Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

MARCUS WEBER, KONSTATIN FACKELYDEY, CHRISTOF SCHÜTTE

Set-Free Markov State Model Building

ZIB Report 17-10 (March 2017)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

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

Set-Free Markov State Model Building

Marcus Weber, Konstantin Fackeldey, Christof Schütte

February 2017

1 Introduction

Molecular dynamics simulations are often used to estimate transition probabilities in phase space (or state space, respectively). In principle one is interested in the following quantity

$$P(t, A, B) = \mathbb{P}_\mu\Big(X_t \in B | X_0 \in A\Big). \tag{1}$$

In this formula A and B are subsets of the state space. $P(t, A, B)$ denotes the conditional probability for a molecular process starting in A to end up in B after a certain time span t. This quantity is not meant for one single trajectory, it is an expectation value over *all* possible trajectories starting in A. Usually, A and B represent metastable conformations of the system, such that these transition probabilities are small and can not be sampled efficiently by performing long-time trajectories of the molecular system. More precisely, estimating $P(t, A, B)$ based on molecular simulation of many trajectories starting in A and rarely reaching B in a given time t, is not a good idea. In this article we recall, that (1) can be seen as a Galerkin discretization of a transfer operator T_t. The (time dominating) slowest processes of the molecular system are correlated with the highest eigenvalues and their eigenfunctions of T_t. Thus, finding dominating molecular processes can be turned into solving an eigenfunction problem of a (self-adjoint) operator T_t. The sampling problem turns into an eigenfunction problem.

Is this class of problems easier to solve? It is a good question, because the state space is high-dimensional and we try to solve a function approximation problem in this space. Furthermore, for a Galerkin discretization we need to compute inner products, which means we have to solve high-dimensional integrals. Both of these problems have been addressed in literature[22]. For the solution of a function approximation problem we will propose an adaptive meshless discretization method. Whereas, for the quadrature problem we will apply a Monte-Carlo approach which will be based on biased stochastic MD simulations. There is a good hope to overcome the sampling problem by this approach, because the approximated eigenfunctions have a rather simple structure (nearly piece-wise constant). The more rarely transitions between A and B occur, the easier the eigenfunctions become. On the other hand, there is no need for the simulation of trajectories which connect A and B. For Monte-Carlo quadrature the trajectories are only needed for the evaluation of (localized) inner products.

2 Theory

2.1 Molecular Dynamics

We consider a molecular system for which we would like to know transition probabilities (1) according to a given stochastic or deterministic molecular dynamics. Let $\mathbb{X}$ denote its phase space or state space, respectively, depending on the form of MD. The result of a MD simulation is a trajectory (x_t), where t runs through all discrete time steps of the simulation and $x_t \in \mathbb{X}$ for all t. Typically this trajectory results from (temporal) discretization of a set of equations of motion governed by the (potential) energy function V that in principle have time-continuous trajectories $(X_t)_{t\in\mathbb{R}}$ as their solutions. For canonical ensemble MD one typically assumes that (x_t) is (approximately) distributed according to the stationary distribution or invariant measure

$$\mu(x) = \frac{1}{Z}\exp(-\beta V(x)),$$

resulting from the ergodicity of (X_t) which guarantees that for an appropriate macro-observable f we have

$$\lim_{T\to\infty}\frac{1}{T}\int f(X_t)dt = \int f(x)\mu(x)dx.$$

When one does very long MD simulations repeatedly starting from the same initial state, then the resulting trajectories differ. This may be caused by the chaotic properties of thermostatted MD [16] for long enough timescales or it may result from the fact that the underlying equations of motion are stochastic like in the case of Langevin or Smoluchowski dynamics. For the following it does not matter which of the two cases we consider; we will just assume that there is a probability distribution $p(t, x, y)$ for observing trajectories (X_t) that start in $X_0 = x$ and end in $X_t = y$. The *transfer operator* T_t associated with the dynamics is defined by

$$T_t f(x) = \mathbb{E}\Big(f(X_t)|X_0 = x\Big) = \int f(y)p(t, x, y)dy,$$

where the condition $X_0 = x$ means that all trajectories start in x. The transfer operator tells us how the expectation value of an observable f is evolving under the dynamics. Specific forms of the transfer operator, e.g., for thermostatted Hamiltonian dynamics, can be found in literature[16, 20].

By introducing the scaled scalar product

$$\langle f, g\rangle_\mu = \int f(x)g(x)\mu(x)dx$$

we can write equilibrium correlation functions as

$$\begin{aligned}\langle f(X_t)g(X_0)\rangle &= \mathbb{E}\Big(f(X_t)g(X_0)|X_0 \sim \mu\Big)\\ &= \int g(x)Tf(x)\mu(x)\,dx = \langle g, Tf\rangle_\mu\\ &= \langle T^* f, g\rangle_\mu.\end{aligned}$$

where T^* denotes the adjoint transfer operator. Similarly, we find that the transition probability between sets A and B in time t can be expressed via the transfer operator also:

$$\begin{aligned} P(t,A,B) &= \frac{1}{\mu(A)}\int_A\int_B p(t,x,y)\mu(x)dxdy \\ &= \frac{1}{\mu(A)}\langle \mathbf{1}_A, T\mathbf{1}_B\rangle_\mu \\ &= \frac{1}{\mu(A)}\langle \mathbf{1}_B(X_t)\mathbf{1}_A(X_0)\rangle, \end{aligned} \tag{2}$$

where $\mu(A) = \int_A \mu(x)dx = \langle \mathbf{1}_{\mathbb{X}}, \mathbf{1}_A\rangle_\mu$ and $\mathbf{1}_A$ denotes the characteristic function of set A, i.e., $\mathbf{1}_A(x) = 1$ if $x \in A$ and $= 0$ otherwise.

