C = \max(tm^{3/2}, (1-\eta^{k-1})/(1-\eta))$.

A critical commentary. Careful analysis of the above error uncovers another disadvantage of standard full partition MSMs: For a given standard MSM, reduction of the transport error below a certain threshold is only possible by further refining the partition in the transition region between the metastable sets where sampling is rare and its improvement computationally expensive, see [5] (theory) and in particular the extensive discussion of the practical consequences in [28]. This problem cannot be circumvented for standard MSMs but only by core set MSMs where the committors automatically incorporate the optimal discretization of the transition region. This explains their superior approximation quality of core-set MSMs in comparison to standard MSMs.

In many articles on practical applications of MSM building the discretization error is *completely ignored*: While the dependence of the overall error on the lagtime τ is checked carefully in many publications on MSM building, starting with [7, 8, 36], the discretization error often is not considered at all. Instead tests for Markovianity or simular consistancy checks are utilized in order to justify the validity of the resulting MSM; however, these tests can show positive results even if the underlying discretization error still is totally off [6, 28]. Thus an explicit warning seems appropriate: Checking for Markovianity and sufficiently long lagtime does not guarantee that the resulting MSM is accurately reproducing the longest timescales of the underlying MD process.

4.3 Sampling Error

Like the discretization error, also the sampling error can be estimated, at least for standard full partition and core set MSMs: Given available trajectory data $(x_i)_{i=0,\dots,K}$ one considers its discretized version $D = (y_i)_{i=0,\dots,K}$ with $y_i = j$ if $x_i \in A_j$. Then the probability $\mathbb{P}(D|\tilde{T}^{(K)})$ that this data has been produced by the transition matrix $\tilde{T}^{(K)}$ of a Markov chain is given by

$$\mathbb{P}(D|\tilde{T}^{(K)}) = \prod_{i=0}^{K-1} \tilde{T}^{(K)}_{y_i,y_{i+1}}.$$

Given a prior distribution ρ on transition matrix space, Bayes formula allows us to write down the posterior probability distribution of transition matrices [37]

$$\mathbb{P}(\tilde{T}^{(K)}|D) = \mathbb{P}(D|\tilde{T}^{(K)})\rho(\tilde{T}^{(K)}),$$

from which the sampling error can be computed. Several algorithms for sampling $\mathbb{P}(\tilde{T}^{(K)}|D)$ have been designed [37, 38, 39] so that a posteriori estimation of the sampling error is possible, e.g., by using the standard deviation of $\mathbb{P}(\tilde{T}^{(K)}|D)$ as an estimator for $\|QT_tQ - \tilde{T}^{(K)}\|$, or by computing the posterior distribution of the leading eigenvalues and eigenvectors from $\mathbb{P}(\tilde{T}^{(K)}|D)$.

Remark. Although the total error decomposes as in (4.1), it is important to note that the two error contributions are not independent, in general. Clearly, a very fine full-partition of state space will guarantee a small projection error, but increases the sampling effort dramatically because of the increasing number of transition probabilities that need to be estimated [31].

5 Pros and Cons

Let us now take a step back from all the details and review the advantages and disadvantages of MSM building based on Galerkin discretizations. The main pros and cons seem to be the following:

+ MSM building allows for error-controlled approximation of the longest timescales of the underlying MD process; reliable schemes for error estimation are available.
+ This approximation can be very precise if the discretization is chosen appropriately and the underlying exploration of state space based on short MD trajectory is sufficient.
+ The construction of MSMs is a not too sophisticated procedure, requires short MD trajectories only, can be easily parallelized, and is applicable to most molecular systems with dominant metastable sets.
– The process of finding appropriate discretizations (full box partition or good core sets) is still not automated but essentially case- and expert-dependent.
– In order to get acceptably small error for high dimensional systems additional information about the system at hand is needed: For choosing not-too-large full partitions in high dimensions, information on the reaction coordinates of the system is required; this can be avoided in core-set MSM but then the choice of good core sets requires knowledge on the main metastable regions in state space.
– Guaranteeing sufficient sampling still is the main bottleneck for large molecular systems. Integration of enhanced sampling methods into MSM building is possible but cannot fully cure the problem. As a result, for most really large molecular systems, reliable estimation of the MSM approximation error is infeasible since one cannot guarantee that all important parts of the state space have been explored.

