Meshless Methods in Conformation Dynamics

Dissertation
am Fachbereich Mathematik und Informatik
der Freien Universität Berlin
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften

vorgelegt von

Diplom-Mathematiker Marcus Weber

Berlin, Februar 2006

Meinem Großvater Heinrich Watermeyer gewidmet

Contents

1 Introduction

1.1 Motivation

The title of this thesis consists of the two parts "meshless methods" and "conformation dynamics". In conformation dynamics one is interested in the identification of metastable molecular conformations and the transition probabilities between them. For a better understanding, let me explain the basic concepts of classical geometrical conformation analysis and conformation dynamics, before showing the necessity of meshless methods in this application area.

Geometrical conformation analysis. A molecule is defined by its n atoms, their types and bonds. *Position states* q of a molecule are defined as a certain arrangement of the atoms in configuration space, i.e. a $3n$-dimensional vector of atom coordinates $q \in \mathbb{R}^{3n} = \Omega$. Different position states q also lead to different chemical properties. The algorithmic framework of software packages based on local minimization[1] of potential energy for geometrical conformation analysis is explained in Leach [77]. Although these algorithms are very fast, there are some problems connected to this kind of approach.

- Unfortunately, the potential energy landscape $V : \Omega \to \mathbb{R}$ of e.g. drug-like molecules is extremely rough including many local minima $q \in \Omega$ with similar structure and properties. A listing of all these local minima does not contain any valuable information.

- In order to filter some important and dissimilar position states out of the set of local minima, a geometry based cluster algorithm is necessary. But chemical similarity is not connected to the euclidian distance measure in Ω, which is often the base of clustering in commercial software.

Instead of the euclidian distance, in a more realistic model, the dissimilarity of two position states corresponds to the height of the free energy barrier between them, i.e. the possibility of one conformation to reach the catchment area of the other conformation via molecular dynamics in a heat bath (this will lead to definition (12) for *metastable* conformations). Again, there is a couple of slower algorithms [6, 74, 12], which find saddle points (or an ensemble of transition paths) between two local minima of the energy landscape in order to define the energy barrier i.e. the dissimilarity between the minima. However, these

---

[1]Classical molecular conformation analysis in literature is often understood as a search for local minima of a potential energy surface $V : \Omega \to \mathbb{R}$ in order to examine the variety and the importance of these position states. Therefore, computational chemistry software like Cerius[2] and Catalyst by Accelrys Software Inc., Concord and Corina by Molecular Networks GmbH Computerchemie, FANTOM by the Seal Center for Structural Biology of the University of Texas and many others, often include a local minimization routine combined with a sophisticated method, which generates suitable starting points for an exhaustive search in Ω.

algorithms may fail, if the potential energy surface is like the solid line in Figure 1(a): The conformations may consist of multiple local minima separated by transition regions, which also contain many local minima and saddle points. Note that Fig. 1(a) is only a one dimensional example, whereas in higher dimensions there usually exists a very complex network of minima and transition regions.

Thermodynamical ensemble.   At a certain temperature T the probability for the occurrence of a position state $q \in \Omega$ in the so-called *canonical ensemble* is proportional to the Boltzmann expression $\exp(-\beta V(q))$, where $\beta \propto T^{-1}$ is the corresponding Boltzmann factor, see also Fig. 1(a). For $T \to 0$ only the global minimum $q \in \Omega$ of V has a significant density in the ensemble, for $T \to \infty$ the position states are more and more equally distributed. The total density function can be seen as a linear combination of n_C partial densities, which are defined in different regions of the potential energy surface of the molecule separated via high-energy barriers, see Fig. 1(b) dashed and dotted line. The thermodynamical weight $\widetilde{w}_i$ of such a *metastable conformation* $i \in \{1, \ldots, n_C\}$ is given by the probability for the molecule to be in conformation i, i.e. the value of the integral over the corresponding partial density function. In the point concept and in commercial software for classical conformation analysis, the weights of the conformations are estimated by the difference of the potential energy values at the corresponding local minima. As one can see in Fig. 1(b), the correlation between the energy value at a local minimum and the thermodynamical weight $\widetilde{w}_i$ may be misleading: Although the right local minimum is deeper, the "left" conformation has a higher thermodynamical weight, because it is broader. For all these reasons, we now leave the classical point-concept algorithms in the present work. Unfortunately, we also leave the possibility to use fast local minimization routines and have to apply extensive algorithms for a thermodynamically correct exploration of the conformational space Ω, like Markov chain Monte Carlo sampling [101] combined with umbrella strategies [123, 124, 127] or other biased sampling methods [88, 9, 84, 10, 35, 129, 113, 65, 1]. In the present thesis, an algorithm is presented which combines the sampling of position states via umbrella strategies and the corresponding analysis of conformations in the sense of Fig. 1(b) and (c).

Complexity reduction.   A direct approach to conformation analysis with a correct thermodynamical weight computation can be done, for example, by an approximation of the partial density functions in Fig. 1(b). Examples for this approach are the *probability density estimation* [53, 119] and the *mixture model computation* [21, 86, 108]. By using this kind of algorithms, the complex shape of the partial density functions connected to the rough energy landscape V becomes displeasing. The first step of model reduction is to apply a clustering approach. As mentioned above, the definition of similarity has to take an examination of the dynamical behavior of the molecule into account. There are a couple of algorithms for clustering on the basis of molecular data and molecular

(a) Potential energy (—) and corresponding Boltzmann density (−−)

(b) Splitting of density into two conformations (−−) and (···)

(c) Complexity reduction via membership functions (−−) and (···)
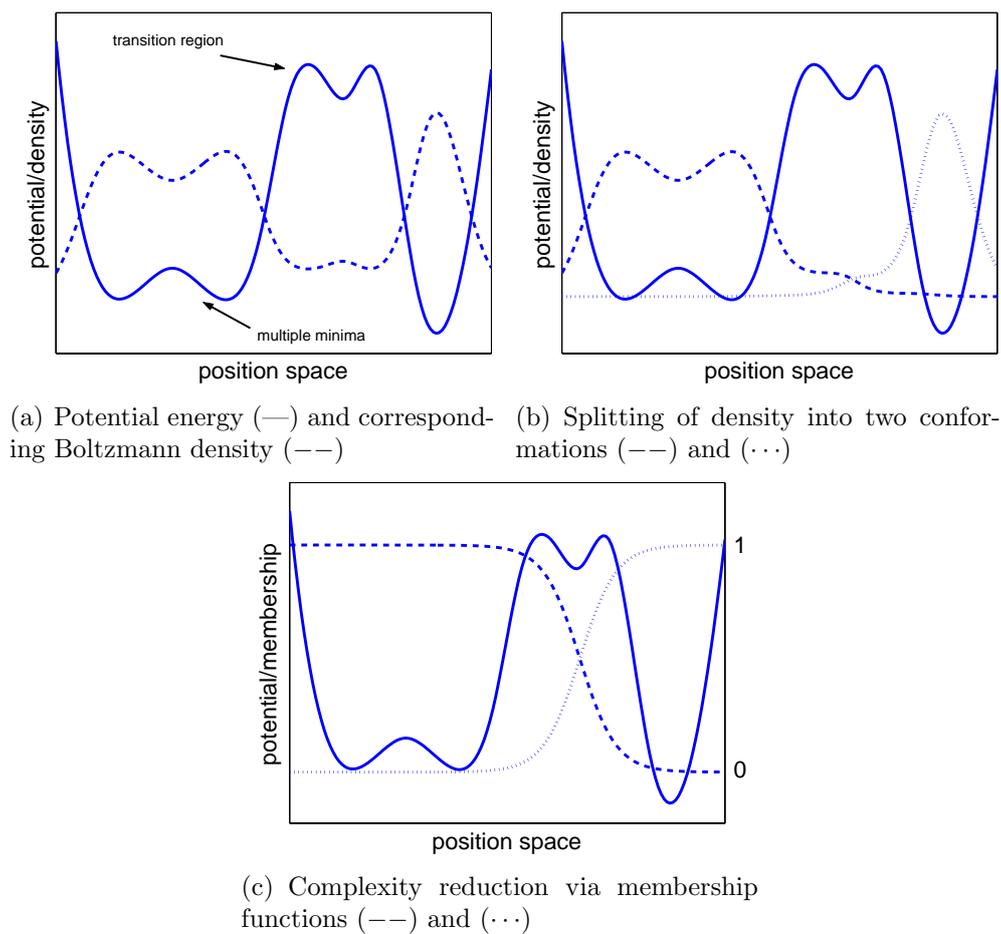
Figure 1: Example for a potential energy function together with (a) its total Boltzmann density of the spacial states and (b) a splitting of the Boltzmann density into two parts, which are separated by an high potential energy barrier. Complexity reduction (c) via membership functions instead of approximating the partial densities themselves.

dynamics time series. However, algorithms which indirectly use a mixture model for
the decomposition of the conformational space into different clusters like *mixture model
clustering* [29, 85] have to contain enough basis functions for the approximation of the
partial density functions, which may lead to a "curse of dimensionality". Also *Hidden
Markov Models* [102] are based upon certain assumptions about the shape of the partial
density functions.

Deuflhard, Schütte et al. [27, 24, 28, 111, 26] invented the set-based approach to
conformation dynamics, which was inspired by ideas of Dellnitz et al. [18, 20]. In the
present text we change the point of view from their set-based concept to a function-based
clustering. For the clustering approach of Fig. 1(c) the Robust Perron Cluster Analysis
with a meshless function basis is used, which does not make any assumptions about the
shape of the density functions in Fig. 1(b). The total density function in Fig. 1(a), dashed
line, is a positive sum of partial densities. For each conformation $i \in \{1, \ldots, n_C\}$ there is
a membership function $\chi_i : \Omega \to [0, 1]$, which determines the portion of the partial density
w.r.t. the total density function. In order to sum up to the total density, these membership
functions χ_i form a partition of unity. In Fig. 1(c) the corresponding functions are shown
for the example from Fig. 1(b). Since local minima of a complex transition region cannot
be assigned to one single conformation, these membership functions are continuous and
overlapping. The problem of approximation of these sigmoidal membership functions is
much easier than the approximation of the partial densities themselves and can be done
with a small number s of basis functions. The algorithms for clustering are based on
spectral methods proposed by Dellnitz and Junge [20]. A good example for complexity
reduction like the one shown in Fig. 1(c) is epigallocatechin on page 99, which has two
conformations with multiple minima. A good example for a molecule with a pronounced
transition region is cyclohexane on page 94. These examples show, that the number s of
basis functions can be much lower than the number of local minima of the molecule.

Meshless methods. Conformation analysis in the view of Fig. 1(c), dashed and dotted
line, is a function approximation problem in a high dimensional space. We do not have
to know the function values exactly for each point q in Ω, because most position states
are physically irrelevant due to a high potential energy value $V(q)$ and a low Boltzmann
factor $\exp(-\beta V(q))$. Mathematically, a Boltzmann-weighted function approximation er-
ror is suitable for this kind of problem, see Section 5.2.1 equation (83). For weighted
function approximation many theoretical results have shown that only randomized al-
gorithms or meshless particle based methods might break the "curse of dimensionality"
[105, 98, 80, 79]. The notation "meshless methods" in this context only means that it
is computationally impossible to create a compactly supported structured function basis
in high-dimensional spaces in order to approximate χ, e.g. triangulation or regular grids,
because the number of basis functions used for the approximation of χ in mesh-based
methods increases exponentially with increasing dimension of Ω. In the present work, we

therefore use a meshless approach to conformation analysis, see Section 5.1.

Conformation dynamics.   Due to the exhaustive sampling in Ω not only thermody-
namical weights $\widetilde{w}_i, i = 1, \ldots, n_C$ of the conformations are available, but also the dynam-
ical behavior between these conformations. Conformation dynamics is directly connected
with the definition of dissimilarity in the clustering step of the algorithm, because con-
formations should be clustered in a way that molecular dynamics is unlikely to transcend
between different conformations. Therefore, in the present work we aim at a maximiza-
tion of metastability (i.e. minimization of transitions between conformations) in Robust
Perron Cluster Analysis, see Section 3.4.2.

1.2   Outline

The thesis is organized as follows: In Section 2 a sampling method for the computation
of observables in configuration space and the transfer operator approach to conformation
dynamics is derived. In Section 3 Robust Perron Cluster Analysis is explained, which is
the most important tool for the meshless identification of molecular conformations. This
analysis is directly based on the transfer operator approach. In Section 4 Robust Perron
Cluster Analysis is extended to a meshless discretization of the configuration space. A
meshless partitioning method is derived for the computation of the relevant matrices used
for the cluster analysis. In Section 5 the algorithmic details of the corresponding code
ZIBgridfree [94] are designed. An error analysis is included, which balances the generation
of sampling points in local minima and transition regions of Ω in a problem adapted way.
Finally, in Section 6 the "proof of concept" is also shown numerically.

2   Physical and Statistical Models

2.1   Molecular Dynamics

For an algorithmic introduction into molecular dynamics see also Allen and Tildesley [2] Chapter 3.

The configuration space of a molecular system having n atoms can be expressed by $3n$ Cartesian coordinates $q \in \Omega = \mathbb{R}^{3n}$. For a classical description of molecular motion with conservative forces, Newton's 2nd law is given by:

$$\mathbf{M}\ddot{q} = f(q).$$

In this equation $\mathbf{M} \in \mathbb{R}^{3n \times 3n}$ is a diagonal matrix of atomic masses and $f(q) \in \mathbb{R}^{3n}$ is a vector of internal and external forces acting on the atoms of a position state q. f in general is the result of an averaging process: Internal forces arise from an averaged electronic configuration of the molecular system due to the well-known Born-Oppenheimer approximation. External forces for example arise from the statistical effects of the not explicitly modeled part of the solvent. The computation of f is the most time-consuming procedure in numerical routines. In the present text, molecular systems only consist of a single molecule in vacuum, i.e. without modeling external forces. For more realistic results the dynamics simulation has to be extended accordingly. In most of the applications, the internal forces are expressed by the negative gradient of an energy function $V : \Omega \to \mathbb{R}$,

$$f(q) = -\nabla V(q),$$

where $V \in C^1(\Omega)$ is split into additive components like covalent interactions and non-covalent interactions. The correct parameterization of different analytical forms of V is a very important task in molecular dynamics simulation, because V is the link between numerical methods and the model of a realistic molecule. The parameterization and invention of so-called *molecular force fields* is a rapidly changing and profitable market. For an excellent introduction into molecular modeling and molecular force fields see Leach [77]. In our case, for the numerical routines in Section 6 the Merck Molecular Force Field [50, 51] has been implemented, which is suitable for drug-like molecules. With these preparations the *Hamiltonian form* of the equation of motion is

$$\begin{aligned} \dot{q} &= \mathbf{M}^{-1}p \\ \dot{p} &= -\nabla V(q), \end{aligned} \tag{1}$$

where $p \in \mathbb{R}^{3n}$ are the momenta coordinates of the molecule. For given initial momenta $p(0)$ and spatial coordinates $q(0)$, the equation of motion (1) uniquely defines a trajectory $x(t) = (q(t), p(t)), t \in (-\infty, \infty)$, in the state space $x \in \Omega \otimes \mathbb{R}^{3n}$. The *Hamiltonian* or *total energy* of a molecular state $x = (q, p)$ is given by the expression

$$H(q, p) = \underbrace{\frac{1}{2}p^\top \mathbf{M}^{-1} p}_{=K(p)} + V(q),$$

where $K(p)$ is the kinetic energy part. Computing the time derivative of the total energy and inserting into (1) implies that for a molecular dynamics trajectory the total energy $H(q(t), p(t)) \equiv H(q(0), p(0))$ is constant for $t \in (-\infty, \infty)$.

Reversibility. A substitution of variables $p \to -p$ and $t \to -t$ leaves the equations of motion unchanged. I.e. if an initial state (q, p) is transferred to $(\widetilde{q}, \widetilde{p})$ in time span τ, then the initial state $(\widetilde{q}, -\widetilde{p})$ is transferred to $(q, -p)$ in time span τ, which is an often used result in the following.