The dynamics is called reversible if (X_t) satisfies the detailed balance condition $\mu(x)p(t,x,y) = \mu(y)p(t,y,x)$. If this is the case then the transfer operator is self-adjoint with respect to the scaled scalar product, i.e, we have $T_t = T_t^*$ such that all of its eigenvalues are real-valued [19, 20]. Then, since the largest (in modulus) eigenvalue of any transfer operator for ergodic dynamics is $\lambda = 1$, the eigenvalues of T_τ for an arbitrary time $\tau > 0$ can be ordered,

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots .$$

Let u_j denote the eigenfunction of T_τ associated with λ_j,

$$T_\tau u_j = \lambda_j u_j.$$

Because of the self-adjointness of T_τ these eigenfunctions are orthogonal wrt $\langle\cdot,\cdot\rangle_\mu$ [16] such that we find that [2, 13, 14]

$$\langle f(X_{n\tau})g(X_0)\rangle = \sum_{j=1}^{m} \lambda_j^n \langle g, u_j\rangle_\mu \langle u_j, f\rangle_\mu.$$

That is, for large enough times (=large enough n) the largest eigenvalues of T_τ completely dominate the kinetic relaxations in the molecular system on associated dominant time scales [20, 2]

$$t_j = -\frac{\tau}{\log\lambda_j}, \quad j > 1$$

with $t_1 = \infty$ describing the asymptotic decay to the stationary distribution at infinite time. The observation, that the dominant eigenvalues of the transfer operator describe the long time relaxation kinetics and transition probabilities of the molecular system, can also be verified for non-reversible systems like Langevin dynamics [20].

The easiest example for the equations of motion in molecular dynamics is diffusion in the energy landscape V, i.e.,

$$\dot{x}_t = -\nabla_x V(x_t) + \sqrt{2\beta^{-1}}\dot{B}_t, \tag{3}$$

where B_t denotes Brownian motion. This kind of dynamics is just a simplistic example for reversible MD but will provide a means for illustration in the following. The dynamics given by (3) is reversible and ergodic wrt μ [20]. In this case the transfer operator has the form [19]

$$T_t = \exp(tL), \quad L = \beta^{-1}\Delta_x - \nabla_x V(x)\cdot\nabla_x,$$

where Δ_x denotes the Laplacian operator.

2.2 Generalized Eigenvalue Problem

The main idea of MSM building is to construct a (small) transition matrix whose leading eigenvalues and -vectors are very good approximation of the dominant eigenvalues and -functions of the transfer operator T_τ of the system for some preselected timescale τ. Therefore we have to consider the eigenvalue problem

$$Tu = \lambda u \tag{4}$$

of the full transfer operator $T = T_\tau$ of the dynamics.

As a starting point for the discretization of the eigenvalue problem (4) we consider a set of m non-negative ansatz functions $\Phi_1, \ldots, \Phi_m$ that form a partition of unity, i.e., for all states $x \in \mathbb{X}$ we have that

$$\sum_{j=1}^{m} \Phi_j(x) = 1.$$

Each ansatz function Φ_j represents the weight

$$\hat{\mu}_j = \int \Phi_j(x)\mu(x)dx = \langle \Phi_j, \mathbf{1} \rangle_\mu > 0,$$

since the partition of unity property guarantees that $\sum_j \hat{\mu}_j = 1$.

All functions that can be expressed as linear combinations of these ansatz functions form the linear subspace

$$D = \{u : \quad u = \sum_{j=1}^{m} \bar{u}_j \Phi_j, \ \bar{u}_j \in \mathbb{R}\}.$$

Based on the ansatz space D, Galerkin discretization of the eigenvalue problem (4) assumes [6, 20]

$$u = \sum_{j=1}^{m} \bar{u}_j \Phi_j, \quad \text{i.e.,} \quad \sum_{j=1}^{m} \bar{u}_j (T\Phi_j - \lambda \Phi_j) = 0,$$

then multiplies the last equation from the left with $\Phi_k/\hat{\mu}_k$, using the scalar product $\langle \cdot, \cdot \rangle_\mu$, which results in

$$\sum_{j=1}^{m} \frac{1}{\hat{\mu}_k} \Big(\langle \Phi_k, T\Phi_j \rangle_\mu - \lambda \langle \Phi_k, \Phi_j \rangle_\mu \Big) \bar{u}_j = 0. \tag{5}$$

Introducing the two $m \times m$ matrices P and M with entries

$$P_{kj} = \frac{1}{\hat{\mu}_k} \langle \Phi_k, T\Phi_j \rangle_\mu, \quad M_{kj} = \frac{1}{\hat{\mu}_k} \langle \Phi_k, \Phi_j \rangle_\mu,$$

and the coefficient vector $\bar{u} = (\bar{u}_j)_{j=1,\ldots,m}$ we can write equation (5) in the form of a generalized eigenvalue problem [17, 20]

$$P\bar{u} = \lambda M \bar{u}. \tag{6}$$

It has been shown [6, 20] that the solution of this generalized eigenvalue problem is the best possible approximation of the solution of the full eigenvalue problem (4) if restricted to the ansatz space D. P is a stochastic matrix with stationary distribution $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_m)$. In fact it can be shown that the discretized eigenvalue problem (6) inherits most of the structural properties of (4), e.g., no eigenvalue has absolute value larger 1 and all its eigenvalues are real-valued if the dynamics is reversible [20].

2.3 Set-Based MSMs

The most prominent example for a partition of unity results from set complete partition of state/phase space: Assume that the disjoint sets $A_1, \ldots, A_m$ decompose state space, i.e., $\cup_j A_j = \mathbb{X}$. Then their characteristic functions $\Phi_j = \mathbf{1}_{A_j}$ form a partition of unity and we find that [20]

$$P_{kj} = P(\tau, A_k, A_j)$$

and $M_{kj} = \delta_{kj}$, the standard identities used in set-based MSM building. In this case, MSM building reduces to (A) computing the transition matrix P and (B) solving the eigenvalue problem $P\bar{u} = \lambda\bar{u}$ for the dominant eigenvalues and eigenvectors. Since the entries of P are just the transition probabilities between the discretization sets A_j, it can be seen easily that the computation of P_{kj} only requires a couple of trajectories of length τ starting in A_k and setting P_{kj} to the fraction of these trajectories that ends up in A_j. This procedure has been studied to quite some extent [16, 19, 17, 20, 21] and has been applied to a variety of molecular systems. Its main disadvantage is that a reduction of the error of the discretization often requires the unwanted refinement of sets in the transition region between the metastable sets which can lead to an explosion of the number of sets needed [15].

Therefore, alternative forms of set-based MSM building do not use complete set partitions of state space but require some disjoint core sets or milestones only that have to be placed at the attractive core of the main metastable sets [17]. Then the partition of unity is formed by the committor functions associated with these core sets. Again, the computation of the discretization matrices P and M can be reduced to transition counts based on trajectories [20, 18]. However, depending on the size and location of the core sets quite long trajectories may be needed which can create severe limitations [18].