The combination of the obstacles of choosing good discretizations and getting sufficient sampling in high dimensions motivates an alternative approach to MSM building that will be discussed in the next section.

6 The Future: Collocation Discretization of Transfer Operators

The discretization by collocation is a partial remedy for the two main problems of the Galerkin approach to MSM building. It is based on the idea of satisfying the eigenvalue problem of T_τ just on some preselected collocation points: Again assume that we can represent the eigenvalue u by ansatz functions $\{\phi_j\}$, so

$$u(x) = \sum_{j=1}^{N} a_j \phi_j(x). \tag{6.1}$$

Applying the transfer operator $T_\tau = T_\tau^*$ gives

$$T_\tau u(x) = \sum_{j=1}^{N} a_j \mathbb{E}_x\Big(\phi_j(X_t)\Big).$$

Now we only require that the eigenvalue problem for T_τ is satisfied at N collocation points $x_1, \ldots, x_N$

$$T_t u(x_k) = \lambda u(x_k), \quad k = 1, \ldots, N.$$

This again yields a finite eigenvalue problem

$$Ea = \lambda Ba$$

with

$$E_{kj} = \mathbb{E}_{x_k}\Big(\phi_j(X_t)\Big), \quad B_{kj} = \phi_j(x_k). \tag{6.2}$$

This way, the computation of the matrix entries (6.2) for the collocation method makes only use of probability measures of the form $\mathbb{P}_{x_k}$, where the process is always initially started at a collocation point. That is, no probabilities of the form $\mathbb{P}_\mu$ have to be computed and no explicit information on μ is needed.

Is this still an MSM? Collocation discretization will in general *not* lead to stochastic discretization matrices in stark contrast to the cases discussed in most articles on practical MSM building. It thus may seem contra-intuitive if we still use the name "Markov State Model" for the resulting finite dimensional problem. However, the matrices resulting from collocation discretization allow for an aggregation into a Markov model in a post-processing step. We will discuss this issue further in our numerical example in Sec. 6.2.

Computation from trajectories. Accurate sampling of the matrix E from (6.2) can be achieved by starting K independent trajectories of length τ from each collocation point x_j. If $y_i(x_j)$ denotes the end point of the ith trajectory of length τ starting in x_j, then

$$\tilde{E}^{(K)}_{kj} = \frac{1}{K}\sum_{i=1}^{K} \phi_k(y_i(x_j))$$

is a maximum likelihood estimator for the matrix E. Note that for this sampling no ergodicity of the trajectory has to be assumed since knowledge of the invariant density is not needed.

6.1 Ansatz Functions

We cannot go into details on the choice of ansatz functions and collocation points for collocation discretization of transfer operators. We just want to add some comments on the fundamental opportunities that result from the fact that we have complete freedom in choosing the ansatz functions.

Trigonometric collocation. Whenever we look at torsion or peptide angles in molecular dynamics, we face periodic potentials. Hence, periodic ansatz functions like the basis of the discrete Fourier transform would be an appropriate choice for approximation. That is, we could choose $N = 2n + 1$ and

$$\psi_1 = 1/2, \quad \psi_{2k} = \cos(kx), \quad \psi_{2k+1} = \sin(kx), \quad k = 1, \ldots, n$$

together with uniform collocation points $x_k = (k-1)2\pi/n$, $k = 1, \ldots, n$, such that

$$S = \{f : \quad f(x) = \frac{1}{2}a_1 + \sum_{k=1}^{n} \Big(a_{2k}\cos(kx) + a_{2k+1}\sin(kx)\Big).$$

Additionally, the unitarity of the discrete Fourier transform guarantees that for any function in L^2_μ our approximation will converge if we let N tend to infinity.

Polynomial collocation. If one considers bonding potentials for non-angular coordinates, one might want to choose polynomial ansatz functions. This becomes obvious if one considers the prototypical example of a harmonic potential, where the eigenfunctions of the transfer operator in one dimension are given by the Hermite polynomials H_n, $n = 1, 2 \ldots$. We thus propose the use of the ansatz space

$$S = \{f : \quad f(x) = \sum_{j=1}^{n} a_j H_j(x)\},$$

with Hermite polynomial centered to the equilibrium position of the bonding potential and with collocation points resulting from the zeros of H_n.