Symplectic flow. Let Φ^τ denote a solution of (1) for a finite time span τ and initial conditions $(q(0), p(0))$:
$$(q(\tau), p(\tau)) = \Phi^\tau(q(0), p(0)).$$
In terms of the Hamiltonian $H(x)$ and the state vector $x = (q, p)$ the molecular dynamics equation can also be written as
$$\dot{x} = J \cdot \nabla_x H(x)$$
with the skew-symmetric matrix
$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$
where I is the $3n$-dimensional unit matrix. (1) has a *symplectic* structure. For a detailed mathematical analysis of the symplectic structure of molecular dynamics see Hofer and Zehnder [57]. The symplectic property is used in the proof of self-adjointness of a transfer operator corresponding to the Hamiltonian flow in Theorem 2.2. Symplecticity is defined as conservation law
$$(D\Phi^\tau)^\top \cdot J \cdot (D\Phi^\tau) = J,$$
for all τ and the Jacobian $D\Phi^\tau$ of the Hamiltonian flow. Many other conservation laws can be derived from the symplectic property. The area-preserving property can schematically be shown via
$$\det((D\Phi^\tau)^\top \cdot J \cdot (D\Phi^\tau)) = \det(J) \Rightarrow \det(D\Phi^\tau)^2 = 1 \Rightarrow |\det(D\Phi^\tau)| = 1.$$

Moreover, Mackey and Mackey [82] have collected some proofs that $\det(S) = 1$ for a symplectic matrix S. A symplectic numerical integrator Φ_h^τ, which is a discretization of Φ^τ with time-step h, meets
$$\left(\frac{\partial \Phi_h^h x(0)}{\partial x(0)}\right)^\top J \left(\frac{\partial \Phi_h^h x(0)}{\partial x(0)}\right) = J$$

for a one step computation and, therefore, shares the area-preserving property with the real flow Φ^τ. The invention of symplectic integrators is not in the scope of the present

context. The most famous example for a symplectic numerical integrator having constant step size h is the Störmer-Verlet integrator [125]. This integrator is also reversible. Both properties are essential for the construction of a sampling scheme in Section 2.3. For a constant step-size integrator the choice of h is determined by the shortest oscillation period of bonds in a molecule, which is only some femtoseconds.

Lyapunov instability.   Not only the time discretization h is sensitive. For molecular dynamics simulation, the total time-span τ has also to be chosen carefully. Assume an initial state $x(0)$ and a perturbed initial state $x^*(0)$. If these states are evolved by Φ^t, then the error $\|x(t) - x^*(t)\|$ exponentially depends on the length t of the trajectory:

$$\|x(t) - x^*(t)\| \sim \|x(0) - x^*(0)\| \, \exp(\lambda_{max} \, t),$$

where λ_{max} is the maximal Lyapunov characteristic exponent. Deuflhard et al. [25] have shown that for molecular dynamics λ_{max} may be very large, such that only trajectories of some 10^{-13} seconds are correlated with the initial state. Molecular dynamics as a point to point concept has a chaotic behavior. Therefore, we will change our point of view from ill-conditioned deterministic to statistical dynamics and investigate ensembles of molecular systems.

2.2   Ensemble and Boltzmann Distribution

Statistical weights inside an (n, v, T)-ensemble.   In an (n, v, T)-ensemble each sub-system, in our case a single molecule in vacuum, has the same macroscopic properties, volume v and temperature T. Furthermore, chemical reactions do not take place, which means that the number n of particles is also kept constant. The subsystems can only exchange energy with their surroundings and therefore have different states $x = (q, p)$, for which Boltzmann statistics can be applied. For a detailed derivation of the Boltzmann distribution, which is defined as the most probable distribution of states in an (n, v, T)-ensemble and is also called *canonical ensemble* μ_{can}, see Schäfer [106] pp. 5–11. Via the Hamiltonian $H(\cdot)$, the probability that a molecule attains the state x is

$$\mu_{can}(x) = \frac{1}{Z} \exp(-\beta \, H(x)),$$

where Z is a corresponding normalization. β is the inverse temperature

$$\beta = \frac{1}{k_B T},$$

where T is measured in Kelvin and $k_B \approx 1.38066 \cdot 10^{-23} J \, K^{-1}$ is the Boltzmann constant.

Partition functions. For separable Hamiltonian functions, the Boltzmann distribution of molecular states $x = (q, p)$ in the canonical ensemble $\mu_{can}(x) = \pi(q)\eta(p)$ has the following splitting into a distribution of momenta and position states:

$$\eta(p) = \frac{1}{Z_p} \exp(-\beta K(p)), \quad \pi(q) = \frac{1}{Z_q} \exp(-\beta V(q)).$$

$\eta(\cdot)$ represents a Gaussian distribution for each of the $3n$ momenta coordinates, because K is a quadratic function with a diagonal matrix $\mathbf{M}$. Therefore, the sampling[2] of momenta according to the Boltzmann distribution is simple, see e.g. [2]. The spacial factor $\pi(\cdot)$ is a more complex distribution function. Whereas the exponential function $\exp(-\beta V(q))$ can be computed pointwise, the unknown normalization constant (also denoted as *spatial partition function*[3]) is

$$Z_q = \int_\Omega \exp(-\beta V(q)) \, dq. \tag{2}$$

For a sampling of points $q \in \Omega$ according to a distribution, which is known except for a normalization constant, the Metropolis-Hastings algorithm can be applied, see Section 2.3.

Bracket notation. A macroscopic measurement is always carried out for a snapshot of a molecular ensemble, where molecular states are distributed according to the Boltzmann distribution. A spatial observable $\langle A \rangle_\pi$ of a function $A : \Omega \to \mathbb{R}$ in configuration space is therefore measured as an ensemble mean, i.e. an expectation value

$$\langle A \rangle_\pi = \int_\Omega A(q) \, \pi(q) \, dq, \tag{3}$$

where A is μ-Lebesgue integrable w.r.t. the measure $\mu(dq) := \pi(q) \, dq$. For the spatial Boltzmann distribution we define the weighted spaces

$$L^r(\pi) := \{u : \Omega \to \mathbb{R}, \; \int_\Omega |u(q)|^r \, \pi(q) \, dq < \infty\}, \quad r = 1, 2.$$

For $r = 2$ we get a Hilbert space with scalar product

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

The scalar product is used for a Galerkin discretization in Section 3.2 and in Section 4.1. In the following, these bracket notations for observable and inner products will be often used abbreviations.

---

[2] "sampling" means: creating a sequence of momenta numerically, such that their distribution converges against η.

[3] Z_q is a function of temperature, volume and number of particles of an ensemble. For the canonical ensemble Z_q is a constant. The total partition function $Z = Z_q Z_p$ is the key to calculating **all** macroscopic properties of the system.

Importance sampling. Solving the integral (3) numerically is an important task in computational chemistry. Deterministic quadrature formulas are not practical for high-dimensional spaces Ω, see Novak [98]. Therefore, one applies Monte Carlo integration methods with an importance sampling routine, i.e.

$$\langle A \rangle_\pi \approx \frac{1}{N} \sum_{i=1}^{N} A(q_i), \quad q_i \propto \pi, \tag{4}$$

where position states $q_i \in \Omega$ are sampled according to their Boltzmann distribution π. For introductory literature of Monte Carlo integration methods see [52, 103].

Molecular simulation. Generating a set of independent and identically distributed position states in (4) is not easy. Only a few points out of Ω are physically important, the most part of Ω has a high potential energy value V and is therefore not relevant. Instead of creating independent points q, in practice one starts with a physically relevant position state $q_1 \in \Omega$ and generates further states via a Markov chain

$$q_1 \rightarrow q_2 \rightarrow q_3 \rightarrow \ldots q_N. \tag{5}$$

This is called a *Monte Carlo Simulation.*

2.3   Hybrid Monte Carlo Method (HMC)

2.3.1   Requirements

Detailed balance. A sufficient condition for a correct sampling via Monte Carlo Simulation in the situatuion of (5) is the detailed balance condition with the desired Boltzmann distribution π. This condition holds, if the conditional probability density function $\mathbf{P}(q \rightarrow \widetilde{q})$ for a transition $q \rightarrow \widetilde{q}$ in (5) meets

$$\pi(q)\,\mathbf{P}(q \rightarrow \widetilde{q}) = \pi(\widetilde{q})\,\mathbf{P}(\widetilde{q} \rightarrow q). \tag{6}$$

In this case, the occurrence of a certain position state q in (5) is proportional to $\pi(q)$. For the sufficient and necessary "balance condition", which is less rigorous than (6), see Manousiouthakis and Deem [83].

Metropolis-Hastings algorithm. In the following we use a Metropolis-Hastings type algorithm [90], where the transition probability density function $\mathbf{P}$ in (6) is split into two factors

$$\mathbf{P}(q \rightarrow \widetilde{q}) = \mathbf{P}_{pr}(q \rightarrow \widetilde{q})\,\mathbf{P}_{ac}(q \rightarrow \widetilde{q}). \tag{7}$$

Here is the corresponding sampling scheme:

- In equation (7), $\mathbf{P}_{pr}$ is a proposal probability density function, i.e. the probability that after $q \in \Omega$ a position state $\widetilde{q} \in \Omega$ is proposed as candidate for the next step in the chain (5).

- With an acceptance probability of $\mathbf{P}_{ac}$ the next step in the chain is $\widetilde{q}$, with a probability of $1 - \mathbf{P}_{ac}$ the step q is repeated in (5). For a numerical realization of the acceptance probability, one computes a uniformly distributed random number $r \in [0, 1]$ and accepts $\widetilde{q}$ if $r \leq \mathbf{P}_{ac}(q \to \widetilde{q})$.

With equation (6) a sufficient condition for a correct sampling according to this scheme is:
$$\pi(q)\,\mathbf{P}_{pr}(q \to \widetilde{q})\,\mathbf{P}_{ac}(q \to \widetilde{q}) = \pi(\widetilde{q})\,\mathbf{P}_{pr}(\widetilde{q} \to q)\,\mathbf{P}_{ac}(\widetilde{q} \to q).$$

For a given (ergodic) proposal probability density function $\mathbf{P}_{pr}$, a possible choice for $\mathbf{P}_{ac}$ satisfying the latter equation is for example the Metropolis dynamics:

$$\mathbf{P}_{ac}(q \to \widetilde{q}) = \min \left\{ 1, \frac{\pi(\widetilde{q})\,\mathbf{P}_{pr}(\widetilde{q} \to q)}{\pi(q)\,\mathbf{P}_{pr}(q \to \widetilde{q})} \right\}. \tag{8}$$

Metropolis dynamics provides the chain of form (7) with minimal asymptotic variance and is therefore the most popular one, see Peskun's theorem [101].

Combination of MCMC and MD.   The transition probabilities $\mathbf{P}(q \to \widetilde{q})$ need not have any physical meaning in order to meet (6), but for a good acceptance ratio $\mathbf{P}_{ac}$, a combination of the Metropolis-Hastings algorithm for Markov chain Monte Carlo integration (MCMC) with molecular dynamics simulations (MD) is useful.

Starting in 1980, a variety of hybrid methods have been developed, which take the advantages of both MD and MCMC [3, 30]. These so-called HMC algorithms where originally developed for quantum chromo-dynamics, but they have been used successfully for condensed-matter systems [88, 16, 64, 37, 46] and also for biomolecular simulations [55, 35, 138]. HMC combine the large steps of MD in phase space with the property of MCMC to ensure ergodicity and to eliminate inaccuracies in the numerical computation of the Hamiltonian dynamics. In the following the HMC method is explained according to Fischer [33].

Proposal step.   HMC is a Metropolis algorithm, in which the proposal step $q \to \widetilde{q}$ is based on a molecular dynamics simulation with simulation length τ. To compute $\widetilde{q}$ out of a given q, we first determine a start momentum vector $p \in \mathbb{R}^{3n}$, where n is the number of atoms. The start momentum vector is taken from the Boltzmann distribution $\eta(p)$ according to the simulation temperature, see Allen and Tildesley [2] Section 5.7.2 for algorithmic details. Then, with some numerical integrator a trajectory of total length τ is computed. The starting point is given by (q, p), let the end point be denoted as $(\widetilde{q}, \widetilde{p}) = \Phi_h^\tau(q, p)$.

Due to determinism in the integration scheme the probability for a proposition $\mathbf{P}_{pr}(q \to \widetilde{q})$ only depends on the choice of the initial momenta p, which is $\mathbf{P}_{pr}(q \to \widetilde{q}) \propto \exp(-\beta\,K(p))$. If the numerical integrator is momentum-reversible, then for the reverse step $\widetilde{q} \to q$, we have to choose the start momentum vector $-\widetilde{p}$, which transfers the starting point $(\widetilde{q}, -\widetilde{p})$ to $(q, -p)$ in time span τ. This means, $\mathbf{P}_{pr}(\widetilde{q} \to q) \propto \exp(-\beta\,K(-\widetilde{p}))$. In equation (6) the two terms are integrated over the Lebesgue measure $dq \wedge d\widetilde{q}$. After transformation into the momenta space the left hand side depents on $dq \wedge dp$ and the right had side on $d\widetilde{q} \wedge d\widetilde{p}$. But this does not change anything in equation (8), because the mapping $(\widetilde{q}, \widetilde{p}) = \Phi_h^\tau(q, p)$ is area preserving and the two measures are equivalent. Inserting these results into (8) yields

$$
\begin{aligned}
\mathbf{P}_{ac}(q \to \widetilde{q}) &= \min\left\{1, \frac{\pi(\widetilde{q})\,\exp(-\beta\,K(-\widetilde{p}))}{\pi(q)\,\exp(-\beta\,K(p))}\right\} \\
&= \min\left\{1, \frac{\exp(-\beta\,V(\widetilde{q}))\,\exp(-\beta\,K(\widetilde{p}))}{\exp(-\beta\,V(q))\,\exp(-\beta\,K(p))}\right\} \\
&= \min\left\{1, \exp(-\beta\,(H(\widetilde{q}, \widetilde{p}) - H(q, p)))\right\}, \quad (9)
\end{aligned}
$$

i.e. the acceptance probability of the HMC proposal step is based on the change of the total energy during a numerical integration of the Hamiltonian. A reversible and area-preserving numerical integrator is necessary and sufficient for a correct sampling, for a rigorous proof see also Mehling et al. [88] Section III.

Hastings' class of algorithms.   Other acceptance rules instead of (8) may be interesting. The reason is that HMC is trapped inside local minima of V, because MD with small initial momenta does not overcome energy barriers. An alternative example is the *Barker dynamics* given by

$$
\mathbf{P}_{ac}(q \to \widetilde{q}) = \frac{\pi(\widetilde{q})\,\mathbf{P}_{pr}(\widetilde{q} \to q)}{\pi(\widetilde{q})\,\mathbf{P}_{pr}(\widetilde{q} \to q) + \pi(q)\,\mathbf{P}_{pr}(q \to \widetilde{q})} = \frac{1}{1 + \exp(\beta\,\Delta H)}. \quad (10)
$$

Metropolis dynamics intensifies the trapping effect, because going from higher to lower energies in (9) is always accepted. The acceptance probability (10) of Barker dynamics is always smaller than that of Metropolis dynamics, especially for "downhill" steps, which can antagonize the trapping effect of HMC. For a modern interpretation of the Metropolis-Hastings algorithm and a description of all possible acceptance rules see Billera and Diaconis [11], especially Remark 3.2.

Sampling convergence.   Using Markov Chain Monte Carlo integration theory the order of convergence of the chain (5) is $\mathcal{O}(N^{-\frac{1}{2}})$, which can be shown by a central limit

theorem, see [14] Section 16, Theorem 1. For more details about HMC and necessary assumptions[4] for the next theorem cf. [33].

Theorem 2.1 *Let $\langle A \rangle_\pi$ be a real-valued spatial observable and $q_i, i = 1, \ldots N$ the Markov chain (5) of an HMC method, then the mean value*

$$\frac{1}{N} \sum_{k=1}^{N} A(q_k)$$

is asymptotically normally distributed with expectation value $\langle A \rangle_\pi$ and variance σ^2, i.e. the following probability

$$\lim_{N \to \infty} \mathbf{P}\Big(\Big| \frac{1}{N} \sum_{k=1}^{N} A(q_k) - \langle A \rangle_\pi \Big| \leq c \frac{\sigma}{\sqrt{N}} \Big) = \Theta(c) \tag{11}$$

converges with

$$\Theta(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} \exp(-\frac{t^2}{2}) dt.$$

It is worth mentioning that the order of convergence $\mathcal{O}(N^{-\frac{1}{2}})$ does not depend on the dimension of the space Ω, which is very important for high-dimensional integration and cannot be achieved with deterministic methods [105, 98]. The price which we have to pay is: The error estimation in Theorem 2.1 is only a random variable and not a fixed value. For the special choice $c = 1$ in Theorem 2.1 the error estimation in (11) holds with a probability of

$$\Theta(1) \approx 0.84,$$

which will be used in Section 5.2.2 below. The speed of convergence in HMC depends on the correlation length of the realization of (5), see [88]. The shorter the correlation length is the faster MCMC converges. In this case, the chaotic behavior and the poor condition number of Hamiltonian dynamics becomes benefiting.

2.3.2 Choice of the Numerical Integrator

Table 1 shows some algorithmic details of HMC and related methods. It shows, which Hamiltonian the methods are based upon, the acceptance probabilities and eventually a necessary pointwise re-weighting of the sampling trajectory in order to be mathematically rigorous.

---

[4]Obviously, the transition probabilities in (5) must be ergodic, in order to sample the whole configuration space Ω.