2.4 Set-Free MSMs

The form of partition of unity that we will discuss next is constructed using radial basis functions: Based on a collection of m so-called *base points* $q_j \in \mathbb{X}$, we define

$$\Phi_j(x) = \gamma(x) \exp\Big(- \alpha_j d(x - q_j)^2 \Big), \tag{7}$$

where α_j is a positive constant, d a distance measure (d-proximity), for example the Euclidean measure $d(x)^2 = \sum_{k=1}^n x_j^2$, and γ is chosen so that the partition of unity condition is satisfied, i.e,

$$\gamma(x)^{-1} = \sum_{j=1}^m \exp\Big(- \alpha_j d(x - q_j)^2 \Big).$$

The efficient computation of the discretization matrices P and M based on the ansatz space D associated with a radial basis function based partition of unity has rarely been discussed[22] up to now and will form the main part of the next section.

2.5 Discretization Error

In order to estimate the discretization error of the eigenvalue problem we first have to understand that the Galerkin discretization can be understood as an orthogonal projection Q (orthogonal wrt. $\langle \cdot, \cdot \rangle_\mu$) from the full function space to the ansatz space D,

$$Qv \;=\; \sum_{k,j=1}^{n} (M^{-1})_{kj} \, \langle \Phi_k, v \rangle_\mu \, \Phi_j. \tag{8}$$

Let the dominant eigenvalues of the full transfer operator $T = T_\tau$ be denoted

$$1 = \lambda_1 > \lambda_2 \geq ... \geq \lambda_m, \tag{9}$$

with associated eigenfunctions u_j, $j = 1, \ldots, m$. Moreover let the leading eigenvalues of the generalized eigenvalue problem (6) be denoted

$$1 = \hat{\lambda}_1 > \hat{\lambda}_2 \geq ... \geq \hat{\lambda}_m.$$

Then the discretization error is characterized by

$$\max_{j=2,\ldots,m} |\lambda_j - \hat{\lambda}_j| \leq 2\lambda_j \delta,$$

where δ denotes the projection error of the leading eigenvectors to D,

$$\delta = \max_{j=1,\ldots,m} \|u_j - Qu_j\|,$$

where $\|v\| = \sqrt{\langle v, v \rangle_\mu}$.

3 Methods

In this section, we will first construct efficient methods for computing the entries of the discretization matrices P and M based on radial basis ansatz functions of the form (7). Second, we will present an algorithm for refining a discretization by adding ansatz functions in order to decrease the error of the discretization.

3.1 Computation of P and M

We first have to discuss how to compute scalar products of the form

$$\langle \Phi_k, u \rangle_\mu = \int \Phi_k(x) u(x) \mu(x) \, dx,$$

for an arbitrary function u that we will specify later. Since $\mu(x) = \exp(-\beta V(x))/Z$ we have

$$\langle \Phi_k, u \rangle_\mu = \int \frac{\gamma(x)}{Z} u(x) \exp\Big(- \beta \, U_k(x) \Big) \, dx,$$

with energy function

$$U_k(x) = V(x) + \frac{\alpha_k}{\beta} d(x - q_k)^2, \tag{10}$$

which is composed out of the original energy function plus an extra harmonic potential. Molecular dynamics simulation runs with the potential U_k instead of V are easily implemented in all standard MD codes and will sample the invariant measure

$$\mu_k(x) = \frac{1}{Z_k} \exp\Big(- \beta\, U_k(x)\Big),$$
$$Z_k = \int \exp(-\beta\, U_k(x))dx,$$

so that

$$\langle \Phi_k, u\rangle_\mu = \frac{Z_k}{Z} \int \gamma(x) u(x) \mu_k(x)\, dx.$$

When performing MD simulations with energy function U_k we will produce sample points $(x_{1,k}, \ldots, x_{L,k})$ that are distributed according to μ_k. For each potential U_k this set of points will be refered to as the *restraint sampling points* in the followings. We remark, that for the procedure which generates the restraint sampling points, *any* rapidly-mixing sampling scheme can be applied independent from the choice of the desired T_t-dynamics. The reason is, that these sampling points only act as quadrature points for a Monte Carlo evaluation of the inner products. In the numerical examples, therefore, a standard Metropolis Hastings (MC) sampling method will be applied.

Using the restraint sampling points, the above integral can be approximated by the respective mean value

$$\langle \Phi_k, u\rangle_\mu \approx \frac{Z_k}{Z} \frac{1}{L} \sum_{i=1}^{L} \gamma(x_{i,k}) u(x_{i,k}),$$

where the error will vanish for $L \to \infty$. This result allows us to calcalute the entries of the mass matrix M,

$$M_{kj} = \frac{\langle \Phi_k, \Phi_j\rangle_\mu}{\langle \Phi_k, \mathbf{1}\rangle_\mu} \approx \tilde{M}_{kj}(L) = \frac{\sum_{i=1}^{L} \gamma(x_{i,k}) \Phi_j(x_{i,k})}{\sum_{i=1}^{L} \gamma(x_{i,k})}. \tag{11}$$

The entries of P, however, require knowledge regarding the action of T on Φ_j:

$$P_{kj} \approx \frac{\sum_{i=1}^{L} \gamma(x_{i,k}) T\Phi_j(x_{i,k})}{\sum_{i=1}^{L} \gamma(x_{i,k})}.$$

In order to approximate $T\Phi_j(x_{i,k})$ we need K trajectories of time length τ of the original molecular dynamics with respect to energy function V, all starting in $x_{i,k}$. Let $z_{i,k,l}$ denote the end point of the lth of these trajectories. Then

$$T\Phi_j(x_{i,k}) = \mathbb{E}\Big(\Phi_j(X_\tau) | X_0 = x_{i,k}\Big) \approx \frac{1}{K} \sum_{l=1}^{K} \Phi_j(z_{i,k,l}),$$

where, again, the error will vanish for $K \to \infty$. Putting things together, we can calculate the entries of P via

$$P_{kj} \approx \tilde{P}_{kj}(L, K) = \frac{1}{K} \frac{\sum_{i=1}^{L} \sum_{l=1}^{K} \gamma(x_{i,k}) \Phi_j(z_{i,k,l})}{\sum_{i=1}^{L} \gamma(x_{i,k})}. \tag{12}$$

Conclusively we found the following algorithm for approximately computing the entries of P_{kj} and M_{kj}, and approximate the dominant eigenvalues λ and eigenfunctions u of T:

1. Assume that the ansatz functions are of the form (7). Select large enough K and L.
2. For each $k = 1, \ldots, m$
   - Compute L sampling points $x_{k,i}$, $i = 1, \ldots, L$, using MD simulation with respect to the energy function U_k given in (10).
   - Compute $\tilde{M}_{kj}(L)$ for all $j = 1, \ldots, m$ using (11).
   - For each $i = 1, \ldots, L$: Compute K trajectories of length τ with end points $z_{k,i,l}$, $l = 1, \ldots, K$, using MD simulation with respect to the original energy function V.
   - Compute $\tilde{P}_{kj}(L, K)$ for all $j = 1, \ldots, m$ using (12).
3. Solve the generalized eigenvalue problem $\tilde{P}\bar{u} = \lambda\tilde{M}\bar{u}$.