6.2 Numerical example

Since collocation discretization will in general *not* lead to stochastic discretization matrices, we will now demonstrate how the MSMs resulting from collocation will look like. As an example we choose a diffusion process in a periodic two-well potential. The process (X_t) is a solution of the following stochastic differential equation

$$dX_t = -\nabla V(X_t)dt + \sigma dB_t,$$

where B_t denotes standard Brownian motion, and the noise intensity σ is associated with an inverse temperature $\beta = 2/\sigma^2$. The potential V is illustrated in Fig. 1 and we use $\beta = 3$.

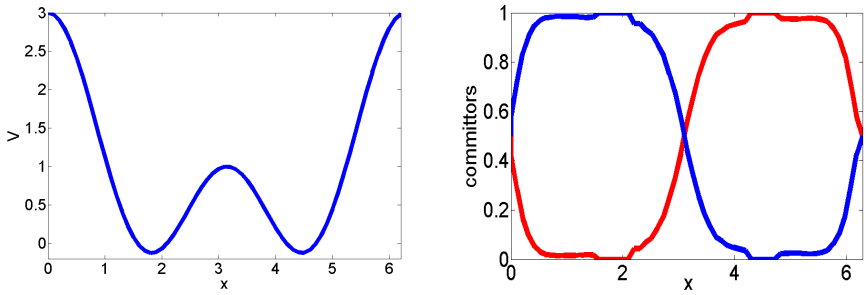

Figure 1: Left: Periodic two-well potential energy function V. Right: Committors for the vicinities of the two minima of the potential energy as core sets C_1, and C_2.

First, we choose $N = 7$ and apply the trigonometric collocation method with equidistant collocation points ($x_j = 2j\pi/n$ for $j = 0, \ldots, n-1$ as usual in FFT). We find the following estimates of the leading three eigenvalues resulting from the generalized eigenvalue problem $Eu = \lambda Bu$ as in (6.2)

$$\tilde{\lambda}_3, \tilde{\lambda}_2, \tilde{\lambda}_1 = 0.1003, 0.9724, 1.0000$$

while the exact eigenvalues of the transfer operator are given by

$$\lambda_3, \lambda_2, \lambda_1 = 0.1138, 0.9682, 1.0000.$$

The matrix $\hat{T} = B^{-1}E$ can also be computed:

$$\hat{T} = \begin{pmatrix} 1.0000 & -0.5216 & -1.2763 & 0.7867 & 0.0211 & 0.0079 & -0.0294 \\ 0 & 0.1311 & -0.1215 & -0.0842 & -0.0040 & 0.0025 & 0.0048 \\ 0 & 0.0233 & 0.0456 & -0.0690 & 0.0167 & -0.0096 & -0.0087 \\ 0 & 0.0467 & -0.0050 & -0.0493 & 0.0009 & 0.0013 & -0.0036 \\ 0 & 0.0080 & -0.0102 & -0.0078 & 1.0690 & -0.4921 & -0.5049 \\ 0 & -0.0036 & 0.0065 & 0.0079 & -0.0100 & 0.0633 & -0.0607 \\ 0 & 0.0071 & -0.0153 & 0.0129 & 0.2030 & -0.1035 & -0.0752 \end{pmatrix},$$

Note that in contrast to the transition matrix in classical Markov State Modelling this matrix is not a stochastic matrix. At first glance, this might seem to be a disadvantage because a transition matrix directly relates to a Markov chain which can be interpreted as a dynamical approximation of the original process. On the other hand, the limitation to ansatz functions that lead to stochastic matrix approximations is not necessary if we are interested in approximating the variational eigenvalue problem. In particular, information about the eigenvectors and eigenvalues is valuable for constructing a discretization that an accurate Markov State Model can be built on.

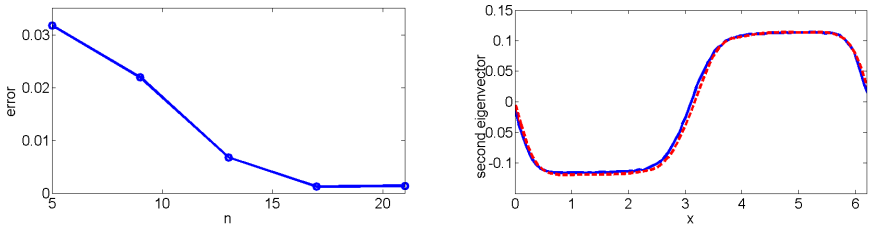

Figure 2: Results for trigonometric collocation method. Left: Error in second eigenvalue for different numbers of collocation points. The error remaining for $n > 15$ is not related to discretization but results from the finite sampling. Right: Comparison of second eigenvector for $N = 7$ with exact second eigenvector.