	Hamiltonian	accept	re-weight
"exact" flow	H	1	no
orig. HMC	H	$e^{-\beta \Delta H} < 1$	no
SHMC (idea)	$\widetilde{H}$	1	$\widetilde{H} \to H$
SHMC	$\widetilde{H}_{app}$	$e^{-\beta \Delta \widetilde{H}_{app}} \approx 1$	$\widetilde{H}_{app} \to H$

Table 1: Possible approaches for a correct sampling according to the HMC method. Algorithmic consequences.

"Exact" flow. As we have seen in the derivation of the acceptance probability, a reversible and area-preserving numerical integrator is necessary and sufficient for a correct sampling, no matter how bad its state space solution is. However, if we could apply an "exact" integrator the total energy would be constant during simulation and, therefore, the acceptance probability according to Metropolis dynamics would be 1. Adaptive integrators can be used with a pre-defined deviation from the exact flow. An example for an adaptive integrator for Hamiltonian dynamics is DIFEX2, an extrapolation method based on Störmer discretization, see Deuflhard et al. [22, 23]. For DIFEX2, area-preservation and reversibility cannot be shown directly, but as the extrapolation method approximates the real flow, it inherits these properties from Φ^{τ}. Other adaptive integration methods and Fortran codes can be found in the book of Hairer et al. [49]. See also the first row of Table 1.

Original HMC. Instead of solving the real dynamics one can apply an arbitrary area-preserving and reversible integrator. In this case the mean acceptance ratio decreases exponentially with system size n and time-step discretization h in the numerical integration Φ_h^{τ}, see Gupta et al. [47] and Kennedy and Pendleton [70] for an analytic study of the computational cost of HMC. It is a general opinion that HMC methods are only suitable for small molecular systems, see e.g. Section 14.2 in [38]. The reason is that in order to keep the mean acceptance probability constant for increasing system size, the time-step h of the symplectic integrator has to decrease accordingly. Instead of time-step refinements, one can also increase the order of the integration method. Using Hamilton's principle of

a stationary action integral, Wendlandt and Marsden [136] derived a systematical scheme for creating so-called *variational integrators*. With these higher-order numerical integrators the acceptance ratio of HMC can be improved via better approximation properties. By omitting the constant step size h in variational integrators, one can even get symplectic *and* energy preserving integration schemes for the price of lower numerical efficiency. For an excellent overview of these methods see Lew et al. [78]. See also the second row of Table 1.

Shadow HMC (SHMC).   Another approach uses the fact that a symplectic numerical integrator solves the dynamics of a modified Hamiltonian $\widetilde{H}$ *exactly*, see [48] or Skeel and Hardy [121]. If one accepts each proposal step of the numerical dynamics simulation this is like computing the density of the modified Hamiltonian, see third row of Table 1. In order to get the right distribution in configuration space, one has to re-weight the resulting position states accordingly. This is only possible, if the modified Hamiltonian is known. Fortunately, $\widetilde{H}$ can be approximated up to arbitrary accuracy. For algorithmic details see Izaguirre and Hampton [65].

Approximated SHMC.   Approximating the modified Hamiltonian as exactly as possible is numerically expensive. Therefore, one would like to truncate the Taylor expansion of $\widetilde{H}$ after a finite number of terms in order to yield $\widetilde{H}_{app}$. This method is part of the TSHMC method of Akhmatskaya and Reich [1]. Again an acceptance rule is introduced, which zeroes out the numerical error of truncation, see the last row of Table 1. This method seems to be very promising for larger molecules, because the acceptance probability is almost 1, the method is mathematically rigorous, and numerically efficient (extra cost for computation of $\widetilde{H}_{app}$ is negligible).

2.4 Transfer Operator Approach

Metastabilities in configuration space.   Now we are interested in the identification of molecular conformations as described in the introduction. Conformations are defined as a decomposition of state space into physically distinguishable parts. Here it is assumed that physical properties mainly arise from the position coordinates $q \in \Omega$ of a molecule. Therefore, for a decomposition only the configuration space fraction Ω of the state space $\Omega \otimes \mathbb{R}^{3n}$ is taken into consideration. Two different position states are physically similar and belong to the same conformation, if molecular motion is likely to transfer one position state into the other[5]. Assume, we aim at a partitioning of Ω into sets $\mathcal{C}_1, \dots, \mathcal{C}_{n_C} \in \mathcal{B}(\Omega)$, where n_C is the number of conformations. Let $\wp : \mathbb{R} \times \Omega \times \mathcal{B}(\Omega) \to [0,1]$ denote the

---

[5]Note that the investigation of dynamically invariant subsets in the state space $\Omega \otimes \mathbb{R}^{3n}$ does not make sense, because this would simply lead to the constant energy level sets of the Hamiltonian $H(x)$, which do not reflect physical properties.

stochastic transition function. $\wp(\tau, q, \mathcal{C}_i)$ is the ratio of trajectories starting in $q \in \Omega$ with Maxwell-Boltzmann distributed initial momenta $p \in \mathrm{I\!R}^{3n}$, which end in $\mathcal{C}_i$ after time-span τ. Thus, with the above considerations a molecular conformation $\mathcal{C}_i$ is defined by the metastability property

$$
\begin{aligned}
\wp(\tau, q, \mathcal{C}_i) &\approx 1 \quad \Leftrightarrow \quad q \in \mathcal{C}_i, \\
\wp(\tau, q, \mathcal{C}_i) &\approx 0 \quad \Leftrightarrow \quad q \notin \mathcal{C}_i.
\end{aligned}
\tag{12}
$$

Transfer operator. The area-preserving property of a symplectic flow implies that molecular dynamics does not have attractors. The clustering into conformations cannot be based on subdivision techniques invented by Dellnitz and Junge [19], but one can extend their transfer operator approach [20] to symplectic flows. The stochastic transition function $\wp$ can be calculated via a transfer operator $P^\tau : L^r(\pi) \to L^r(\pi), r = 1, 2,$

$$
P^\tau u(q) = \int_{\mathrm{I\!R}^{3n}} u(\Pi_1 \Phi^{-\tau}(q, p))\, \eta(p)\, dp,
\tag{13}
$$

where $\Phi^{-\tau}$ is the reverse flow with initial state (q, p) and time span τ according to Hamilton dynamics and Π_1 is the projection of the molecular state $\Phi^{-\tau}(q, p)$ onto its position coordinate, see Schütte et al. [111, 114]. As molecular dynamics is reversible, one gets the position coordinates of the reverse flow by replacing p with $-p$ and computing the forward Hamilton dynamics. Note that the probability density function is symmetric, $\eta(p) = \eta(-p)$. Thus, forward and backward propagation are commutable. For more details about backward and forward transfer operators see [114].

Theoretically speaking, the transfer operator P^τ is self-adjoint for a Hamiltonian differential equation in a canonical ensemble:

Theorem 2.2 (Schütte [111]) *The transfer operator P^τ defined in (13) meets the following properties:*

 i) *$P^\tau e = e$ for the constant function $e : \Omega \to \mathrm{I\!R}$, $e(q) \equiv 1$.*

 ii) *$P^\tau : L^1(\pi) \to L^1(\pi)$ is a linear Markov operator.*

 iii) *$P^\tau : L^2(\pi) \to L^2(\pi)$ is a bounded and self-adjoint operator with $\|P^\tau u\|_\pi \leq \|u\|_\pi$ for $u \in L^2(\pi)$ and the corresponding weighted norm $\|\cdot\|_\pi$. Hence its spectrum satisfies $\sigma(P^\tau) \subset [-1, 1]$.*

Proof: *ii)* is the proposition of Lemma 3.7, *iii)* a consequence of Lemma 3.9 and Lemma 3.10 in [111]. *i)* can be shown by a simple calculation. As self-adjointness of P^τ is an important property in the following, here is the summary of Schütte's proof:

$$
\langle P^\tau u, v \rangle_\pi \;=\; \int_\Omega \Big[\int_{\mathrm{I\!R}^{3n}} u(\Pi_1 \Phi^{-\tau}(q, p))\, \eta(p)\, dp \Big]\, v(q)\, \pi(q)\, dq
$$

$$
\begin{aligned}
&= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, \Phi^{-\tau} x)\, v(\Pi_1\, x)\, \mu_{can}(x)\, dx \\
(1^*)\quad &= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, y)\, v(\Pi_1\, \Phi^{\tau} y)\, \mu_{can}(y)\, dy \\
&= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, (\widetilde{q}, \widetilde{p}))\, v(\Pi_1\, \Phi^{\tau}(\widetilde{q}, \widetilde{p}))\, \mu_{can}(\widetilde{q}, \widetilde{p})\, dy \\
(2^*)\quad &= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, (\widetilde{q}, -\widetilde{p}))\, v(\Pi_1\, \Phi^{\tau}(\widetilde{q}, -\widetilde{p}))\, \mu_{can}(\widetilde{q}, -\widetilde{p})\, dy \\
(3^*)\quad &= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, (\widetilde{q}, \widetilde{p}))\, v(\Pi_1\, \Phi^{-\tau}(\widetilde{q}, \widetilde{p}))\, \mu_{can}(\widetilde{q}, \widetilde{p})\, dy \\
&= \int_{\Omega \otimes \mathbb{R}^{3n}} u(\Pi_1\, y)\, v(\Pi_1\, \Phi^{-\tau} y)\, \mu_{can}(y)\, dy \\
&= \langle u, P^{\tau} v \rangle_{\pi}
\end{aligned}
$$

With $x = \Phi^{\tau} y$ equation (1^*) holds due to invariance of μ_{can} with respect to the Hamiltonian flow, $H(y) = H(\Phi^{\tau} y)$, and due to $|\det(D\Phi^{\tau})| = 1$, because (1) is symplectic. In equation (2^*) the p-direction of integration is switched. Equation (3^*) holds due to reversibility of (1) and $\eta(p)$-symmetry of μ_{can} with $y = (\widetilde{q}, \widetilde{p})$. $\hfill\square$

From set-based to function-based conformations.   The stochastic transition function $\wp$ can be calculated as

$$
\wp(\tau, q, \mathcal{C}_i) = P^{\tau}\, \mathbb{1}_{\mathcal{C}_i}(q),
$$

where $\mathbb{1}_{\mathcal{C}_i} : \Omega \to \{0, 1\}$ is the characteristic function of the set $\mathcal{C}_i$, i.e. $\mathbb{1}_{\mathcal{C}_i}(q) = 1$ if $q \in \mathcal{C}_i$ and $\mathbb{1}_{\mathcal{C}_i}(q) = 0$ otherwise. The metastability property (12) of a conformation $\mathcal{C}_i$ can be rewritten as

$$
P^{\tau}\, \mathbb{1}_{\mathcal{C}_i}(q) \approx \mathbb{1}_{\mathcal{C}_i}(q), \tag{14}
$$

which will be discretized in the following. Via this transfer operator description of conformations, the set-based approach can be extended to a pure function-based definition of conformations. This will be done in three steps:

1. Assume a decomposition of Ω into pairwise disjoint sets $\mathcal{X}_1, \ldots, \mathcal{X}_s$ with

$$
\mathbb{1}_{\mathcal{C}_j} = \sum_{i=1}^{s} \chi_{disc}(i, j)\, \mathbb{1}_{\mathcal{X}_i}, \quad j = 1, \ldots, n_c,
$$

where $\chi_{disc} \in \{0, 1\}^{s \times n_C}$ and $\chi_{disc}(i, j) = 1$ if $\mathcal{X}_i$ belongs to conformation $\mathcal{C}_j$ and $\chi_{disc}(i, j) = 0$ otherwise. In order to find this clustering, Deuflhard et al. [26] invented the so-called Perron Cluster Analysis. However, this method is not robust

against small perturbations of the transfer operator, because a perturbation of P^τ is a continuous process, whereas changes of χ_{disc} can only switch between 0 and 1. This algorithm is not part of this thesis.

2. A robust version of the Perron Cluster Analysis is possible, see Deuflhard and Weber [28], if we allow $\chi_{disc}(i,j)$ to attain values between the extremes 0 and 1. In this case $\mathbb{1}_{\mathcal{C}_1}, \ldots, \mathbb{1}_{\mathcal{C}_{n_C}}$ are not "characteristic" any more, but replaced by *almost* characteristic functions $\chi_1, \ldots, \chi_{n_C}$. In terms of cluster analysis this is a transition from a so-called *crisp* clustering into a *fuzzy* clustering. The Robust Perron Cluster Analysis is the main step from a set-based conformation analysis to a function-based one and it is described in Section 3.

3. In Section 4 we aim at a meshless discretization of Ω, which means that the discretization of Ω into sets (i.e. characteristic functions $\mathbb{1}_{\mathcal{X}_1}, \ldots, \mathbb{1}_{\mathcal{X}_s}$) is replaced by an arbitrary function basis $\xi_1, \ldots, \xi_s$, in order to get trial functions for a global Galerkin discretization of (14). The advantages of this approach and its impact on the HMC sampling method can be seen in Section 4.4. The construction of this function basis is part of Section 5.1, for which an error analysis is derived in Section 5.2.

As we will abandon the set-based concept of conformations by and by, in the following we will not use the abbreviation $\mathbb{1}_{\mathcal{C}_i}$ for characteristic functions any more. If characteristic functions occur, it will be stated explicitly.

3   Robust Perron Cluster Analysis

In this section a robust algorithm for the identification of an almost characteristic decomposition of Ω into almost invariant membership functions $\chi_1, \ldots, \chi_{n_C}$ is derived. First, the physical meaning of these functions χ_l as conformations is shown. Afterwards, an optimization problem is derived, which leads to an optimal decomposition of Ω with regard to uniqueness or metastability of the conformations. The algorithm for clustering needs a spectral analysis of the transfer operator for eigenvalues near the unit circle which is based on ideas of Dellnitz and Junge [20].

3.1   Almost Characteristic Functions

Like in Section 2.4 a conformation l of a molecule is defined as an *almost characteristic function* (also called *membership function* in literature) $\chi_l : \Omega \to [0, 1]$. A *decomposition* of Ω into n_C conformations $\chi_1, \ldots, \chi_{n_C}$ in this context means that these functions form a *partition of unity*. For the discretization of χ_l a partition of unity function basis is used:

Definition 3.1 *A set of functions* $\xi_1, \ldots, \xi_s : \Omega \to \mathbb{R}$ *is called a* membership basis *if it is a* partition of unity, *i.e. for all* $q \in \Omega$

$$\sum_{i=1}^{s} \xi_i(q) = 1, \tag{15}$$

and the functions ξ_j *are nonnegative, i.e.*

$$\xi_i(q) \geq 0 \tag{16}$$

for all $q \in \Omega$ *and* $i = 1, \ldots, s$.

In Section 3, we investigate membership basis functions $\xi_1, \ldots, \xi_s$, which are characteristic functions of a decomposition of $\Omega = \bigcup_{i=1,\ldots,s} \mathcal{X}_i$ into pairwise disjoint subsets $\mathcal{X}_i \subset \Omega$, i.e. $\xi_i(q) = 1$ if $q \in \mathcal{X}_i$ and $\xi_i(q) = 0$ otherwise. In this case, the membership basis is a set of pairwise orthogonal functions in $L^2(\pi)$. In Section 4, the choice of membership basis functions is generalized to arbitrary non-orthogonal systems. Because of the partition of unity and the non-negativity assumption in Definition 3.1, a molecular conformation is a *convex combination* of the membership basis functions, i.e.

$$\chi_l(q) = \sum_{i=1}^{s} \xi_i(q) \, \chi_{disc}(i, l),$$

where $q \in \Omega$, and $\chi_{disc} \in \mathbb{R}^{s \times n_C}$ is a row-stochastic rectangular matrix. In the following, we want to find an optimal matrix χ_{disc} of linear combination factors. Like in set-based conformation dynamics, thermodynamical information for drug design can be calculated via this function-based concept of conformations as follows:

Partition functions.   Each conformation χ_i of the molecule has its own partition function $\widetilde{w}_i > 0$ with

$$\widetilde{w}_i = \langle \chi_i \rangle_\pi, \quad i = 1, \ldots, n_C. \tag{17}$$

$\widetilde{w}_i$ denotes the part of the ensemble which corresponds to conformation χ_i. The partition function is an important term to calculate thermodynamic properties of molecular systems. In Section 5.2.2 an error estimation for the computation of $\widetilde{w}_i, i = 1, \ldots, n_C$ related to the approximation error of χ_i is derived.