3.2 Finding metastable sets

Finding the eigenvalues λ_i of the transfer operator T is only one part of the problem. The other part is: identifying the sets in (1) , such that the transition probabilities are small. We will use PCCA+[5, 8] in order to characterize the metastabilities of the dynamical system. PCCA+ is based on the following concept: A membership function $\chi : \mathbb{X} \to [0, 1]$ represents a kind of "fuzzy"-set, it assigns a grade of membership $\chi(x)$ to every state $x \in \mathbb{X}$ of the system. If $T\chi \approx \chi$, then this set is metastable with regard to the dynamics encoded in the transfer operator T. If we try to find a set of n membership functions $\chi_1, \ldots, \chi_n$ which are linearily independent and which are all as metastable as possible, then this can be written in terms of a variational principle and the functions χ are optimally given by a linear combination of the leading eigenfunctions of T. In the above algorithm we have already computed an approximation of these eigenfuntions. Given the n leading generalized eigenvectors $\bar{u}_1, \ldots, \bar{u}_n \in \mathbb{R}^m$ with $\tilde{P}\bar{u}_j = \lambda\tilde{M}\bar{u}_j$ according to the above algorithm, we construct approximated membership functions $\tilde{\chi}_1, \ldots, \tilde{\chi}_n$ by the linear combination approach in the space of ansatz functions Φ:

$$\tilde{\chi}_i(x) = \sum_{j=1}^{m} c_{ij}\Phi_j(x).$$

If the vectors $c_i = (c_{i1}, \ldots, c_{im})^T$ are non-negative and form a partition of unity, such that $\sum_{i=1}^{n} c_{ij} = 1$, then $\tilde{\chi}$ is non-negative and forms a partition of unity, because it is a convex combination of the ansatz functions. In order to compute the vectors c_i, we denote them as linear combination of the eigenvectors $\bar{u}_j$ via $c_{ik} = \sum_{j=1}^{n} a_{ij}\bar{u}_{jk}$. PCCA+ is used to find an optimal set of linear combination factors a_{ij}.[5, 8]

3.3 Constructing an initial discretization

In the beginning of the meshless approach an initial sampling of the state space $\mathbb{X}$ is needed. This sampling need not be trajectory-based, it even need not represent the invariant measure μ. Additionally, it need not be completly covering $\mathbb{X}$. In the numerical example, we will start with a very localized sampling only having discovered one out of three metastable states of the system. Out of this initial sampling $s_1, \ldots, s_I$, we will pick a predefined number of initial base points. The *picking algorithm* is done in the following way [7].

- In the first step an arbitrary base point q_1 is picked. Then the sampling point q_2 with the maximal distance to q_1 is picked.
- In the next steps, always the sampling point q_k is picked which has the maximal minimal distance to all given base points $q_1, \ldots, q_{k-1}$. That is, q_k is the point s out of the set of sampling points $s_1, \ldots, s_I$ which maximizes the following expression $\max_s \min_{j=1,\ldots,k-1; s \neq q_j} d(s - q_j)^2$.
- The last base point q_m picked in this way has the maximal minimal distance d_{min} to all other base points.

The basis functions Φ_k are now almost constructed; last we only need to define α_k. The more dense the base points are located, the higher the α-values should be, because the restraint simulations should be localized around their individual base points. In Theorem 5.10 in the doctoral thesis[22] it is shown that the error $\|\chi - \tilde{\chi}\|_{L^1(\mu)}$ between the true membership functions χ and the best approximation $\tilde{\chi}$ (based on the meshless discretization) is small only if α is sufficiently large. In order to assure a small overlap between restraint samplings for base points having a greater distance than $2d_{min}$, the scalar factors α_j are (equally) chosen to be proportional to d_{min}^{-2} (formula (80) in the doctoral thesis[22]). For a predefined parameter σ, we define $\alpha_j = \sigma / d_{min}^2$ for $j = 1, \ldots, m$. The question remains, how to sufficiently cover the relevant part of $\mathbb{X}$ with an initial disrectization. One could apply existing methods for a good initial sampling of $\mathbb{X}$ (ConCoord[4], GLAT[25], taboo search[1], or continuation methods[12]). Alternatively, the above picking algorithm could be used (and will be used in the numerical example) to "fill" $\mathbb{X}$: After we have constructed the basis functions Φ_k, we perform the restraint simulations according to the penalty potentials U_k. A major challenge in molecular simulation is that the trajectories of the system run into the next closest minimum, which is also known as trapping problem. This fact can hinder the system to explore additional parts of the state space $\mathbb{X}$. If we have picked a large number of base points, then α_k will be large and the penalty potentials diminishes a trapping. We simply restart the picking algorithm on the basis of all sampling points stemming from the restraint simulations and pick again the same number m of base points out of this set. If d_{min} has extended compared to the initial picking, then new parts of the state space have been found and we have to restart the restraint simulations and picking again. This iteration is done, until d_{min} does not extend anymore. The result of this procedure is our initial basis for the refinement steps.