In Fig. 2 one can see that the approximation of the second eigenvalue by the periodic ansatz functions of the discrete Fourier transform converges quickly, but also gives a good result with only few collocation points used. The second eigenvector tells us how to coarse grain the collocation MSM further: It exhibits a sign change at $x_0 \approx \pi$ so that we can coarse grain using two macrostates, $A = [0, x_0]$, and $B = [x_0, 2\pi]$. By using these sets and the available trajectory information (which has been collected while computing the collocation discretization for $N = 7$), we get an estimate for the respective 2×2 MSM transition matrix

$$\hat{T} = \begin{pmatrix} 0.9876 & 0.0124 \\ 0.0124 & 0.9876 \end{pmatrix},$$

with eigenvalues 0.9752 and 1. In order to get the same accuracy by using a standard full partition MSM with uniform boxes, we need $N = 19$ boxes.

In comparison, the core set approach with 2 sets (ball of radius 0.3 around the two minima of the potential, see right hand panel in Fig. 1) gives the estimate $\tilde{\lambda}_2 = 0.9685$ for the second eigenvalue which is three digits accurate. In comparison, the standard full partition MSM approach only achieves this accuracy if more than $n = 30$ boxes are used. This demonstrates that the core set approach –provided good choices for the core sets are available– has superior approximation properties.

6.3 Sparse Tensor Approximation in High Dimensions

Assume now that we are dealing with two dimensions, an angular one, x_1, and a bond-length like dimension, x_2. Then, we consider approximations in which we first expand $f(\cdot, x_2)$ for fixed x_2 in terms of trigonometric ansatz functions, and then the x_2-dimension in Hermite polynomial

$$\begin{aligned} f(x_1, x_2) &= \sum_{i_1=1}^{n_1} c_{i_1}(x_2)\psi_{i_1}(x_1) \\ &= \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2} a_{i_1,i_2}\psi_{i_1}(x_1)H_{i_2}(x_2). \end{aligned}$$

In d dimensions this kind of tensor approximation leads to the ansatz space

$$S_{n_1,\ldots,n_d} = \{f(x_1,\ldots,x_d) = \sum_{i_1,\ldots,i_d=1}^{n_1,\ldots,n_d} a_{i_1,\ldots,i_d}\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)\}$$

where ϕ_i is either trigonometric or polynomial depending on the dimension x_i. The dimension of this ansatz space is $n_1 \times \ldots \times n_d$ and thus exponential in d. In order to avoid the curse of dimensions, *sparse tensor approximation* considers ansatz spaces like

$$\begin{aligned} S_n &= \{f(x_1,\ldots,x_d) = \sum_{i_1,\ldots,i_d\in I_n} a_{i_1,\ldots,i_d}\phi_{i_1}(x_1)\cdot\ldots\cdot\phi_{i_d}(x_d)\} \\ I_n &= \{(i_1,\ldots,i_d) : \prod_{k=1}^{d}\max(1,i_k) < n\} \end{aligned}$$

The literature on sparse tensor approximation shows that S_n still has sufficient approximation properties [40] and that the dimension of S_n just grows like $n(\log n)^{d-1}$.

It has been shown in [23] that the dominant eigenvectors of the transfer operator are almost constant in the direction of *fast* degrees of freedom. Therefore, any tensor ansatz can be limited to $n_i = 1$ if x_i is one of the fast dimensions. Numerical schemes for identifying such dimensions in the framework of sparse tensor approximation are available (so-called adaptive dimension schemes [41]). In this way, collocation MSMs based on S_n in combination with adaptive dimension schemes open the opportunity for the construction of MSM building schemes that can be applied in an automatic way even in very high dimensions.

6.4 Pros and Cons of Collocation

Let us very shortly discuss the advantages and disadvantages of MSM building based on collocation:

+ Collocation schemes allow for sparse tensor approximations that avoid the curse of dimensions and thus in principle would allow for automated MSM building in very high dimensions.

+ Collocation schemes do *not* require that the invariant measure has to be sampled. The collocation points result from the form of ansatz functions used.

– Presently no validation (apart mathematical approximation theory and demonstrations for test system) of collocation-based MSM building for realistic molecular systems has been given.