Observables.   Spatial observables of a function $A : \Omega \to \mathbb{R}$ like in (3) can also be computed separately for each function-based conformation $\chi_1, \ldots, \chi_{n_C} : \Omega \to [0, 1]$ via

$$\langle A \rangle_{\pi, \chi_i} := \frac{1}{\widetilde{w}_i} \langle A \, \chi_i \rangle_\pi = \frac{1}{\widetilde{w}_i} \int_\Omega A(q) \, \chi_i(q) \, \pi(q) \, dq, \quad i = 1, \ldots, n_C. \tag{18}$$

From this point of view, a conformation χ_i, which is a function in configuration space, can also be seen as a molecular state. Oršič and Shalloway [99] therefore denote this kind of conformation as *macrostate*. A function-based conformation can be seen as

1. a macrostate i, which is described via a modified distribution in state space, i.e. if π is the Boltzmann distribution of spacial coordinates, then the macrostate is defined by the distribution $\frac{1}{\widetilde{w}_i} \chi_i \pi$, see (18), or

2. as a macrostate i, which is totally described via a modified potential energy function $V - \frac{1}{\beta} \ln(\chi_i)$, because

$$\frac{1}{\widetilde{w}_i} \chi_i \pi = \frac{1}{Z_q \widetilde{w}_i} \chi_i \exp(-\beta V) = \frac{1}{Z_q \widetilde{w}_i} \exp(-\beta(V - \frac{1}{\beta} \ln(\chi_i))),$$

   see also (56).

Transition probabilities.   The behavior of a single molecule in the ensemble can be seen as a continuous-time Markov jump process, $\bar{x}(t) \in \{1, \ldots, n_C\}$, where a molecule can jump between the n_C different macrostates, see also Section 4.3. Neglecting the exchange of energy with the surroundings during the dynamics simulation, the corresponding transition probabilities for a discretization of time into intervals of length τ can be computed as proposed in [114]:

Definition 3.2 *The stochastic coupling matrix* $\mathcal{P} \in \mathbb{R}^{n_C \times n_C}$, *which provides the transition probabilities between the conformations* $\chi_1, \ldots, \chi_{n_C} : \Omega \to [0, 1]$, *is defined as*

$$\mathcal{P} = \left(\frac{\langle \chi_i, P^\tau \chi_j \rangle_\pi}{\langle \chi_i \rangle_\pi} \right)_{i,j=1,\ldots,n_C},$$

where P^τ *is the transfer operator defined in equation (13). The trace of* $\mathcal{P}$ *is referred to as* metastability *of the conformations* $\chi_1, \ldots, \chi_{n_C}$.

3.2   Galerkin Discretization

Metastable conformations.   With the transfer operator approach (14) a metastable almost characteristic function $\chi_l : \Omega \to [0,1]$ has the following defining property:

$$\chi_l \approx P^\tau \chi_l, \quad l = 1, \ldots, n_C. \tag{19}$$

For a set-based discretization of Ω the membership basis functions meet $\xi_i(q) \in \{0,1\}$ for $i = 1, \ldots, s$ and therefore $\xi_i^2 = \xi_i$. This together with the partition of unity implies that the product of two different membership basis functions vanishes, $\xi_i \xi_j = 0$ for $i \neq j$. With these preparations the Galerkin discretization of the condition (19) is:

$$\chi_l \approx P^\tau \chi_l$$

$$\Leftrightarrow \quad \sum_{j=1}^s \xi_j \, \chi_{disc}(j,l) \approx P^\tau \sum_{j=1}^s \xi_j \, \chi_{disc}(j,l)$$

$$\Rightarrow \quad \int_\Omega \sum_{j=1}^s \xi_i(q) \, \xi_j(q) \, \chi_{disc}(j,l) \, \pi(q) \, dq \approx \int_\Omega \sum_{j=1}^s \xi_i(q) \, P^\tau \xi_j(q) \, \chi_{disc}(j,l) \, \pi(q) \, dq$$

$$\Leftrightarrow \quad \langle \xi_i \rangle_\pi \, \chi_{disc}(i,l) \approx \sum_{j=1}^s \langle \xi_i, P^\tau \xi_j \rangle_\pi \, \chi_{disc}(j,l), \quad i = 1, \ldots, s$$

$$\Leftrightarrow \quad D\chi_{disc} \approx \overline{P}\chi_{disc},$$

where $D = \mathrm{diag}(w_1, \ldots, w_s)$ is the diagonal matrix of the statistical weights of the membership basis functions $w_i = \langle \xi_i \rangle_\pi, i = 1, \ldots, s$ and $\overline{P}$ is the Galerkin discretization of the transfer operator P^τ, i.e. $\overline{P}(i,j) = \langle \xi_i, P^\tau \xi_j \rangle_\pi$. Self-adjointness of P^τ implies that $\overline{P} \in \mathbb{R}^{s \times s}$ is a symmetric matrix. The diagonal entries of D equal the row sums of $\overline{P}$, therefore, $P = D^{-1}\overline{P}$ is a stochastic matrix. Like in Definition 3.2, where $\mathcal{P}$ defines the transition probabilities for the conformations, the entry $P(i,j)$ can be interpreted as transition probability between the subsets of Ω given by the characteristic functions ξ_i and ξ_j, see also Schütte [111]. With these preparations the discretization of (19) is

$$P\chi_{disc} \approx \chi_{disc}. \tag{20}$$

Eigenvalues and eigenvectors of P.   For Robust Perron Cluster Analysis the eigenvectors $X \in \mathbb{R}^{s \times s}$ of P play an important role. In order to exploit the Hamiltonian structure for the computation of the eigenvectors and the eigenvalues $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_s)$, one can transform the eigenvalue problem of P into a symmetric form:

$$P\,X = X\,\Lambda$$

$$\Leftrightarrow \quad (D^{-1/2}\overline{P}D^{-1/2})\,D^{1/2}\,X = D^{1/2}\,X\Lambda. \tag{21}$$

Instead of solving the eigenvalue problem for P one can solve the symmetric eigenvalue problem for the nonnegative matrix $(D^{-1/2}\,\overline{P}\,D^{-1/2})$ and rescale its eigenvectors $Y \in \mathrm{I\!R}^{s \times s}$ with $X = D^{-1/2}\,Y$. The eigenvalues are identical, which also shows that P has a real-valued spectrum and generalized symmetric eigenvectors with $I = Y^\top Y = X^\top D X$, where $I \in \mathrm{I\!R}^{s \times s}$ is the s-unit matrix. A simple Gerschgorin estimation yields $\lambda_i \in [-1, 1]$, where $\lambda_1 = 1$ always occurs: For a stochastic matrix P the eigenvector $X(:, 1)$ corresponding to the eigenvalue $\lambda_1 = 1$ is a constant vector $X(:, 1) = (1, \dots, 1)^\top \in \mathrm{I\!R}^s$, which will be important for Robust Perron Cluster Analysis, see Section 3.4.1.

Basic idea of Perron Cluster Analysis. As P has a real-valued spectrum with sorted eigenvalues $\lambda_1 \geq \dots \geq \lambda_s$ and an eigenvector matrix X, where the i-th column of X is an eigenvector corresponding to λ_i, the clustering can be computed in terms of the basis X as

$$\chi_{disc} = X\mathcal{A}, \tag{22}$$

with $\mathcal{A} \in \mathrm{I\!R}^{s \times n_C}$. As equation (20) is aimed at, one can restrict equation (22) to eigenvectors, which correspond to eigenvalues $\lambda_1, \dots, \lambda_{n_C}$ near the maximal eigenvalue 1, i.e. $X \in \mathrm{I\!R}^{s \times n_C}$, $\mathcal{A} \in \mathrm{I\!R}^{n_C \times n_C}$ and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_{n_C})$. The spectral analysis is also the basis for identifying the number of conformations n_C, see Section 3.5. The justification for this restriction of the full basis expansion (22) to only n_C eigenvectors can be given by perturbation analysis, see Section 3.3. This approach reduces the number of unknowns from $\dim(\chi_{disc}) = n_C \cdot s$ to $\dim(\mathcal{A}) = n_C^2$, because if $\mathcal{A}$ is given χ_{disc} can be computed via (22). This is a real simplification, because in most applications $n_C \ll s$. In this context, cluster analysis is equivalent to finding a basis transformation $\mathcal{A}$ of n_C eigenvectors X, such that $\chi_{disc} = X\mathcal{A}$ is a row-stochastic matrix. In general, there are many matrices $\mathcal{A}$ which meet this condition, therefore, in Section 3.4 an optimization problem for the entries of $\mathcal{A}$ is derived.

3.3   Perturbation Analysis

In order to understand why the optimal clustering χ_{disc} can be seen as a linear combination of the first n_C eigenvectors of P, we first examine the "ideal case", where the matrix P is block-structured. For perturbation analysis, this block-structured matrix is denoted as $\widetilde{P}$. The row sum of a stochastic matrix $\widetilde{P} \in \mathrm{I\!R}^{s \times s}$ is 1 for each row, which can be expressed via $\widetilde{P}e = e$ for the constant vector $e = (1, \dots, 1)^\top \in \mathrm{I\!R}^s$. If $\widetilde{P}$ is a transition matrix for a non-ergodic Markov chain having n_C disjoint communicating index subsets $\mathcal{C}_1, \dots, \mathcal{C}_{n_C}$, then row sums can be restricted to the corresponding index subsets. In this case $\widetilde{P}$ is a block-diagonal matrix[6]. For a matrix $\widetilde{\chi}_{disc} \in \mathrm{I\!R}^{s \times n_C}$, with $\widetilde{\chi}_{disc}(i, j) = 1$ if $i \in \mathcal{C}_j$ and

---

[6]possibly with a permutation of rows and columns.

$\widetilde{\chi}_{disc}(i, j) = 0$ otherwise, this non-ergodicity can be expressed in matrix notation:

$$\widetilde{P}\widetilde{\chi}_{disc} = \widetilde{\chi}_{disc}.$$

This is an eigenvector equation. For every regular transformation matrix $\widetilde{\mathcal{A}} \in \mathbb{R}^{n_C \times n_C}$ the matrix $\widetilde{X} \in \mathbb{R}^{s \times n_C}$, with $\widetilde{X} = \widetilde{\chi}_{disc}\widetilde{\mathcal{A}}^{-1}$, provides a basis of eigenvectors of $\widetilde{P}$ corresponding to $\lambda = 1$. In this case, $\widetilde{\mathcal{A}}^{-1}$ is only a shift of the piecewise constant components of $\widetilde{\chi}_{disc}$. Thus, for a reducible stochastic matrix $\widetilde{P}$ the n_C communicating index subsets are given by the constant level pattern of the first n_C eigenvectors $\widetilde{X}$. Vice versa, for a given set of eigenvectors $\widetilde{X}$ there is a regular transformation matrix $\widetilde{\mathcal{A}}$, such that the "solution" $\widetilde{\chi}_{disc}$ of the cluster problem is given as

$$\widetilde{\chi}_{disc} = \widetilde{X}\widetilde{\mathcal{A}}.$$

This is a justification of the approximation (22) for the solution of the cluster problem for the irreducible matrix P via a linear mapping of the eigenvectors, which correspond to eigenvalues near $\lambda_1 = 1$. Now it is shown that convergence of χ_{disc} towards $\widetilde{\chi}_{disc}$ holds when P becomes more and more blocky. As the eigenvalue problem for P can be transformed into a symmetric one, see (21), perturbation analysis can be applied. Assume the regularity conditions of Theorem 6.1 from Kato [69]: Let the analytic expansion

$$P(\epsilon) = \widetilde{P} + \epsilon P^{(1)} + \epsilon^2 P^{(2)} + \dots$$

define a family of matrices, such that $\widetilde{P}$ is reducible with n_C communicating index subsets and $P(\epsilon)$ is generalized symmetric and stochastic for $\epsilon_{max} \geq \epsilon \geq 0$. Furthermore, let $P(\epsilon)$ for $\epsilon > 0$ be primitive, i.e. the Perron root $\lambda_1(\epsilon) = 1$ is simple and each $P(\epsilon)$ has a unique left positive eigenvector w corresponding to λ_1 (w is called the *invariant distribution*). With these regularity conditions, the eigenvalues $\lambda_i(\epsilon), i = 1, \dots, n_C$, and eigenvectors of $P(\epsilon)$ are analytic for $\epsilon \geq 0$ with the following expansion

$$X = \widetilde{X} + \epsilon X^{(1)} + \mathcal{O}(\epsilon^2), \tag{23}$$

which has been shown in [114]. As P has strong structural properties, the following lemma holds.

Lemma 3.3 ([28]) *There exists a matrix $B \in \mathbb{R}^{n_C \times n_C}$ such that the first order term $X^{(1)}$ in the expansion (23) can be simplified to*

$$X^{(1)} = \widetilde{\chi}_{disc} B.$$

As we have shown, $\widetilde{X}$ is a matrix of piece-wise constant eigenvectors of $\widetilde{P}$. Lemma 3.3 shows that the matrix $X^{(1)}$ shares the same piece-wise constant level pattern like $\widetilde{X}$. The main conclusion is that $\widetilde{X}$ and $\widetilde{X} + \epsilon X^{(1)}$ define the same n_C-subspace and therefore lead to the same solution in terms of χ_{disc}. This conlusion is part of the next

Theorem 3.4 ([28]) *In terms of the perturbation parameter $\epsilon > 0$, the following approximation result holds:*

$$\chi_{disc} = \widetilde{\chi}_{disc} + \mathcal{O}(\epsilon^2).$$

Assume that for the transformation matrix $\mathcal{A}$, the diagonal matrix

$$\Lambda = \Lambda(\epsilon) = \mathrm{diag}(\lambda_1(\epsilon), \ldots, \lambda_{n_C}(\epsilon)),$$

and the n_C-unit matrix I the following inequality

$$\|\mathcal{A}^{-1}\Lambda\mathcal{A} - I\|_1 < 1$$

is satisfied. Then metastability, see Definition 3.2, can be bounded via

$$\mathrm{trace}(\Lambda) - \mathcal{O}(\epsilon^2) \leq \mathrm{trace}(\mathcal{P}) \leq \mathrm{trace}(\Lambda).$$

Proof: The complete proof can be found in [28]. An important step in the proof of the metastability bounds is a reformulation of $\mathcal{P}$ via

$$\mathcal{P} = \mathcal{S}\,\mathcal{A}^{-1}\,\Lambda\,\mathcal{A}, \tag{24}$$

where $\mathcal{S}$ is defined as:

$$\mathcal{S} = \left(\frac{\langle \chi_i, \chi_j \rangle_\pi}{\langle \chi_i \rangle_\pi}\right)_{i,j=1,\ldots,n_C}.$$

Due to the $\mathcal{O}(\epsilon)$-result for χ_{disc} (and therefore also for χ_l, $l = 1, \ldots, n_C$) this matrix is a perturbation of the unit matrix I, i.e. $\mathcal{S} = I - \mathcal{O}(\epsilon^2)$, which implies that $\mathrm{trace}(\mathcal{P}) = \mathrm{trace}(\Lambda) - \mathcal{O}(\epsilon^2)$. $\qquad\square$

From Theorem 3.4 the robustness of the approach (22) with almost characteristic functions becomes clear. The more blocky the transition matrix P is, the more characteristic are the resulting conformations and $\mathcal{O}(\epsilon^2)$-convergence towards the unperturbed clustering holds. In practice, only the perturbed transition matrices P are available, the unperturbed matrix $\widetilde{P}$ is only artificial. In Section 3.5.1 a construction method for $\widetilde{P}$ is shown, such that $\widetilde{P}$ has the same invariant distribution like P, i.e. the eigenvectors of P and $\widetilde{P}$ are orthogonal with respect to the same diagonal matrix D.

In the following, the term *Perron cluster eigenvalues* and *Perron cluster eigenvectors* is used for the eigenvalues $\lambda_1, \ldots, \lambda_{n_C} \approx 1$ and the corresponding eigenvectors $X \in \mathbb{R}^{s \times n_C}$. From this denotation the abbreviations PCCA for Perron Cluster (Cluster) Analysis and PCCA+ for Robust Perron Cluster (Cluster) Analysis is derived, because these *cluster* methods use the Perron *cluster* eigenvalues, see also [28].

3.4 Optimization Problem

Cluster analysis in terms of (22) is not unique. There is an uncountable number of transformation matrices $\mathcal{A}$ such that $\chi_{disc} = X\mathcal{A}$ is a row-stochastic matrix. For example, there is always a matrix $\mathcal{A}$, such that every entry of χ_{disc} is identical to $1/n_C$, i.e. the conformations are identical constant functions. Therefore, we have to introduce some more properties in order to get an "optimal" clustering in terms of separation or metastability.