3.4 Solving the Generalized Eigenvalue Probem & Refinement

The general eigenvalue problem (6) in general is ill-conditioned, especially, when solving it by inverting $\tilde{M}$. We, therefore, tread a different path by only computing the leading eigenvalues of $\tilde{P}$ i.e.

$$\tilde{P}\bar{u}_{P,i} = \lambda_{P,i}\bar{u}_{P,i}. \tag{13}$$

This eigenvalue problem typically is well-conditioned. Using PCCA+ the coefficients c_j, $j = 1, ..., n$ are computed as a linear combination of the leading eigenvectors of (13), i.e., $c_{ik} = \sum_{j=1}^{n} a_{ij}\bar{u}_{P,jk}$, which we can also denote using the linear space spanned by the leading eigenvectors, $c_j \in \text{span}(\{\bar{u}_{P,i}\}), \quad j = 1, ..., n$. If

$$\tilde{M}c_j \in \text{span}(\{\bar{u}_{P,i}\}), \quad j = 1, ..., n. \tag{14}$$

then the eigen vectors $(\bar{u}_i)$ of (6) meet

$$c_j \in \text{span}(\{\bar{u}_i\}), \quad j = 1, ..., n,$$

which is equivalent to

$$\tilde{M}c_j - \Pi_{\mu}^{\bar{u}_j}\left(\tilde{M}c_j\right) = 0, \quad j = 1, ..., n.$$

In this equation $\Pi_{\mu}^{\bar{u}_j}$ is the μ weighted orthogonal projection onto the eigenvector $\bar{u}_j$ which in its matrix form can be denoted by $\Pi_{\mu}^{\bar{u}_j} = \bar{u}\bar{u}^T W$ with $W = \text{diag}(\hat{\mu}_k)$. Summing up we have the following algorithm for solving the generalized eigenvalue problem (6):

1. Compute the n leading eigenvectors $\{\bar{u}_{P,i}\}$ of (13).
2. Evaluate the coefficients c_j, $j = 1, ..., n$ via PCCA+.
3. Compute for each basis function Φ_k, $k = 1, ..., m$ the error

$$r_k = \sum_{i=1}^{n} |\Big(\tilde{M}c_i - \Pi_{\mu}^{\bar{u}_j}(\tilde{M}c_i)\Big)_k|.$$

4. For some given tolerance TOL:

$r_k < \text{TOL}$	break
$r_k \geq \text{TOL}$	add new basis function $(m \to m+1)$ and go to 1.

We remark, that (14) can be interpreted as a condition on the discretization, which guarantees that solving the generalized eigenvalue problem (6) is not ill-conditioned. More precisely if (14) holds, then the eigenvalues of (6) are given by $\lambda_i = \lambda_{P,i}/\lambda_{M,i}$ where $\lambda_{P,i}$ are the eigenvalues of $\tilde{P}$ and $\lambda_{M,i}$ are the eigenvalues of $\tilde{M}$, respectively. The algorithm can also be considered as an adaptive refinement strategy with indicator r_k. Within each basis function $\Phi_k, k = 1, ..., m$, we check if the restraint sampling points $(x_{1,k}, ..., x_{L,k})$ are distributed according to the local invariant measure μ_k. For this purpose we apply the Gelman-Rubin [9] criterion like described in literature [3]. One further refinement strategy[22] had been proposed based on a perturbation of the shape parameter α. Common of all refinement strategies is that they require to add new basis functions, which can be done in an efficient way as described in the next section.

3.5 Adding new base points

In order to refine the basis, m' new base points have to be added to the existing ones. We take the union set of all sampling points according to all basis functions which are identified to be refined by the above algorithm and call this union the *picking set*. For choosing the base points for the additional basis functions, we observe that sampling points with a large distance to the base points q_k have a large statistical weight due to the structure of U_k and the weight γ. Consequently, the selection of the additional base points out of the picking set is then done by the picking algorithm in Section 3.3: New base points are iteratively added having maximal minimal distance to *all* base points selected so far – including the initial base points. After adding the predefined number of new base points, the scalar parameters $\alpha_k, k = m+1, \ldots, m+m'$, are again determined according to the smallest distance (only for the *added* basis functions $\alpha_j = \sigma/d^2_{min}$). Thus, adding new basis functions only affects the normalization function γ and has no impact on the already sampled restraint potentials U_k. The only change is, that the computations of P and M in equations (11) and (12) have to be applied with regard to the new normalization function γ.

Since the distances between the base points are assumingly decreasing in each refinement step, α will increase. New simulations according to the restraints $U_k, k = m+1, \ldots, m+m'$, are performed. In order to assure optimal acceptance ratios, the step length of the MC sampling method is necessary. Usually, the step length will decrease leading to less "exploration" of the conformational space within the predefined number of simulated points. In this way the adding of new base points is a real refinement.

3.6 Exploring additional transitions

For performing the described meshless algorithm and the refinements it is not mandatory that there exists one simulated trajectory which covers all parts of the state space $\mathbb{X}$. Even the initial discretization can be based on unconnected samplings of different parts of the conformational space. The advantage of a meshless approach with global radial basis functions (instead of sets) is the assured construction of dense matrices $\tilde{P}$ and $\tilde{M}$. The "transition matrix" $\tilde{P}$ is dense and also accounts for d-proximity of the base points (cp. (7)). More precisely, the computed membership functions $\tilde{\chi}$ account for transitions where they occur and for d-proximity where transitions do not occur. The discrimination between dynamical and geometrical proximity is coded in the difference between $\tilde{M}$ and $\tilde{P}$, because $\tilde{M}$ only includes the pure proximity information. The extension based stopping criterion in Section 3.3 probably leads to resolve only main transitions between the metastable states in the beginning. However, in the refinement procedure later on it is likely to find additional transitions between the identified conformations. The reason is the following. The added base points will be concentrated in those regions of the state space $\mathbb{X}$ where the transitions between the conformations occur. These regions are related to level sets of $\tilde{\chi}_i \approx 0.5$. If the refinement is perfomed "completely along" these level sets, then all transitions between the corresponding conformations are identified in the end. In the illustrative example we will show, that the refinement procedure identifies a transition region between two conformations which was not explored in the beginning.

4 Numerical Experiments

4.1 Illustrative Example

As an illustrative example the following potential energy function is analyzed:

$$\begin{aligned} V(x) \quad = \quad & 3\,\exp\big(-x_1^2-(x_2-\frac{1}{3})^2\big) \\ & -3\,\exp\big(-x_1^2-(x_2-\frac{5}{3})^2\big) \\ & -5\,\exp(-(x_1-1)^2-x_2^2) \\ & -5\,\exp(-(x_1+1)^2-x_2^2) \\ & +0.2\,x_1^4+0.2\,(x_2-\frac{1}{3})^4. \end{aligned} \tag{15}$$

Contour lines of the potential energy function are plotted in Fig. 1. In the numerical experiments the inverse temperature is $\beta = 4$, which leads to very rare transitions between the three metastable regions of the potential energy function. On average only every 100,000th step of the selected Langevin dynamics overcomes energetic barriers and is able to leave one of the major metastable regions. Long-term trajectories are, thus, inefficient in order to sample transition probabilities between the three wells, because the transitions occur rarely. There exists a eigenvalue solution based on the infinitesimal generator of that potential (Eq. (5.3) in[23]). From this solution we get that $\lambda_2 \approx 0.99998$ and $\lambda_3 \approx 0.997$.