– Presently no theory for estimating the approximation error of collocation-based MSMs has been developed.

7 Concluding Remarks

Based on the pros and cons already discussed above we will restrict our concluding remarks to the following three:

1. Almost all practical applications of MSM building use standard full partition MSMs, cf. [16, 10]. In view of the rich variety of alternative MSM constructions and their respective advantages, this is rather surprising. The explanation may be that standard full partition MSMs provide trivial interpretation in form of a Markov process that jumps between the discretization sets with just the right probability which gives a direct kinetic meaning to it. However, all the alternative MSMs allow for the same construction if wanted but deliver superior approximation properties.

2. Galerkin-based MSMs do not seem to be the optimal solution for construction of accurate approximations and reliable sampling in very high dimensions, in particular if one cannot identify appropriate reaction coordinates to which discretization can be limited. Instead sparse tensor approximation of transfer operators need to be developed further. Collocation discretization of transfer operators seems to allow for this development because it opens the door for using dimension-adaptive tensor approximations, cf. [41].

3. Most of the theory of MSM building is based on the assumption that the molecular dynamics process considered is *in equilibrium* and has a unique invariant measure. It covers scenarios in which the relaxation of a molecular system back to equilibrium is described but it does *not* fully apply to nonequilibrium molecular dynamics. There are some first approaches, like [26], that start to develop the fundamentals of MSM building for nonequilibrium MD but the field is still rather unexplored.

Despite the concept of MSM building in molecular dynamics is more than 15 years old already and has developed into a toolkit that is successfully and widely utilized in molecular research, its potential has not been fully explored and it seems that many interesting developments can well be expected.

References

[1] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999. Special Issue on Computational Biophysics.

[2] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

[3] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 161(184), 2005. 398 Special issue on matrices and mathematical biology.

[4] N. Djurdjevac, M. Sarich, and Ch. Schütte. Estimating the eigenvalue error of Markov state models. *Multiscale Modeling & Simulation*, 10(1):61–81, 2012.

[5] Marco Sarich, Frank Noé, and Christof Schütte. On the approximation error of Markov state models. *SIAM Multiscale Model. Simul.*, 8:1154–1177, 2010.

[6] Ch. Schütte and M. Sarich. *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, volume 24 of *Courant Lecture Notes*. American Mathematical Society, December 2013.

[7] William C. Swope, Jed W. Pitera, and Frank Suits. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B*, 108:6571–6581, 2004.

[8] J. Chodera, N. Singhal, V. S. Pande, K. Dill, and W. Swope. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics*, 126, 2007.

[9] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl. Constructing the full ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. USA*, 106:19011–19016, 2009.

[10] V. Pande, K. Beauchamp, and G. Bowman. Everything you wanted to know about Markov state models but were afraid to ask. *Methods*, 52(1):99–105, 2010.

[11] Kai J Kohlhoff, Diwakar Shukla, Morgan Lawrenz, Gregory R Bowman, David E Konerding, Dan Belov, Russ B Altman, and Vijay S Pande. Cloud-based simulations on google exacycle reveal ligand modulation of gpcr activation pathways. *Nature chemistry*, 6(1):15–21, 2014.

[12] C. R. Schwantes, R. T. McGibbon, and V. S. Pande. Perspective: Markov models for long-timescale biomolecular dynamics. *The Journal of Chemical Physics*, 141(9):–, 2014.

[13] C. Schütte. Conformational dynamics: Modelling, theory, algorithm, and application to biomolecules, 1998. Habilitation Thesis.

[14] P. Deuflhard, M. Dellnitz, O. Junge, and C. Schütte. Computation of essential molecular dynamics by subdivision techniques. In *Lecture Notes in Computational Science and Engineering*, pages 98–115. Springer, 1999.

[15] B. G. Keller, J.-H. Prinz, and F. Noé. Markov models and dynamical fingerprints: Unraveling the complexity of molecular kinetics. *Chem. Phys.*, 396:92–107, 2012.

[16] G. R. Bowman, V. S. Pande, and F. Noé, editors. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, volume 797 of *Advances in Experimental Medicine and Biology*. Springer, 2014.

[17] M. Senne, B. Trendelkamp-Schroer, A. S. J. S. Mey, Ch. Schütte, and F. Noé. Emma - A software package for Markov model building and analysis. *J. Chem. Theory Comput.*, 8:2223–2238, 2012.