3.4.1 Feasible Set

For the set $\{\chi_1, \ldots, \chi_{n_C}\}$ of almost characteristic functions, the partition of unity and the non-negativity constraints should hold. In terms of the discretization factors χ_{disc} for a membership basis $\xi_1, \ldots, \xi_s$, this means that $\chi_{disc} \in \mathbb{R}^{s \times n_C}$ has to be a row-stochastic rectangular matrix. Additionally it is assumed that $\chi_{disc} = X\mathcal{A}$ is a linear combination $\mathcal{A} \in \mathbb{R}^{n_C \times n_C}$ of the eigenvector data. This assumption restricts the problem to n_C^2 unknowns, because by fixing the n_C^2 entries of $\mathcal{A}$ one also determines the set of conformations. In the following the set $F_{\mathcal{A}}$ of feasible transformation matrices is characterized. For given eigenvector data $X \in \mathbb{R}^{s \times n_C}$ the non-negativity constraint is

$$\chi_{disc}(l,i) \geq 0 \;\Leftrightarrow\; \sum_{j=1}^{n_C} \mathcal{A}(j,i)\, X(l,j) \geq 0, \quad i = 1, \ldots, n_C, \; l = 1, \ldots, s, \qquad (25)$$

and the partition of unity can be expressed as

$$\sum_{i=1}^{n_C} \chi_{disc}(l,i) = 1 \;\Leftrightarrow\; \sum_{i=1}^{n_C}\sum_{j=1}^{n_C} \mathcal{A}(j,i) X(l,j) = 1, \quad l = 1, \ldots, s. \qquad (26)$$

Equation (26) can be written in matrix form using the constant vectors

$$\widetilde{e} = (1, \ldots, 1)^\top \in \mathbb{R}^{n_C}, \quad e = (1, \ldots, 1)^\top \in \mathbb{R}^s$$

and the 1st unit vector $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{n_C}$:

$$X\mathcal{A}\widetilde{e} = e \;\Leftrightarrow\; \mathcal{A}\widetilde{e} = e_1 \;\Leftrightarrow\; \sum_{i=1}^{n_C} \mathcal{A}(j,i) = \delta_{j,1}, \; i = 1, \ldots, n_C,$$

where δ is the Kronecker symbol and it is used that the first eigenvector of a stochastic matrix is constant, i.e. the first column of X equals e. Together with (25) we arrive at

the following constraints[7] for the feasible set $F_{\mathcal{A}}$:

$$
(1) \quad \mathcal{A}(j,1) \;=\; \delta_{j,1} - \sum_{i=2}^{n_C} \mathcal{A}(j,i), \quad j = 1, \ldots, n_C \;,
$$

$$
(2) \quad \mathcal{A}(1,i) \;\geq\; - \sum_{j=2}^{n_C} \mathcal{A}(j,i) X(l,j), \quad i = 1, \ldots, n_C, \quad l = 1, \ldots, s.
$$

(27)

The set $F_{\mathcal{A}}$ has at least the feasible point $\mathcal{A}^*(j,i) := \delta_{j,1} n_C^{-1}$ and is therefore not empty.

Vertices of $F_{\mathcal{A}}$.   Note that the constraints for $F_{\mathcal{A}}$ are linear, which also implies that $F_{\mathcal{A}}$ is convex. In Lemma 3.7, the vertex set[8] $v(F_{\mathcal{A}})$ of $F_{\mathcal{A}}$ plays an important role. Because of $F_{\mathcal{A}} \subset \mathbb{R}^{n_C^2}$, a vertex $\mathcal{A} \in v(F_{\mathcal{A}}) \subset F_{\mathcal{A}}$ is determined by n_C^2 active linear constraints out of (27). A subset of $2n_C$ active constraints can be characterized by the following

Lemma 3.5 *Let $\mathcal{A} \in F_{\mathcal{A}}$ be regular and $\chi_{disc} = X\mathcal{A}$. If there exists an index $i \in \{1, \ldots, n_C\}$, such that for all $l = 1, \ldots, s$*

$$
0 < \chi_{disc}(l,i)
$$

holds, then $\mathcal{A} \notin v(F_{\mathcal{A}})$.

Proof: Without loss of generality $i = 1$ in the above lemma. Define

$$
\mathcal{A}^* := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n_C \times n_C},
$$

and

$$
0 < \delta := \min_{j=1,\ldots,s} \chi_{disc}(j,1) < 1, \quad \mathcal{B} := \frac{\mathcal{A} - \delta \mathcal{A}^*}{1 - \delta},
$$

where χ_{disc} and $\mathcal{A}$ are given by the lemma. Then, we have $\mathcal{A}^*, \mathcal{B} \in F_{\mathcal{A}}$ and $\mathcal{A}$ is not a vertex of the convex set $F_{\mathcal{A}}$, because $\mathcal{A} = (1 - \delta)\mathcal{B} + \delta \mathcal{A}^*$. $\square$

Because of Lemma 3.5, at a vertex of $F_{\mathcal{A}}$, for every $i = 1, \ldots, n_C$, there exists an $l \in \{1, \ldots, s\}$ such that $\chi_{disc}(l,i) = 0$. This means, for every $i = 1, \ldots, n_C$ at least one

---

[7]The presentation of the theory of Robust Perron Cluster Analysis slightly differs from [28]. The reason is that in the present text the transformation of (27) into (28) and the introduction of unconstrained minimization in Algorithm 3.11 without any "side condition" seems to be more directly.

[8]In [131] the notion "indecomposable membership function" was introduced, which is nothing else but a discretization matrix $\chi_{disc} = X\mathcal{A}$ for $\mathcal{A} \in v(F_{\mathcal{A}})$. Therefore, Lemma 3.5 can be compared to Lemma 3.3(iii) in [131].

linear constraint out of (27.2) is active. Therefore, one can define a subset $F'_{\mathcal{A}} \subset F_{\mathcal{A}}$ with $v(F_{\mathcal{A}}) \subset F'_{\mathcal{A}}$ by the following constraints:

$$
\begin{aligned}
(1) \quad & \mathcal{A}(j,1) \;=\; \delta_{j,1} - \sum_{i=2}^{n_C} \mathcal{A}(j,i), \quad j = 1, \ldots, n_C \,, \\[2mm]
(2) \quad & \mathcal{A}(1,i) \;=\; - \min_{l=1,\ldots,s} \sum_{j=2}^{n_C} \mathcal{A}(j,i) X(l,j), \quad i = 1, \ldots, n_C.
\end{aligned}
\tag{28}
$$

This subset $F'_{\mathcal{A}}$ of the feasible set $F_{\mathcal{A}}$, which includes an optimal vertex $\mathcal{A}$ of $F_{\mathcal{A}}$, is very important in the following, because it is defined via equations, which reduces the dimension of the search space. Except for (28.1) in the case $j = 1$, every equality in (28) is invariant against positive scaling $\mathcal{A} \to \gamma \mathcal{A}, \gamma > 0$. This partial invariance against scaling is used for a construction of feasible matrices $\mathcal{A}$ in Algorithm 3.10 and for a transformation into an unconstrained optimization problem in Algorithm 3.11 below.

3.4.2 Objective Functions

Every matrix $\mathcal{A} \in F_{\mathcal{A}}$ leads to a feasible clustering $\chi_{disc} = X\mathcal{A}$ in terms of almost characteristic functions. We now specify some objective functions, which should be maximized for an *optimal* transformation matrix $\mathcal{A}$. Here are two natural objective functions together with their upper bounds:

- For each conformation $\chi_i, i = 1, \ldots, n_C$ there should be a point $q \in \Omega$ with maximal degree of membership $\chi_i(q) \approx 1$. This optimization problem can be expressed by the objective function

$$
I_1(\mathcal{A}) = \sum_{j=1}^{n_C} \max_{l=1,\ldots,s} \chi_{disc}(l,j) = \sum_{j=1}^{n_C} \max_{l=1,\ldots,s} \sum_{i=1}^{n_C} X(l,i)\,\mathcal{A}(i,j) \le n_C. \tag{29}
$$

A justification for this objective function is given in (34) in Section 3.5.3 below.

- Metastability of the conformations $\chi_1, \ldots, \chi_{n_C}$ should be maximal, see Definition 3.2. With $\widetilde{D} = \operatorname{diag}(\widetilde{w}_1, \ldots, \widetilde{w}_{n_C})$ and

$$
\mathcal{P} = \widetilde{D}^{-1} \chi_{disc} \overline{P} \chi_{disc} = \widetilde{D}^{-1} \mathcal{A}^{\top} X^{\top} \overline{P} X \mathcal{A} = \widetilde{D}^{-1} \mathcal{A}^{\top} \Lambda \mathcal{A},
$$

the corresponding objective function is

$$
I_2(\mathcal{A}) = \operatorname{trace}(\mathcal{P}) = \operatorname{trace}(\widetilde{D}^{-1} \mathcal{A}^{\top} \Lambda \mathcal{A}) \le \sum_{i=1}^{n_C} \lambda_i.
$$

In order to get a formulation of metastability I_2 in terms of $\mathcal{A}$, we proof that the partition functions can be expressed by the first column of a feasible transformation matrix:

Lemma 3.6 *The partition function $\widetilde{w}_i$ of conformation χ_i in equation (17) can be obtained from the transformation matrix $\mathcal{A}$ via*

$$\widetilde{w}_i = \mathcal{A}(1, i), \quad i = 1, \ldots, n_C.$$

Proof: From (17) we get

$$\widetilde{w}_i = \langle \chi_i \rangle_\pi = \sum_{j=1}^{s} \chi_{disc}(j, i) \langle \xi_j \rangle_\pi = \sum_{j=1}^{s} \sum_{k=1}^{n_C} X(j, k) \mathcal{A}(k, i) \langle \xi_j \rangle_\pi.$$

Due to (21) the eigenvectors $X(:, k)$ for $k = 2, \ldots, n_C$ are orthogonal to the first one, $X(:, 1) = e$, w.r.t a generalized dot product, i.e.

$$0 = X(:, k)^\top D e = \sum_{j=1}^{s} X(j, k) \langle \xi_j \rangle_\pi.$$

Hence

$$\begin{aligned}
\widetilde{w}_i &= \sum_{k=1}^{n_C} \mathcal{A}(k, i) \sum_{j=1}^{s} X(j, k) \langle \xi_j \rangle_\pi = \mathcal{A}(1, i) \sum_{j=1}^{s} \underbrace{X(j, 1)}_{=1} \langle \xi_j \rangle_\pi \\
&= \mathcal{A}(1, i) \sum_{j=1}^{s} \langle \xi_j \rangle_\pi = \mathcal{A}(1, i).
\end{aligned}$$

$\square$

Via Lemma 3.6 metastability I_2 becomes:

$$I_2(\mathcal{A}) = \sum_{i=1}^{n_C} \lambda_i \sum_{j=1}^{n_C} \frac{\mathcal{A}(i, j)^2}{\mathcal{A}(1, j)}.$$

Note that for I_1 and I_2 the eigenproblem data X and λ is fixed and $\mathcal{A}$ consists of the optimization variables. There are many other possibilities for objective functions. However, for a transformation of the corresponding optimization problem with feasibility constraints $\mathcal{A} \in F_{\mathcal{A}}$ into an unconstrained optimization problem in Section 3.4.4 it is only necessary that the objective function is convex. This leads to a combinatorial optimization problem:

Lemma 3.7 *The set $F_{\mathcal{A}}$ in (27) of feasible transformation matrices $\mathcal{A} \in F_{\mathcal{A}}$ is a convex polytope. The objective functions I_1 and I_2 are convex in $F_{\mathcal{A}}$. For I_1 and I_2 there are optimal vertices of $F_{\mathcal{A}}$ maximizing the objective functions, i.e. there is a solution $\mathcal{A}$ of the optimization problem with $\mathcal{A} \in v(F_{\mathcal{A}}) \subseteq F'_{\mathcal{A}}$. For the definition of $F'_{\mathcal{A}}$ see (28).*

Proof: The constraints which determine $F_\mathcal{A}$ are linear, therefore $F_\mathcal{A}$ is a convex polytope. The convexity of I_1 is easy to show: The maximum of a set of linear functions is convex and also the sum of these convex functions, Theorem 1.9 in [59]. Convexity of I_2 follows from the fact that the Hessian of $f(a,b) := a^2\, b^{-1}$ has non-negative eigenvalues for $b > 0$, whereas positivity of $b = \mathcal{A}(1,j)$ holds due to Lemma 3.6. I_1 and I_2 are continuous bounded functions, suprema exist. A supremum of a convex function in a closed convex polytope is attained at a vertex of the polytope, see Chapter 3 in [59]. $\qquad\square$

3.4.3 Linear Program

The existence of maxima for I_1 and I_2 has been shown in Lemma 3.7. *Maximization of convex functions*[9] I_1 and I_2 in a linear bounded feasible set is a non-trivial global optimization problem. In the following the convex function I_1 is transformed into a linear function, i.e. optimization can be done by solving a linear program.

Simplex structure. The transformation $\mathcal{A} \in \mathbb{R}^{n_C \times n_C}$ from the matrix of Perron cluster eigenvectors $X \in \mathbb{R}^{s \times n_C}$ to the matrix $\chi_{disc} = X\mathcal{A} \in \mathbb{R}^{s \times n_C}$ of the convex combination factors is linear. Because of the positivity and partition of unity property, the rows of χ_{disc} can be seen as points $\chi_{disc}(i,:) \in \mathbb{R}^{n_C}, i = 1, \ldots, s$, which lie inside a simplex σ_{n_C} spanned by its vertices – the n_C unit vectors $e_1, \ldots, e_{n_C} \in \mathbb{R}^{n_C}$:

$$x \in \sigma_{n_C} \Leftrightarrow \exists \gamma_1, \ldots \gamma_{n_C} \in [0,1], \ x = \sum_{j=1}^{n_C} \gamma_j e_j, \ \sum_{j=1}^{n_C} \gamma_j = 1.$$

A linear mapping $\mathcal{A}^{-1}$ maps the simplex σ_{n_C} to another simplex $\widetilde{\sigma}_{n_C} = \sigma_{n_C} \mathcal{A}^{-1}$ with vertices

$$v_j^\top = e_j^\top \mathcal{A}^{-1} = \mathcal{A}^{-1}(j,:), \quad j = 1, \ldots, n_C. \tag{30}$$

This means, the rows of X as points in $\mathbb{R}^{n_C}$ lie inside a simplex with the vertices $v_1, \ldots, v_{n_C} \in \mathbb{R}^{n_C}$, and the rows of $\mathcal{A}^{-1}$ equal these vertices. Assume (*maximality assumption*) that for each conformation j in the solution of Robust Perron Cluster Analysis there exists a set ξ_i with maximal membership $\chi_{disc}(i,j) = 1$, i.e. $\chi_{disc}(i,:) = e_j^\top$. Then the i-th row of χ_{disc} is a vertex of σ_{n_C}, which implies that the i-th row of X is a vertex $v_j = X(i,:)^\top$ of $\widetilde{\sigma}_{n_C}$. Via (30) this gives us:

$$\mathcal{A}^{-1}(j,:) = X(i,:). \tag{31}$$

In other words, the maximality assumption implies that the vertices v_j can be found among the rows of X. With an appropriate index mapping

$$ind : \{1, \ldots, n_C\} \to \{1, \ldots, s\}$$

---

[9]Equivalent: Minimization of concave functions. This problem is known as "Concave Programming". The general form is NP-hard, see [58, 59].
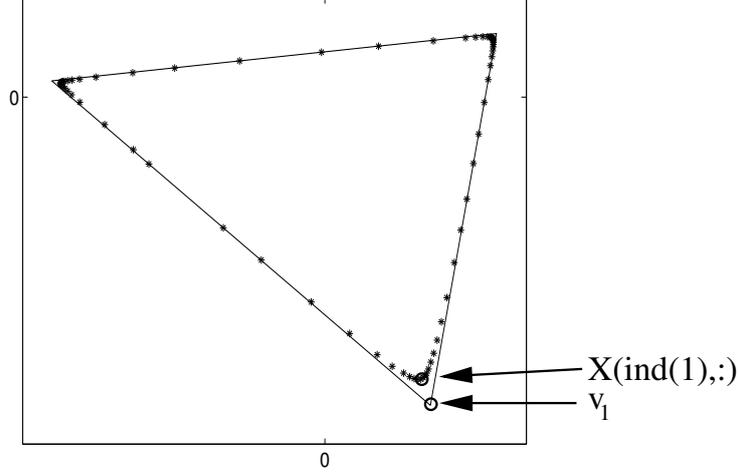
Figure 2: Example for $n_C = 3$, taken from [28]. By omitting the first (constant) component, the eigenvector data $X(l,:) \in \mathbb{R}^3$ for $l = 1,\dots,60$ is projected to $\mathbb{R}^2$. The I_1-optimal simplex $\widetilde{\sigma}_3$ spanned by the vertices $v_1,\dots,v_{n_C}$ is shown. Algorithm 3.8 provides $X(ind(1),:)$ as an approximation of $v_1^\top = \mathcal{A}^{-1}(1,:)$.

we get $v_j = X(ind(j),:)^\top, j = 1,\dots,n_C$. If the maximality assumption nearly holds, see Fig. 2, then the following routine [28] finds an index mapping for an approximation of the simplex $\widetilde{\sigma}_{n_C}$. Its "vertices" are determined successively among the rows of X: Once one has found a vertex subset, the hyperplane of minimal dimension is constructed, which includes these vertices. The next vertex is a point having maximal distance to this hyperplane. The distance can be calculated from an orthogonalization routine.

Algorithm 3.8 Index mapping for approximate vertices of $\widetilde{\sigma}_{n_C}$.