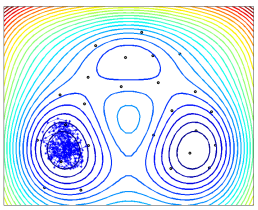

Figure 1: Contour lines of a three-well potential with two symmetric metastabilities (on the left and on the right) and one less metastable region (top). The blue trajectory shows an initial sampling of the state space $\mathbb{X}$ starting in the left metastable region. 500 steps of a discretized Langevin dynamics have been computed. This number is too low to expect a transition into another metastable region. However, by the restricted samplings, one transition path from the left (via the top) to the right metastable region is detected. The black circles correspond to the intial nodes.

Two different settings. In this first example we investigate the insertion of new basis functions as described in Section 3.5. More precisely, by setting $\alpha_j = \sigma/d_{\min}$ and using different values for σ we control the influence of new

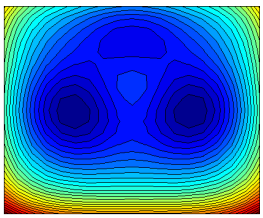

Figure 2: Contour lines of a three-well potential with two symmetric metastabilities (on the left and on the right) and one less metastable region (top).

basis functions onto the shape of all basis functions. By setting firstly $\sigma = 1$, the shape parameter α increases moderately, such that a very deep refinement of the state space is realized, see Fig. 3 and Fig. 5. Secondly we chose $\sigma = 3$ such that α increases rapidly, and only a few refinement steps are performed. Below the two choices of σ are discussed based on the results of the rough discretization.

Initial discretization for $\sigma = 1$. The intitial discretization of the state space $\mathbb{X}$ has been done according to the above algorithm in Sec. 3.3. For an initial sampling, we start 500 steps of a diffusion process as described in eq. (3) in the left minimum (cf. Fig. 1). In the next step 25 initial base points are selected out of this initial sampling. For each of the corresponding basis functions we compute the restraint subsampling points according to the restraint potential $U_k, k = 1, .., 25$. We then employ the picking algorithm (Sec 3.3) and subsample the new basis functions. Repeating this procedure enables us to cover three minima with an inital discretization. This corresponds to the black points in Fig. 1.

Refinement for $\sigma = 1$. If we apply the refinement procedure in Sec. 3.4 determining the basis functions that have to be refined and if we interpret the potential as a 2-metastabilities-situation, then mainly the region which directly connects the two deep minima is refined, see Fig. 3. In this region the membership functions have the largest gradients and need to be represented by a finer "grid" of meshless basis functions. If we plot circles with the r_k-values as radii and base points as centers in conformational space (Fig.4), we can directly see that in this region the mandatory subspace argument for the well-conditioned computation of λ_i via $\lambda_i = \lambda_{P,i}/\lambda_{M,i}$ is violated at most.

Results for $\sigma = 1$. In the above example, the potential energy function is regarded as a 2-metastabilities-situation with a direct transition between the two deep minima and a (lower minimum) transition region at the top. The refinement procedure stopped, when $\alpha > 60$. The computation of the second largest eigenvalue according to $\lambda_i = \lambda_{P,i}/\lambda_{M,i}$ yields $\lambda_2 = 0.99996$. Both transition pathways between the main metastabilities have been detected. However, the potential energy function can be seen as a 3-metastabilities-situation, too. In that case we aim at computing 3 generalized eigenvectors instead of 2. Let us rephrase, that condition (14) is fundamental for a well-conditioned computation of leading eigenvalues of T. For instance, if we would try to compute the third eigenvalue based only on the refinement in Fig. 3, then the approximation yields $\lambda_3 = 1.0237$ which is obviously wrong, because T can not have eigenvalues

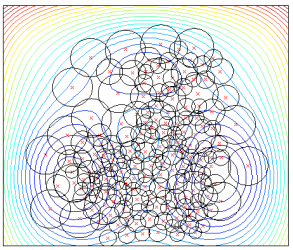

Figure 3: In this plot we try to identify 2 metastabilities $n = 2$. Mainly the direct transition region between the deep minima is refined. The red crosses represent the selected nodes, while the black circles represent the "size" or variance of the penalty potentials. The smaller the circles, the higher α.

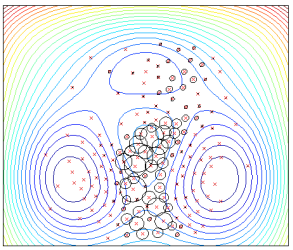

Figure 4: The error evaluation for the 2-metastabilities-situation. Bigger black circles have higher error values r_k. Clearly, the subspace condition is mostly violated in the direct transition region between the two deep minima.

greater than 1. Thus, the target invariant subspace condition of the refinement procedure is mandatory.

If we apply the refinement procedure for 3 eigenvectors, then the refinement will not only take place in the region where the two main metastable regions meet each other, it takes place inside the small top basin, too. This can be seen in Fig. 5. The resulting leading eigenvalues are $\lambda_2 = 0.99994$ and $\lambda_3 = 0.99695$.

Results for $\sigma = 3$. Using a higher value for σ leads to a faster increase of α. In the initial phase of the algorithm, a higher value of α leads to an exploration of regions with an higher energy. Thus, both transition pathways between the main metastabilities are already included in the initial discretization. The final discretization is already found after two refinement steps, see Fig. 6. Only near the direct transition region between the two main metastabilities additional base points were located. The resulting eigenvalues of the approximated Galerkin projection are $\lambda_2 = 0.99991$ and $\lambda_3 = 0.99733$. This is already a very good approximation compared to the very fine discretizations. In the numerical experiments, the second eigenvalue mainly increases during the refinement, whereas, the third eigenvalue decreases. There is a reason for this. On the one hand, there is the (Rayleigh's) variational principle [11]. The better the discretization can resolve the leading eigenfunctions, the higher the

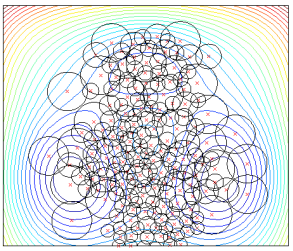

Figure 5: If the refinement procedure is based on 2 metastabilities, then mainly the direct transition region between the deep minima is refined. If the refinement, however, is based on 3 metastabilities, then the top region of the state space $\mathbb{X}$ is refined, too. This is shown here.