1. *Find the row i^*, which maximizes the norm $\|X(i,:)\|_2, i = 1,\dots,s$.*

   Set $ind(1) = i^*$, $\tilde{v}_1 = X(ind(1),:)$.

   For $i = 1,\dots,s$ set $X(i,:) \leftarrow X(i,:) - \tilde{v}_1$.

2. *FOR $j = 2,\dots,n_C$ find further vertices via:*

   Find the row i^*, which maximizes the norm $\|X(i,:)\|_2, i = 1,\dots,s$.

   Set $ind(j) = i^*$ and $X \leftarrow X/\|X(ind(j),:)\|_2$ for $i = 1,\dots s$.

   Set $\tilde{v}_j = X(ind(j),:)$.

   For $i = 1,\dots,s$ set $X(i,:) \leftarrow X(i,:) - \tilde{v}_j X(i,:)^\top \tilde{v}_j$.

   NEXT loop index

Objective function.   Algorithm 3.8 supplies the indices $ind(1), \ldots, ind(n_C)$ of points $X(ind(j), :), j = 1, \ldots, n_C$, which are nearly vertices of $\widetilde{\sigma}_{n_C}$. In a feasible solution the vertices v_j of $\widetilde{\sigma}_{n_C}$ are mapped to the vertices of σ_{n_C}. In this case, the points $X(ind(j), :) \approx v_j, j = 1, \ldots, n_C$ are nearly mapped to the unit vectors $e_1, \ldots, e_{n_C}$. In other words, for a fixed $j = 1, \ldots, n_C$, the membership $\max_{i=1,\ldots,s} \chi_{disc}(i, j)$ is maximal for $i = ind(j)$. Using this information, I_1 in (29) becomes:

$$I_1(\mathcal{A}) = \sum_{j=1}^{n_C} \sum_{i=1}^{n_C} X(ind(j), i)\, \mathcal{A}(i, j).$$

This is a linear function. Maximization of a linear function over a polytope is known as linear programming and can be solved efficiently via simplex algoritms or via interior point methods, see Karmarkar [68].

3.4.4   Unconstrained Optimization

Solving the optimization problem with objective I_1 via a linear program is a time-consuming procedure, because there may be a large number of constraints. Note that Algorithm 3.8 can also provide an initial guess $\overline{\mathcal{A}}_0 \in \mathbb{R}^{n_C \times n_C}$ for a local optimization method. In the spirit of equation (30), for a feasible solution of the optimization problem we have to find a simplex $\widetilde{\sigma}_{n_C}$ that includes the rows of X as points in $\mathbb{R}^{n_C}$. Only if the convex hull $co(\mathcal{X})$ of the point set $\mathcal{X} := \{X(i, :);\ i = 1, \ldots, s\}$ is already a simplex, the Algorithm 3.8 finds its vertices and the following algorithm computes a feasible starting guess, see Lemma 3.13 and Theorem 3.14 below. In general the starting guess is infeasible.

Algorithm 3.9 Infeasible initial guess for optimization

1. *Determine the index mapping ind() for finding approximate vertices of $\widetilde{\sigma}_{n_C}$ via Algorithm 3.8.*

2. *Using the maximality assumption, the transformation matrix $\overline{\mathcal{A}}_0$ can be calculated via equation (31):*

$$\overline{\mathcal{A}}_0 = \begin{pmatrix} X(ind(1), 1) & \cdots & X(ind(1), n_C) \\ \vdots & & \vdots \\ X(ind(n_C), 1) & \cdots & X(ind(n_C), n_C) \end{pmatrix}^{-1}.$$

3. *In general, $\overline{\mathcal{A}}_0$ provides an infeasible initial guess $\chi_{disc} = X\overline{\mathcal{A}}_0$ for the optimization problem.*

What kind of infeasibility can occur in Algorithm 3.9? Note that the first column of $\overline{\mathcal{A}}_0^{-1}$ is $\overline{\mathcal{A}}_0^{-1}(i, 1) = 1$ for $i = 1, \ldots, n_C$ by construction. Also note that the first column

of X has this property, too, i.e. $X(l, 1) = 1$ for $l = 1, \ldots, s$. Therefore, from the first column of the equation $\chi_{disc} \overline{\mathcal{A}}_0^{-1} = X$ it follows that χ_{disc} meets the partition of unity property in Algorithm 3.9. Only the positivity constraint may be violated. In most of the applications, Algorithm 3.9 provides very good solutions, which are almost feasible. Therefore, further investigations are not neccessary. But in order to be mathematically rigorous, a feasible solution is derived in the following.

From infeasible to feasible solutions.   We now aim at a projection algorithm from an arbitrary matrix $\overline{\mathcal{A}} \in \mathbb{R}^{n_C \times n_C}$ to a feasible transformation matrix $\mathcal{A}$. Because of Lemma 3.7, theoretically, we can restrict our search of transformation matrices to vertices of the feasible set $\mathcal{A} \in v(F_{\mathcal{A}})$. In practice, it is not easy to compute these vertices explicitly. Therefore, one applies Lemma 3.5 and (28) to extend the search to a superset of matrices $\mathcal{A} \in F'_{\mathcal{A}} \supset v(F_{\mathcal{A}})$. It is easy to verify that the following routine computes a feasible transformation matrix $\mathcal{A}$ out of an arbitrary matrix $\overline{\mathcal{A}}$:

Algorithm 3.10 Feasible transformation matrices

   1. For $j = 2, \ldots, n_C$ replace $\overline{\mathcal{A}}(j, 1)$ with $\overline{\mathcal{A}}(j, 1) := -\sum\limits_{i=2}^{n_C} \overline{\mathcal{A}}(j, i)$.

   2. For $i = 1, \ldots, n_C$ replace $\overline{\mathcal{A}}(1, i)$ with $\overline{\mathcal{A}}(1, i) := -\min\limits_{l=1,\ldots,s} \sum\limits_{j=2}^{n_C} \overline{\mathcal{A}}(j, i) \, X(l, j)$.

   3. The sum of the elements of the first row have to be 1 due to Lemma 3.6:

   $$\gamma := \sum_{i=1}^{k} \overline{\mathcal{A}}(1, i).$$

   Therefore, the feasible transformation matrix $\mathcal{A}$ is computed as follows:

   $$\mathcal{A}(j, i) := \frac{\overline{\mathcal{A}}(j, i)}{\gamma}, \quad i, j = 1, \ldots, n_C.$$

Verification of the Algorithm 3.10:   Step 1 assures feasibility of $\mathcal{A}$ w.r.t. equation (28.1) for $j = 2, \ldots, n_C$, which still is true after rescaling in step 3 of the algorithm. Step 2 is also invariant against positive scaling of $\mathcal{A}$ and assures feasibility of $\mathcal{A}$ according to equation (28.2) for $i = 1, \ldots, n_C$. After step 3 $\mathcal{A}$ is feasible w.r.t. the missing condition (28.1) for $j = 1$. Algorithm 3.10 is obviously a projection algorithm: If the input $\overline{\mathcal{A}}$ is feasible w.r.t. (28), then $\mathcal{A} = \overline{\mathcal{A}}$ is returned. I.e. Algorithm 3.10 is a surjective mapping to $F'_{\mathcal{A}}$ including the optimal transformation matrix. $\qquad\square$

In order to get a feasible transformation matrix in Algorithm 3.10 only a specification of the elements $\overline{\mathcal{A}}(j, i)$ for $j, i \neq 1$ is necessary, because the other elements are specified in

step 1 and 2 of the algorithm. Obviously, there are no constraints[10] for the choice of these $(n_C - 1)^2$ elements, because after step 3 *every* input matrix $\overline{\mathcal{A}}$ is mapped to a feasible one. In other words, the constrained optimization with n_C^2 variables can be transformed into an unconstrained optimization problem in $(n_C - 1)^2$ variables:

Algorithm 3.11 Robust Perron Cluster Analysis

1. *Via Algorithm 3.9 compute an infeasible initial guess $\overline{\mathcal{A}}_0$ for a local optimization. Use Algorithm 3.10 to get a feasible initial guess $\mathcal{A}_0$.*

2. *Perform an iterative local optimization $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \ldots$ of I_1 or I_2 starting with $\mathcal{A}_0$. In each step $\mathcal{A}_k \to \mathcal{A}_{k+1}$ of the local optimization routine only update the elements $\mathcal{A}_k(i, j), i, j \neq 1$, without regarding any constraints. Use Algorithm 3.10 to get a feasible matrix $\mathcal{A}_{k+1}$ before evaluating $I_1(\mathcal{A}_{k+1})$ or $I_2(\mathcal{A}_{k+1})$ respectively.*

3. *If a criterion of convergence is satisfied for some $\mathcal{A}_k$: STOP. The solution is $\chi_{disc} = X\mathcal{A}_k$.*

In Robust Perron Cluster Analysis, via Algorithm 3.10 infeasible matrices $\overline{\mathcal{A}}_k$ are transformed into feasible ones $\mathcal{A}_k \in F_{\mathcal{A}}$ using a non-differentiable routine, i.e. for a local optimization procedure only function values $I_1(\mathcal{A}_{k+1})$ or $I_2(\mathcal{A}_{k+1})$ are available. Deuflhard and Weber [28] therefore proposed to use the derivative-free Nelder-Mead simplex method [97, 76] for optimization. From [76]: "[...] The Nelder-Mead algorithm should not be confused with the (probably) more famous simplex algorithm of Dantzig for linear programming; both algorithms employ a sequence of simplices but are otherwise completely different and unrelated[...]" This statement also holds for the simplices used in the Nelder-Mead routine, which have dimension $(n_C - 1)^2$, and $\widetilde{\sigma}_{n_C}$ or σ_{n_C}, which have dimension n_C.

3.4.5 Unique Solutions

The optimization problems for I_1, I_2 are combinatorial *global* optimization problems, in general with different local maxima. We cannot assume that local optimization in Algorithm 3.11 finds a global maximum of the objective functions. In the present section conditions are derived, for which the set of vertices $v(F_{\mathcal{A}})$ of $F_{\mathcal{A}}$ is in such a manner that once one has found any transformation matrix $\mathcal{A} \in v(F_{\mathcal{A}})$ this matrix is globally optimal. In other words, every vertex of $F_{\mathcal{A}}$ leads to the same function value of I_1 or I_2.

---

[10]Exception: For evaluation of I_2 there should not exist any column $i = 2, \ldots, n_C$ with $\overline{\mathcal{A}}(j, i) = 0$ for all $j = 2, \ldots, n_C$.

Definition of uniqueness. As we have seen, the optimal solution of Robust Perron Cluster Analysis is attained at a vertex $\mathcal{A} \in v(F_{\mathcal{A}})$ of the feasible set $F_{\mathcal{A}}$. "Uniqueness of the solution" means that there is only one "real" vertex of $F_{\mathcal{A}}$. In other words, a regular transformation matrix $\mathcal{A} \in v(F_{\mathcal{A}})$ is called *unique* if for every transformation matrix $\mathcal{B} \in v(F_{\mathcal{A}})$ the columns of $\mathcal{A}$ and $\mathcal{B}$ are identical except for permutation. A result of Deuflhard and Weber provides a sufficient condition for uniqueness.

Theorem 3.12 ([28]) *For a matrix $\chi_{disc} \in \mathbb{R}^{s \times n_C}$ and the matrix of Perron cluster eigenvectors $X \in \mathbb{R}^{s \times n_C}$ always three out of the following four conditions are satisfiable:*

i) For all $l = 1, \ldots, s$: $\sum_{i=1}^{n_C} \chi_{disc}(l, i) = 1$,

ii) for all $i = 1, \ldots, n_C$ and $l = 1, \ldots, s$: $\chi_{disc}(l, i) \geq 0$,

iii) $\chi_{disc} = X\mathcal{A}$ with $\mathcal{A}$ regular,

iv) for all $i = 1, \ldots, n_C$ there exists $l \in \{1, \ldots, s\}$ with $\chi_{disc}(l, i) = 1$.

If all four of the conditions hold, then $\chi_{disc} = X\mathcal{A}$ is a solution of the Robust Perron Cluster Analysis, where $\mathcal{A} \in v(F_{\mathcal{A}})$ is unique.

Theorem 3.12 implies that a clustering algorithm in terms of almost characteristic functions always has one "disadvantage". For example:

- The "inner simplex algorithm" by Weber and Galliat [131], which is identical to Algorithm 3.9, provides a clustering, which meets the triple "i),iii),iv)", but the solution may have negative elements, $\chi_{disc} \not\geq 0$.

- The sign-structure PCCA algorithm by Deuflhard et al. [26] provides a crisp clustering, i.e. the triple "i),ii),iv)" is satisfied, but the solution is not a linear combination of eigenvectors, i.e. an important condition for robustness of the clustering does not hold.

- The Robust Perron Cluster Analysis by Deuflhard and Weber [28] meets "i),ii),iii)", where condition "iv)" turns into an optimization problem for I_1. If I_1 is maximal, then the solution is unique.

I_1 as indicator of uniqueness. Whereas Theorem 3.12 provides a sufficient condition for uniqueness, in the following a necessary and sufficient condition is shown, which leads to a better understanding of the relation between the initial guess in Algorithm 3.9 and the objective function I_1. As a pre-requisite, for a characterization of uniqueness we first proof the following

Lemma 3.13 (compare [131]) *For a matrix $X \in \mathbb{R}^{s \times n_C}$ of Perron cluster eigenvectors define the set $\mathcal{X} := \{X(l, :) \in \mathbb{R}^{n_C}; l = 1, \ldots, s\}$. Then, every regular transformation matrix $\mathcal{A} \in v(F_{\mathcal{A}})$ is unique, if and only if the convex hull $co(\mathcal{X})$ of $\mathcal{X}$ is a simplex.*

Proof: Feasibility of $\mathcal{A}$ can be expressed, like in Section 3.4.3, by the condition that the convex simplex $\widetilde{\sigma}_{n_C}$ spanned by the vertices $v_i^\top = \mathcal{A}^{-1}(i, :) \in \mathbb{R}^{n_C}$ is a superset of $\mathcal{X}$. Equivalently, $\widetilde{\sigma}_{n_C} \supset co(\mathcal{X})$. If a point $f \in \mathcal{X}$ meets $f \in \partial \widetilde{\sigma}_{n_C}$, then $f \in \partial co(\mathcal{X})$, because $co(\mathcal{X})$ is the smallest convex set including $\mathcal{X}$.
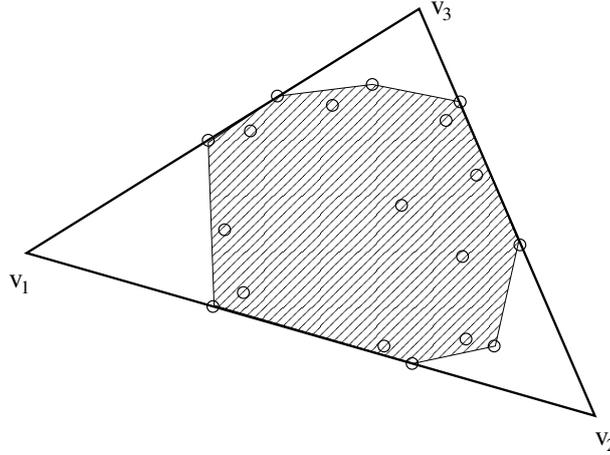
Figure 3: A vertex $\mathcal{A} \in v(F_{\mathcal{A}})$ of the feasible set of transformation matrices is characterized by the fact that each facet of $\widetilde{\sigma}_{n_C}$ includes a facet of $co(\mathcal{X})$ (elements of $\mathcal{X}$ are plotted as circles), where $\widetilde{\sigma}_{n_C}$ is spanned by the vertices $v_i^\top = \mathcal{A}^{-1}(i, :) \in \mathbb{R}^{n_C}$, $i = 1, \ldots, n_C$. The figure shows the case $n_C = 3$ with a projection to $\mathbb{R}^2$.

$\mathcal{A} \in v(F_{\mathcal{A}})$ if and only if $n_C(n_C - 1)$ linear independent inequalities in (27.2) are active, i.e. $n_C(n_C - 1)$ points out of $\mathcal{X}$ are mapped via $\mathcal{A}$ onto facets of σ_{n_C}, where membership according to one cluster is zero. A dimension argument shows that this is the case if and only if each of the n_C facets of $\widetilde{\sigma}_{n_C}$ includes $(n_C - 1)$ points out of $\mathcal{X}$. These points have to be part of $\partial co(\mathcal{X})$, too, and therefore determine a facet of $co(\mathcal{X})$. Because of the convexity of $co(\mathcal{X})$, finally, a vertex $\mathcal{A} \in v(F_{\mathcal{A}})$ is characterized by the fact that a facet of the corresponding simplex $\widetilde{\sigma}_{n_C}$ includes a facet of $co(\mathcal{X})$. Therefore, the choice of such facets is unique (except for permutation) if and only if $co(\mathcal{X})$ is a simplex, see also Figure 3. □

With these preparations, the main result of this section is

Theorem 3.14 *For a transformation matrix $\mathcal{A} \in \mathbb{R}^{n_C \times n_C}$ and a matrix $X \in \mathbb{R}^{s \times n_C}$ of Perron cluster eigenvectors, the following statements are equivalent:*

i) $\mathcal{A}$ is (except for permutation of columns) the result of Algorithm 3.9 and $\mathcal{A} \in F_{\mathcal{A}}$, i.e. the initial guess $\chi_{disc} = X\mathcal{A}$ is feasible.

ii) $\mathcal{A} \in v(F_{\mathcal{A}})$ *is unique.*

iii) $\mathcal{A} \in F_{\mathcal{A}}$ *and* $I_1(\mathcal{A}) = n_C$, *i.e. for every* $i = 1, \ldots, n_C$ *there is an* $l \in \{1, \ldots, s\}$ *with maximal degree of membership* $\chi_{disc}(l, i) = 1$.

In most of the applications, Algorithm 3.9 provides an almost feasible solution of the cluster problem, which means that Theorem 3.14 is not a trivial exception but the generic case, see also [132] and [135].

Proof: "i)$\Rightarrow$ iii)": In Algorithm 3.9 for a feasible solution we get $I_1(\mathcal{A}) = n_C$, because for $i = 1, \ldots, n_C$ maximailty $\chi_{disc}(ind(i), i) = 1$ holds via construction, see Section 3.4.3.

"iii)$\Rightarrow$ ii)": If "iii)" holds, then due to maximality of I_{n_C} there is a feasible transformation matrix $\mathcal{A}$, which maps n_C points $\{v_1, \ldots, v_{n_C}\} \subset \mathcal{X}$ directly to the vertices $e_1, \ldots, e_{n_C}$ of the standard simplex σ_{n_C}, where $\mathcal{X}$ is defined in Lemma 3.13. In this case, the convex hull of $\mathcal{X}$ is a simplex spanned by these vertices $v_1, \ldots, v_{n_C}$. Lemma 3.13 implies that "ii)" holds.

"ii) $\Rightarrow$ i)": Algorithm 3.9 applies Algorithm 3.8 for an identification of so-called extreme points of $\mathcal{X}$. Extreme points are in fact vertices of the convex hull $co(\mathcal{X})$, see Section 3.1.1. in [100]. If the convex hull of $\mathcal{X}$ is a simplex due to Lemma 3.13, then Algorithm 3.8 already finds *all* vertices of the convex hull of $\mathcal{X}$ and therefore $\mathcal{A}$ is feasible, because the rows of $\mathcal{A}^{-1}$ span a simplex including all points of $\mathcal{X}$. Note that a permutation of rows of $\mathcal{A}^{-1}$ is equivalent to a permutation of columns of $\mathcal{A}$. $\qquad\square$

Theorem 3.14 shows that if $F_{\mathcal{A}}$ has only one "real" vertex $\mathcal{A}$, then this transformation matrix can be found via the starting guess algorithm. An indicator of uniqueness is the value of the objective function I_1. See also Section 3.5.3, where it will be shown that this indicator I_1 is an upper bound for the metastability I_2. Especially for the case $n_C = 2$ an important consequence of Theorem 3.14 can be derived:

Corollary 3.15 *For the case* $n_C = 2$, *Algorithm 3.9 provides a feasible unique solution of Robust Perron Cluster Analysis.*

Proof: (see (4.26)-(4.28) in [28]) The solution of Robust Perron Cluster Analysis is $\chi_{disc} = X\mathcal{A}$ with a transformation matrix $\mathcal{A} \in v(F_{\mathcal{A}}) \subset F'_{\mathcal{A}}$. Therefore, if $\mathcal{A}$ is unique for $F'_{\mathcal{A}}$, then it is also unique for $v(F_{\mathcal{A}})$. For the case $n_C = 2$ equation (28) for $F'_{\mathcal{A}}$ is:

$$\mathcal{A}(1,1) = 1 - \mathcal{A}(1,2), \qquad\qquad \mathcal{A}(2,1) = -\mathcal{A}(2,2),$$

$$\mathcal{A}(1,1) = -\min_{l=1,\ldots,s} \mathcal{A}(2,1)X(l,2), \quad \mathcal{A}(1,2) = -\min_{l=1,\ldots,s} \mathcal{A}(2,2)X(l,2). \tag{32}$$

Introducing the following notations

$$a = \max_{l=1,\ldots,s} X(l,2), \quad b = -\min_{l=1,\ldots,s} X(l,2),$$

for $\mathcal{A}(2,2) > 0$ we get a unique solution of (32):

$$\mathcal{A} = \begin{pmatrix} \frac{a}{a+b} & \frac{b}{a+b} \\ -\frac{1}{a+b} & \frac{1}{a+b} \end{pmatrix}.$$

For $\mathcal{A}(2,2) < 0$ we get the same solution, but with a permutation of the columns of $\mathcal{A}$. For $\mathcal{A}(2,2) = 0$, the solution $\mathcal{A}$ is not regular. Summarizing, in the case $n_C = 2$ the feasible transformation matrix $\mathcal{A}$ is unique. A short calculation shows $I_1(\mathcal{A}) = 2 = n_C$, i.e. Theorem 3.14 "iii)$\Rightarrow$ i)" implies that Algorithm 3.9 finds this solution. $\square$

3.5   Indicators for the Number of Clusters

For a given number n_C of clusters the Algorithm 3.11 provides a locally optimal clustering in terms of almost characteristic functions. But how to fix n_C? Here are three methods.

3.5.1   Lower Bound for the Perron Cluster

The simplest method to identify the number of clusters is to prescribe a lower bound for the lowest eigenvalue λ_{n_C} of the Perron cluster. This MINVAL-indicator can be justified by Theorem 3.18, which will be shown in the following, see also Weber [130].

First we have to construct the unperturbed counterpart $\widetilde{P}$ of P. There are many possible construction methods. The following definition is inspired by a disaggregation technique evolved from the Simon-Ando theory [120]. This construction rule preserves the invariant distribution of P.

Definition 3.16 *The* disaggregation *of a stochastic matrix $P = D^{-1}\overline{P}$ is a stochastic, reducible matrix $\widetilde{P} = D^{-1}\overline{\widetilde{P}}$ having n_C uncommunicating blocks with a corresponding index set decomposition $C_1, \ldots, C_{n_C}$. Herein, the non-vanishing entries of $\overline{\widetilde{P}}$ are defined as:*

$$\overline{\widetilde{P}}(i,j) = \begin{cases} \overline{P}(i,j) & , i \neq j \\ \overline{P}(i,i) + \sum\limits_{k=1;\, i \notin C_k}^{n_C} \sum\limits_{l \in C_k} \overline{P}(i,l) & , i = j. \end{cases}$$

This means, the outer-block entries of $\overline{P}$ are added to the diagonal of $\overline{\widetilde{P}}$.

For an eigenvalue estimation of λ_{n_C} we need the following

Lemma 3.17 *(see [44] Corollary 8.1.6.) If A and $A + E$ are symmetric (s,s)-matrices, then*

$$|\lambda_k(A + E) - \lambda_k(A)| \leq \|E\|_2.$$

With these preparations we can estimate the lowest eigenvalue of the Perron cluster from the deviation of P and $\widetilde{P}$.

Theorem 3.18 *For a stochastic matrix $P = D^{-1}\overline{P}$ of a reversible Markov chain, i.e. $\overline{P}$ is symmetric, and its disaggregation $\widetilde{P} = D^{-1}\overline{\widetilde{P}}$ the following estimation holds*

$$\lambda_{n_C}(P) \geq 1 - \|P - \widetilde{P}\|,$$

where $\|\cdot\|$ is any induced sub-multiplicative matrix norm.

This theorem implies that one has to select n_C such that $\lambda_{n_C} \geq 1 - \epsilon$, if ϵ is the maximal allowed deviation $\epsilon = \|P - \widetilde{P}\|$ from block structure.

Proof of Theorem 3.18, see also Weber [130]. By construction $\overline{\widetilde{P}}$ is symmetric and has n_C blocks, i.e. the n_C-th eigenvalue $\lambda_{n_C}(\widetilde{P}) = 1$ is maximal. From (21) we know that

$$D^{-1/2}\overline{P}D^{-1/2} \quad \text{and} \quad D^{-1/2}\overline{\widetilde{P}}D^{-1/2}$$

are symmetric matrices with the same spectra like P and $\widetilde{P}$, respectively. From Lemma 3.17 we get the following estimation for the spectral radius $\rho(P - \widetilde{P})$:

$$
\begin{aligned}
1 - \lambda_{n_C}(P) &= |\lambda_{n_C}(P) - \lambda_{n_C}(\widetilde{P})| \\[4pt]
&= |\lambda_{n_C}(D^{-1/2}\overline{P}D^{-1/2}) - \lambda_{n_C}(D^{-1/2}\overline{\widetilde{P}}D^{-1/2})| \\[4pt]
&\leq \|D^{-1/2}\overline{P}D^{-1/2} - D^{-1/2}\overline{\widetilde{P}}D^{-1/2}\|_2 \\[4pt]
&= \|D^{-1/2}(\overline{P} - \overline{\widetilde{P}})D^{-1/2}\|_2 \\[4pt]
&= \max\{|\lambda_1(D^{-1/2}(\overline{P} - \overline{\widetilde{P}})D^{-1/2})|, |\lambda_s(D^{-1/2}(\overline{P} - \overline{\widetilde{P}})D^{-1/2})|\} \\[4pt]
&= \max\{|\lambda_1(D^{-1}(\overline{P} - \overline{\widetilde{P}}))|, |\lambda_s(D^{-1}(\overline{P} - \overline{\widetilde{P}}))|\} \\[4pt]
&= \max\{|\lambda_1(P - \widetilde{P})|, |\lambda_s(P - \widetilde{P})|\} \\[4pt]
&= \rho(P - \widetilde{P}).
\end{aligned}
$$

The fact that any induced sub-multiplicative matrix norm is an upper bound for the spectral radius of a square matrix completes the proof[11]. $\qquad\square$

---

[11]For arbitrary eigenvalues the proof method of Theorem 3.18 provides $\lambda_j(P) \geq \lambda_j(\widetilde{P}) - 2\|P - \widetilde{P}\|$ for any sub-multiplicative norm. For similar results see also Meerbach et al. [87] Theorem 3.3. Therein, the

Remark 3.19 *For the proof it is only necessary that $\overline{\widetilde{P}}$ is symmetric and has the same row-sums like $\overline{P}$. Therefore, Definition 3.16 is not the only construction rule, which is compatible with Theorem 3.18.* Stochastic complementation *is another example, see [91, 87].*

MINVAL-indicator in practice.   If one knows the maximal allowed error $\epsilon = \|P - \widetilde{P}\|$ a priori, then Theorem 3.18 provides a *necessary* but not sufficient condition for the lowest Perron cluster eigenvalue $\lambda_{n_C} \geq 1 - \epsilon$. The question is: Which value ϵ is reasonable? A simple counter example shows that in the presence of transition states the estimation of Theorem 3.18 is rather bad. Example 2.1 in [131], with $a = b = 0.5$ and $\epsilon = 0.01$, is

$$P = \begin{pmatrix} 0.99 & 0.01 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 0.01 & 0.99 \end{pmatrix}. \tag{33}$$

In this matrix the second row can be seen as a transition state. For $n_C = 2$ we get $\lambda_2(P) = 0.99$, but the construction of $\widetilde{P}$ according to Definition 3.16 at least gives us $\|P - \widetilde{P}\|_1 = 0.51$, i.e. Theorem 3.18 provides the estimation $\lambda_2(P) \geq 0.49$. The bad estimation is implied by the fact that the transition state (2nd row of P) cannot be assigned to one of the other two states keeping ϵ small. In conformation dynamics the occurrence of transition states is generic. In fact, this observation was the reason for introducing a "soft"-clustering method, see [131]. The MINVAL indicator is necessary but not sufficient for fixing n_C. Further investigations are essential. Another motivation for a MINVAL indicator is given in Section 4.3 Theorem 4.4, where a connection between eigenvalues and holding probabilities of conformations is derived.

3.5.2   Spectral Gap

Schütte's conjecture.   Schütte conjectured (Conjecture 4.23 in [111]) that for realistic molecular data the essential spectral radius of the propagator P^τ is very small. This means that in the discrete counterpart P of P^τ there is probably a significant gap between the last Perron cluster eigenvalue λ_{n_C} and the rest of the spectrum. Theorem 4.13 of Huisinga [60] clarifies this conjecture:

Theorem 3.20 *Let $P^\tau : L^1(\pi) \to L^1(\pi)$ denote the propagator corresponding to a stochastic transition function $\wp : \mathbb{R} \times \Omega \times \mathcal{B}(\Omega) \to [0,1]$, where $\mathcal{B}(\Omega)$ is the class of Borel sets in*

---

estimation is
$$\lambda_j(P) \geq \lambda_j(\widetilde{P}) - 2\|P - \widetilde{P}\|_\infty, \quad j = 1, \ldots, s,$$

where $\widetilde{P}$ is constructed like in Definition 3.16 or $\widetilde{P}$ is the so-called *stochastic complement* of P. An alternative proof of Theorem 3.18 uses the well-known result $\|\mathrm{diag}(\lambda_1 - \widetilde{\lambda}_1, \ldots, \lambda_s - \widetilde{\lambda}_s)\| \leq \|A - \widetilde{A}\|$ for Hermitian matrices and any unitarily invariant norm $\|\cdot\|$, which has been shown by Mirsky [95].

Ω. Let μ denote the measure implied by the Boltzmann distribution π. If $p(\cdot, \cdot, \cdot)$ satisfies a μ-a.e. Doeblin-condition, i.e. if there exist constants $\epsilon, \delta > 0$ and $m \in \mathbf{N}$ such that for all $A \in \mathcal{B}(\Omega)$ and μ-almost every $q \in \Omega$

$$\mu(A) \leq \epsilon \Rightarrow \wp^m(\tau, q, A) \leq 1 - \delta,$$

then the essential spectral radius of P^τ is bounded by $r_{ess}(P^\tau) \leq (1 - \delta)^{1/m}$.

This theorem implies that Schütte's conjecture is true, if the probability $\wp(\tau, q, A)$ for a transition in time span τ from $q \in \Omega$ into a set $A \subset \Omega$ of small measure is also small. For example, the essential radius might not be small, if the free energy landscape of the system has a funnel structure with a deep valley. These systems, however, are not very interesting for conformation analysis.

Vice versa, due to the correlation between metastability and the spectrum of P in Theorem 3.4, inside the Perron cluster of eigenvalues there should not be a significant gap. Together with Schütte's conjecture, this assumption implies that it is possible to search for a significant gap inside the spectrum of P in order to identify the number n_C of clusters.

Condition number of Robust PCCA. Spectral analysis is an important field in Markov chain theory and well-investigated. Another justification for the spectral gap indicator is also well-investigated, but usually not mentioned in Perron Cluster Analysis – the condition of the invariant subspace corresponding to the Perron cluster eigenvalues. In most cases the eigenvalues $\lambda_1, \ldots, \lambda_{n_C} \approx 1$ stay close together. This means that the problem of computation of the corresponding eigenvectors is not well-conditioned, see Corollary 7.2.6 in [44]. But in Robust Perron Cluster Analysis we are not interested in each single eigenvector $X_1, \ldots, X_{n_C} \in \mathbb{R}^s$, but only in the invariant subspace spanned by all Perron cluster eigenvectors $X = [X_1, \ldots, X_{n_C}]$, because the solution χ_{disc} is searched in this space, $\chi_{disc} = X\mathcal{A}$. To see that this problem is well-conditioned for a big gap between the last Perron cluster eigenvalue λ_{n_C} and the next one λ_{n_C+1} here is a corresponding adaption of Theorem 8.1.10 in [44]:

Theorem 3.21 *Define the symmetric matrix $A := D^{-1/2} \overline{P} D^{-1/2} \in \mathbb{R}^{s \times s}$ like in equation (21). Sort its eigenvalues $\lambda_1 \geq \ldots \geq \lambda_s$ in descending order. For the corresponding normalized eigenvectors $y_1, \ldots, y_s \in \mathbb{R}^s$ define the matrices $Y_1 := [y_1, \ldots, y_{n_C}] \in \mathbb{R}^{s \times n_C}$ and $Y_2 := [y_{n_C+1}, \ldots, y_s] \in \mathbb{R}^{s \times (s - n_C)}$. If $E \in \mathbb{R}^{s \times s}$ is a symmetric error matrix with*

$$\|E\|_2 \leq \frac{\lambda_{n_C} - \lambda_{n_C+1}}{5}$$

and $E_{21} := Y_2^\top E Y_1$, then there exists a matrix $\hat{Y}_1 \in \mathbb{R}^{s \times n_C}$, such that the columns of $\hat{Y}_1$ define an orthogonal basis for an invariant subspace of $A + E$ with

$$\mathrm{dist}(\mathrm{span}(Y_1), \mathrm{span}(\hat{Y}_1)) \leq \frac{4}{\lambda_{n_C} - \lambda_{n_C+1}} \|E_{21}\|_2.$$

In this inequality $\mathrm{dist}(\mathrm{span}(Y_1), \mathrm{span}(\hat{Y}_1)) := \|Y_2^\top \hat{Y}_1\|_2$ *can be seen as distance measure between two subspaces of* $\mathrm{I\!R}^s$ *and* $4(\lambda_{n_C} - \lambda_{n_C+1})^{-1}$ *as condition number for the sensitivity of* $\mathrm{span}(Y_1)$, *which is small for a big gap between* λ_{n_C} *and* λ_{n_C+1}.