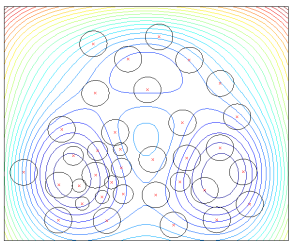

Figure 6: If the parameter σ is high, then α increases rapidly. The algorithm stops after only a few refinement steps of the discretiation. The meshless basis functions are indicated in this plot. A fast convergence could lead to unrevealed transition regions. However, the probability to find different transition pathways in the initial discretization phase is high for $\sigma = 3$.

eigenvalues are of the corresponding Galerkin projection of the transfer operator. This could explain, why λ_2 increases with a finer dicretization. On the other hand, the transfer operator is not analytically given, it is represented by a sampling procedure. This means, as long as not all transition pathways are found, the processes seem to be "more metastable" as they are. This could explain, that a certain level of discretization has to be reached in order to account for all possible transition pathways in the system. This means, that eigenvalues can decrease with a finer discretization, seemingly contradicting the variational principle .

4.2 Molecular Example

The second example is a nine-dimensional molecular sytem. In this system 3 Argon atoms (in three dimensional space) are interacting with each other. The interaction potential is a Lennard-Jones potential

$$V_{LJ}(r) = \epsilon\Big(\Big(\frac{r_0}{r}\Big)^{12} - 2\Big(\frac{r_0}{r}\Big)^{6}\Big),$$

which depends on the pairwise distance r between the Argon atoms. In our example the LJ-energy is $\epsilon = 94.87[kJ/mol]$ and the optimal distance between the Argon atoms is $r_0 = 3.74$[Å]. Additionally to the LJ-potential there is a quadratic penalty potential to restrict each atom coordinate to be in the intervall [-2Å,2Å]. The inverse temperature has been set to $\beta = 0.1$[mol/kJ] in order to have rare transition events. Like in the above example, we set $\sigma = 3$ for the adaptive calculation of the parameter α of the modified potential. We set the trial step size to be 0.03[Å] for the Metropolis Hastings sampling (MC) of the inner products. This assures for an acceptance ratio of about 80% during our simulations. For the refinement we combine the invariant subspace condition of Sec. 3.4 with a standard Gelman-Rubin criterion. The samplings should at least converge. This is reached, if the Gelman-Rubin-indicator is less than 1.3.

We start our simulation with a conformation, where all pair-wise distances between the Argon atoms are optimal, i.e., $r = r_0$. The transfer of the molecular process is given by a diffusion process where the drift is based on the gradient of the LJ-potential, see eq. (3).

This system has four different metastable conformations, see Fig.7. The first one "1-2-3" is the main conformation, where all distances between the atoms are nearly r_0. In the other three conformations one atom is unbound and two atoms form a pair; these symmetric conformations can be denoted by "1-2 3, 1 2-3, 1-3 2". Because of this, we refer to the main conformation as to the doubly bound one and call the other three the singly bound conformations.

The given molecular example has three difficulties. First, the starting point of the simulation 1-2-3 has a very low potential energy such that the transition from that conformation to another conformation is very unlikely. Second, the transitions between the conformations of the system are not only driven by the potential energy term, they are also driven by entropic contributions, because the unbound atom has to "find" its binding partners. Thus, there is no saddle point in the potential energy landscape. And third, the three conformations "1-2 3, 1 2-3, 1-3 2" are equivalent. Thus, the corresponding transition probabilities should be equal.

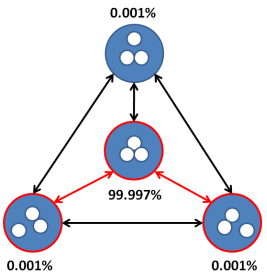

Figure 7: Statistical weights of the conformations of Argon. The red colored circles and arrows are the conformations and transitions that have been found in the initial discretization, i.e. the first row in Table 1

In the initial discretization phase two singly bound conformations are already covered. Furthermore, the transitions between these conformations and the

doubly bound conformation are revealed. Analyzing the transition matrix after that phase reveals only 2 further eigenvalues close to $\lambda_1 = 1$ (the system should have three). The refinement of the conformational space mainly takes part in the entropic region. The transitions between the singly bound states are discovered at a late stage of the refinement. In Table 1 the results of the initial level and the three refinement levels are shown. For measuring the similarity between the subspaces spanned by $\bar{u}_P$ and $M\bar{u}_P$ we computed the angle between the subspaces [10].

After four refinement steps with 65 base points in total, the analysis of the transition matrix reveals all four conformations, i.e., all transitions indicated in Fig. 7.

refinement level	# steps	# propagated	# base points	subspace	$\lambda_{P,2}, \lambda_{P,3}, \lambda_{P,4}$	# cluster
0	37,500	7,500	25	0.031	0.78, 0.77,0.68	3
1	21,000	4,200	14	0.029	0.78, 0.77,0.74	3
2	18,000	3,600	12	0.0602	0.78, 0.78,0.74	4
3	21,000	4,200	14	0.0394	0.79, 0.78,0.74	4
total	135,000	19,500	65			

Table 1: Showing the refinements steps of the procedure explained in Section 3.4. Inititally we started the picking procedure with 37,500 steps (3 iterations of the picking algorithm). The number of conformations (clusters) has been determined with the minChi criterion[24] and increases during the procedure. The angle between the subspaces of the eigenvectors $\bar{u}_P$ and $M\bar{u}_P$ increases within the refinement steps (whenever the number of clusters is unchanged). This shows that the idea of the refinement indicator applies.

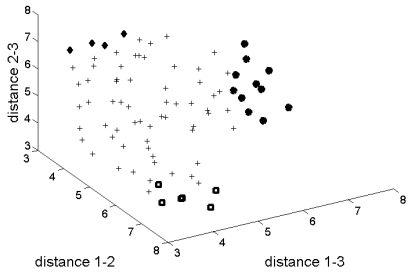

Figure 8: Resulting clustering of the state space (PCCA+-based assignment of the base points to their clusters). The nine-dimensional space is projected on the three pair-wise atom distances. One can see the clustering of the space into the desired four conformations, as indicated by different symbols.

Although the discretization is far from being symmetric, the calculated transition probabilities between the conformations are symmetric: The conditional

transition probabilities from the minor conformations into the main conformation are in the range of $7\% - 8\%$. Thus, the maximal absolute deviation from the symmetry is 1%. The eigenvalues λ_2, λ_3, and λ_4 of M are near 0.84, while the corresponding eigenvalues of P are near 0.77. This indicates a relatively good mixing and a good overlap between the meshless basis functions. Although in the meshless approach the eigenvalues are low, the computation of the eigenvalues of $\tilde{M}^{-1}\tilde{P}$ and of the statistical weights reveals, that the main conformation covers about 99.997% of the conformational space, and transitions from that doubly bound conformation into one of the singly bound ones on average only take place every $30,000th$ time step, if an unrestricted sampling of the diffusion dynamics (3) is applied instead of our meshless discretization approach. In total, we sampled $154,500$ steps for the whole procedure and figured out the statistical weights and all transition probabilities to correct symmetry.

For comparison, we performed a straight forward simulation without any restraints of the dynamics with $300,000$ steps (same stepsize as above). An analysis of the trajectory with PCCA+ classifies the atomic distance of 4[Å] to be "unbound", whereas, in the meshless approach it was 7[Å]. The reason for the difference is that the farest distance between the atoms during the simulation has only been 5[Å]. This explains also why the main conformation had a lower statistical weight in the straight forward simulation: $299,840$ steps were found in the main conformation, which are 99.94% of all steps. In total, the trajectory changed 4 times between the conformations, and only 2 of the 3 unbound conformations were visited. A straight simulation giving the correct symmetry and reaching the conformations where the atoms have a distance of more than 7[Å] seems computationally extremely demanding.

5 Conclusion

In this article, we presented a set-free approach to MSM building utilizing smooth overlapping ansatz functions instead of sets. In addition, an adaptive refinement procedure of this kind of meshless discretization was proposed that allows to improve the quality of the model while exploring state space wile in parallel improving the approximation quality of the MSM by inserting new ansatz functions into the discretization.

Furthermore, the choice of Gaussian-like ansatz functions makes it easy to implement the Galerkin projections underlying the discretization by means of existing sampling software via quadratic penalty potentials. We combined this easily available restraint sampling procedure with a reweighting strategy such that the ansatz functions are approximate indicator functions of sets such that the popular interpretation of the resulting MSM in terms of transition probabilities can again be used.

A generalized eigenvalue problem has to be solved for the resulting non-orthogonal Galerkin projection. In order to avoid ill-conditioned eigenvalue problems, we proposed to adaptively refine the ansatz space such that the eigenvectors of the "transition matrix" P are also invariant with regard to the "mass matrix" M. It turned out, that this approach explores the transition regions of the conformational space, and can reveal conformations of the molecular system which were undiscovered by the initial sampling.

Acknowledgement. This research has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems", Project (A05) "Probing Scales in Equilibrated Systems by Optimal Nonequilibrium Forcing".

References

[1] R. Battiti and G. Tecchiolli. The reactive tabu search. *ORSA J. Comput.*, 6(2):12–140, 1994.

[2] 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.

[3] A. Bujotzek, O. Schütt, A. Nielsen, K. Fackeldey, and M. Weber. ZIBgrid-free: Efficient conformational analysis by partition-of-unity coupling. *J. Math. Chem.*, 52:781–804, 2014.

[4] B.L. de Groot, D.M.F. van Aalten, R.M. Scheek, A. Amadei, G. Vriend, and H.J.C. Berendsen. Prediction of protein conformational freedom from distance constraints. *Proteins*, 29:240–251, 1997.

[5] 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.

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

[7] V. Durmaz. *Free energy calculations of host-guest systems*. PhD thesis, Free University Berlin, 2016.

[8] K. Fackeldey, S. Röblitz, O. Scharkoi, and M. Weber. Soft versus hard metastable conformations in molecular simulations. In E. Onate and D.R.J. Owen, editors, *Particle Methods II, Fundamentals and Applications*, pages 899–909. Barcelona, Spain 26-28 Oct. 2011, 2011.

[9] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[10] G. H. Golub and C.F. Van Loan. *Matrix Computation, chapter 6.4.3*. John Hopkins University Press, 2013.

[11] R.A. Horn and C.A. Johnson. *Matrix Analysis*. Cambridge University Press, 2013.

[12] J. J. Moré and Z. Wu. Issues in Large-Scale Global Molecular Optimization. In *Large-Scale Optimization with Applications*, pages 99–121. Springer, 1997.

[13] Feliks Nüske, Bettina G. Keller, Guillermo Pérez-Hernández, Antonia Mey, and F. Noé. Variational approach to molecular kinetics. *J. Chem. Theory Comput.*, 10:1739–1752, 2014.

[14] G. Perez-Hernandez, F. Paul, T. Giorgino, G. de Fabritiis, and Frank Noe. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.*, 139:015102, 2013.

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

[16] 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.

[17] C. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden. Markov state models based on milestoning. *J. Chem. Phys.*, 134:204105, 2011.

[18] C. Schütte and M. Sarich. A critical appraisal of Markov state models. *The European Physical Journal Special Topics*, 224(12):2445–2462, 2015.

[19] Ch. Schütte and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In *Handbook of Numerical Analysis*, pages 699–744. Elsevier, 2003.

[20] 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.

[21] E. Vanden-Eijnden and M. Venturoli. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.*, 130(19):194101, 2009.

[22] M. Weber. *Meshless Methods in conformation dynamics*. PhD thesis, FU Berlin, 2006.

[23] M. Weber. *A Subspace Approach to Molecular Markov State Models via a New Infinitesimal Generator*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2011.

[24] M. Weber, W. Rungsarityotin, and A. Schliep. An Indicator for the Number of Clusters Using a Linear Map to Simplex Structure. In *From Data and Information Analysis to Knowledge Engineering, 29th Annual Conference of the German Classification Society 2005, March 9-11, Studies in Classification, Data Analysis, and Knowledge*, pages 103–110. Springer, Heidelberg, 2006.

[25] F. Wende, F. Cordes, and T. Steinke. On Improving the Performance of Multi-threaded CUDA Applications with Concurrent Kernel Execution by Kernel Reordering. In *Proceedings of the 2012 Symposium on Application Accelerators in High Performance Computing, SAAHPC 12, Washington, DC, USA*, pages 74–83, 2012.