Proof: See Theorem 8.1.10 with equation (8.1.2) and Corollary 8.1.11 in [44]. □

3.5.3 minChi Indicator for Simplex Structure

The next indicator has first been defined in 2002 [131] as *indicator for uniqueness.* If this indicator is zero, then the initial guess in Algorithm 3.9 is feasible and except for permutation a unique solution of the clustering, see also Theorem 3.14. In the view of perturbation analysis [28] and after comparison of Robust Perron Cluster Analysis with graph partitioning methods by Weber et al. [134, 135] the meaning of the following indicator has changed.

Perfect simplex structure. Robust Perron Cluster Analysis is a cluster method, which is in principle a simple linear mapping from eigenvector data into a simplex. Let $\chi_1, \ldots, \chi_{n_C} \in L^2(\pi)$ a set of almost invariant membership functions, $\chi_l : \Omega \to [0, 1]$, with maximal value $\chi_{l,max} \leq 1$. For the metastability of these functions we have (the maximal eigenvalue of P^τ is 1, see Theorem 2.2 *iii)*):

$$\mathcal{P}(l, l) = \frac{\langle \chi_l, P^\tau \chi_l \rangle_\pi}{\langle \chi_l, e \rangle_\pi} \leq \frac{\langle \chi_l, \chi_l \rangle_\pi}{\langle \chi_l, e \rangle_\pi} \leq \frac{\chi_{l,max} \langle \chi_l, e \rangle_\pi}{\langle \chi_l, e \rangle_\pi} = \chi_{l,max}. \tag{34}$$

Therefore, the maximal value $\chi_{l,max}$ of the membership function l is an upper bound for the metastability $\mathcal{P}(l, l)$ of an almost invariant membership function χ_l, i.e. $I_2 \leq I_1$. We aim at a clustering of Ω into a set $\chi_1, \ldots, \chi_{n_C}$ of *maximal* metastable membership functions. Due to (34) a necessary condition is $\chi_{l,max} \approx 1$, or in the discrete case, the matrix χ_{disc} should have a maximal entry near 1 per column. The corresponding rows in χ_{disc} represent points close to vertices of the simplex σ_{n_C}. In other words: A necessary pre-requisite for metastability is an almost perfect simplex structure of the eigenvector data and its linear mapping, see Section 3.4.3.

Initial guess and minChi indicator. In general, the initial guess of Algorithm 3.9 is infeasible. Whereas feasibility w.r.t. the partition of unity is always satisfied, the interval restrictions for $\chi_{disc}(i, j) \in [0, 1], i = 1, \ldots, s$ and $j = 1, \ldots, n_C$, usually are violated, which means that negative entries occur. The minChi-indicator is defined as the minimal entry of χ_{disc} evaluated for the initial guess in step 3. of Algorithm 3.9. If this entry is almost zero, then the initial guess is almost feasible, then $\chi_{l,max} \approx 1$ can be achieved for each membership function, which is a *necessary* condition for high metastability due to equation (34). Perturbation theory obviously implies $\mathrm{minChi} = \mathcal{O}(\epsilon^2)$.

MinChi for $n_C = 2$.   For the case $n_C = 2$ there is always a unique solution of the Robust Perron Cluster Analysis, see Corollary 3.15. Therefore, the minChi-indicator is always zero in this case. One has to take this fact into account for the determination of n_C. If minChi is almost zero for some $2 < n_C \ll s$ then there is a clustering. But if minChi is only zero for $n_C = 2$ and significantly larger for $n_C > 2$, then there has to be another indicator for the decision whether $n_C = 1$ or $n_C = 2$. The results of spectral analysis can be helpful for this decision. A corresponding improvement of the minChi-algorithm given in [134, 135] is the following:

Algorithm 3.22 Determining the number of clusters

> *0. If the second eigenvalue λ_2 of P is $\lambda_2 < \lambda_{min}$ for a predefined lower bound λ_{min}, then there is no metastability inside P, i.e. $n_C = 1$. STOP*

> *1. For $n_C = 2, \ldots, b$, where $2 < b \ll s$: Compute the initial guess for the Robust Perron Cluster Analysis via Algorithm 3.9 and its $minChi(n_C)$-value.*

> *2. Determine the maximal number n_C for which the corresponding minChi-value $minChi(n_C)$ is higher than a predefined lower bound $\theta < 0$ as the number of clusters.*

Unfortunately, there is always a couple of parameters, which have to be predefined in each of the above algorithms, in order to compute the number n_C of clusters. The question, what is the "correct" number of clusters, is not easy to answer. Therefore, the given indicators are only a kind of justification for a special choice of n_C.

3.6   Normalized Cut and Robust Perron Cluster Analysis

There is a deep relation between Robust Perron Cluster Analysis and spectral clustering methods in graph partitioning [134, 126]. Especially the relaxation of the *normalized cut* minimization problem used by Shi and Malik [118] and the Multicut Algorithm by Meila and Shi [89] are very close to the Perron Cluster Analysis method. This relation will be discussed in the present section.

Minimizing the cut.   The symmetric matrix $\overline{P}$ can be interpreted as weight matrix of a weighted and undirected complete graph G, such that $\overline{P}(i,j)$ is defined as the weight of the edge between the vertices i and j of G. The problem of computing almost invariant subsets in configuration space is equivalent with the problem of optimal graph partitioning of G. The straight forward method for an optimal graph partitioning is to minimize the so-called *cut* of the partitioning, see e.g. Wu and Leahy [137]. If V denotes the set of vertices of G the *cut* of a partition $A \subset V$ is defined as

$$\mathrm{cut}(A) := \sum_{i \in A, j \notin A} \overline{P}(i,j).$$

The sum of the cut values of $A_1, \ldots, A_{n_C}$, which is an n_C-way graph partitioning of G, can be seen as a special norm of the deviation $\overline{\overline{P}} - \overline{P}$, see Section 3.5.1. This value should be minimal for an optimal clustering. Although this optimization problem can be solved, the results are bad, because the minimal cut algorithm tends to separate low weighted parts of G first and to lead to an unbalanced clustering, see [118, 137].

Normalized cut. In order to get a balanced partitioning, the *normalized cut* value of a partitioning is minimized, which is defined as

$$\text{Ncut} := \sum_{i=1}^{n_C} \frac{\text{cut}(A_i)}{\text{assoc}(A_i)},$$

where $\text{assoc}(A_i)$ is the *association value* of A_i, i.e. the sum of the rows of $\overline{P}$ assigned to A_i:

$$\text{assoc}(A_i) := \sum_{j \in A_i} \sum_{k=1}^{s} \overline{P}(j, k).$$

In [118] it has been shown that minimizing the normalized cut value for a weighted graph leads to a balanced n_C-way graph partitioning of G. Unfortunately, solving this problem is NP-complete.

Relaxation of normalized cut. The main result in [118] is: Normalized cut can be relaxed, i.e. Shi and Malik transformed the crisp partitioning algorithm into a partitioning with almost characteristic functions (the same procedure which led from the Perron Cluster Analysis in [26] to the Robust Perron Cluster Analysis in [28]). Then they showed that the solution of this problem for $n_C = 2$ is a linear combination of the eigenvectors corresponding to the lowest two eigenvectors of the following eigenvalue problem:

$$(D - \overline{P})x = \overline{\lambda} D x.$$

A short calculation shows that this eigenvalue problem is equivalent to

$$P x = (1 - \overline{\lambda}) x.$$

Via substitution $\lambda = 1 - \overline{\lambda}$ this turns out to be the eigenvalue problem solved for the Robust Perron Cluster Analysis. In other words: For $n_C = 2$ the Robust Perron Cluster Analysis exactly solves the relaxation of the normalized cut optimization problem.

NP-completeness. Due to the NP-completeness proof for the normalized cut minimization by Papadimitrou in 1997 (see Appendix of [118]) we cannot assume a polynomial-time algorithm for solving this optimization problem in a crisp sense. This is a mathematical reason for changing the point of view in conformation dynamics from strict clustering, i.e. conformations as almost invariant sets, to its relaxation, i.e. conformations as almost characteristic functions.

For the case $n_C > 2$. For an n_C-way clustering with more than two clusters, Shi and Malik suggest to use the sign structure of the last n_C eigenvectors of the above eigenvalue problem. This is very similar to the examination of the sign structure like in the former Perron Cluster Analysis defined by Deuflhard et al. [26]. Note that the eigenvectors used in both algorithms are identical. In [28] Deuflhard and Weber have shown that this kind of algorithm is not robust. In terms of graph partitioning: We get into trouble if we go back from the relaxed normalized cut optimization problem to a crisp clustering. For $n_C > 2$ Meila and Shi [89] suggest the so-called Multicut Algorithm in which they compute the first n_C eigenvectors of the stochastic matrix P and consider the rows of the eigenvector matrix X as points in $\mathbb{R}^{n_C}$, which is equivalent to the proceeding in Robust Perron Cluster Analysis, see Section 3.4.3. But after this step the two methods are different. Meila and Shi simply apply a standard clustering method to this point set, whereas Robust Perron Cluster Analysis makes use of the simplex structure of the eigenvector data. That a simplex structure is necessary for clustering into metastable conformations has been shown in Section 3.5.3.

4   Basic Concept of the Meshless Algorithm

In the following a set-based discretization of the conformational space will be replaced by a global Galerkin discretization. This will lead to the occurence of a *mass matrix* in the discretization of P^τ.

4.1   Global Galerkin Discretization

So far, the solution of the Robust Perron Cluster Analysis assigns degrees of membership to subsets of the configuration space Ω. The discretization of Ω into subsets in order to define a membership basis $\xi_1, \ldots, \xi_s$ often leads to methods depending on meshes.

- The simplest approach for a discretization of Ω is based on a uniform mesh, which suffers from an exponentially increasing number of discretization boxes w.r.t. the number of dimensions: "Curse of dimensionality", see Bellmann [8].

- The "successive Perron Cluster Analysis of dihedrals" by Cordes et al. [17] uses a uniform interval discretization of single intramolecular coordinates. In order to avoid the curse of dimensionality, this method assumes that single internal coordinates can be used in order to decompose Ω into metastable subsets. This assumption may be wrong for some molecules.

- A different discretization approach of Kohonen [73] is based on prototyping, i.e. *particles* $\{q_1, \ldots, q_s\} \subset \Omega$ are used in order to represent a continuous distribution in the high-dimensional space Ω. This method was the basis for Galliats self-organizing box maps (SOBM) [39, 40]. In SOBM the Kohonen particles become centers of boxes for a discretization of Ω. The problem of this method is to keep balance between small overlap of these boxes and preferably complete overlay of the configuration space, which works hardly for high-dimensional spaces.

All these methods use meshes (boxes, intervals) in some way, which can cause the curse of dimensionality. For the solution of high-dimensional partial differential equations via finite element methods this problem is also known, see e.g. [79, 45, 115, 80]. In the last decade meshless particle methods have become of more and more interest, see Li and Liu [79]. Similar to the self-organizing maps of Kohonen, methods including a non-deterministic generation of particles are assumed to be able to break the curse of dimensionality, see Appendix A.3. In the following the meshless discretization approach, i.e. a global Galerkin discretization of P^τ, will be applied to conformation dynamics.

From set-based to function-based discretization.   The set-based discretization of Ω can also be seen as a function-based discretization of Ω with a basis consisting of piecewise constant (characteristic) functions $\xi = [\xi_1, \ldots, \xi_s]$. The conceptual change of cluster

identification from configuration space to function space can be extended to an arbitrary and non-orthogonal set of basis functions, which is the main topic of this thesis. The key is that we do not need orthogonality of ξ or any set-based descriptions of conformations in order to compute important informations for drug design, like partition functions, coupling matrices or visualization of conformations, see Section 3.1 and Fig. 12 on page 95.

For an arbitrary membership basis in Definition 3.1, the condition that χ_l should be almost invariant w.r.t. the propagator P^τ, can be written as

$$\chi_l \approx P^\tau \chi_l$$

$$\Leftrightarrow \quad \sum_{j=1}^{s} \xi_j \, \chi_{disc}(j,l) \approx P^\tau \sum_{j=1}^{s} \xi_j \, \chi_{disc}(j,l)$$

$$\Rightarrow \quad \int_\Omega \sum_{j=1}^{s} \xi_i(q) \, \xi_j(q) \, \chi_{disc}(j,l) \, \pi(q) \, dq \approx \int_\Omega \sum_{j=1}^{s} \xi_i(q) \, P^\tau \xi_j(q) \, \chi_{disc}(j,l) \, \pi(q) \, dq$$

$$\Leftrightarrow \quad \sum_{j=1}^{s} \langle \xi_i, \xi_j \rangle_\pi \, \chi_{disc}(j,l) \approx \sum_{j=1}^{s} \langle \xi_i, P^\tau \xi_j \rangle_\pi \, \chi_{disc}(j,l), \quad i = 1, \ldots, s$$

$$\Leftrightarrow \quad \overline{S} \chi_{disc} \approx \overline{P} \chi_{disc}, \tag{35}$$

where the mass matrix $\overline{S}$ is defined as

$$\overline{S}(i,j) := \langle \xi_i, \xi_j \rangle_\pi, \quad i,j = 1, \ldots, s,$$

and the Galerkin discretization $\overline{P}$ of the transfer operator is defined as

$$\overline{P}(i,j) := \langle \xi_i, P^\tau \xi_j \rangle_\pi, \quad i,j = 1, \ldots, s.$$

$\overline{S}$ is symmetric by construction, whereas $\overline{P}$ is symmetric because P^τ is self-adjoint w.r.t. the Boltzmann distribution π. In most cases, $\overline{S}$ and $\overline{P}$ are positive (semi-)definite, but they need not be. The next Lemma presents some properties of $\overline{S}$ and $\overline{P}$.

Lemma 4.1 *The row-sums of $\overline{P}$ and $\overline{S}$ are identical. For $i = 1, \ldots, s$:*

$$w_i = \sum_{j=1}^{s} \overline{P}(i,j) = \sum_{j=1}^{s} \overline{S}(i,j) = \langle \xi_i \rangle_\pi.$$

If $\overline{P}$ and $\overline{S}$ are positive, then the vector $w = (w_1, \ldots, w_s)^\top \in \mathbb{R}_+^s$ is the unique positive left eigenvector with $\|w\|_1 = 1$ of the stochastic matrices

$$S = D^{-1} \overline{S} \quad \text{and} \quad P = D^{-1} \overline{P},$$

where $D = \mathrm{diag}(w_1, \ldots, w_s)$.

Proof: The identity of the row-sums of $\overline{P}$ and $\overline{S}$ is a simple consequence of the partition of unity property of ξ and $P^\tau e = e$ for the constant function $e : \Omega \to \{1\}$. It follows that the corresponding diagonal matrix D transforms $\overline{P}$ and $\overline{S}$ into stochastic matrices. The uniqueness of the eigenvector w follows from the scaling condition $\|w\|_1 = 1$ and Perron's Theorem, see Theorem 1.2.2. in [7], where $w^\top P = w^\top$ and $w^\top S = w^\top$ respectively. $\qquad \square$

Construction of a membership basis. The choice of the membership basis ξ is very important for meshless conformation dynamics. It depends on technical details of meshless cluster identification and Hybrid Monte Carlo sampling. Therefore, more theory about the construction of a suitable membership basis follows in Section 5.1.

4.2   Meshless Robust Perron Cluster Analysis

Generalized eigenvalue problem. For the Robust Perron Cluster Analysis only the eigenvector data $X \in \mathbb{R}^{s \times n_C}$ for the first n_C eigenvalues is important. In meshless methods, i.e. for a non-orthogonal membership basis, however, the eigenvalue problem becomes generalized, see equation (35). In this situation, the input data X for the clustering is obtained from the solution of

$$\overline{S} X \Lambda = \overline{P} X, \qquad (36)$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_C})$ is the matrix of Perron cluster eigenvalues. In this case it is important to analyze whether $\overline{S}$ is positive definite, strictly positive definite or positive semi-definite.

$\overline{S}$ positive definite. If $\overline{S}$ is positive definite, then the following Lemma can be applied:

Lemma 4.2 *(Corollary 8.7.2 in [44]) In the situation of (36) let $\overline{P}$ be symmetric and $\overline{S}$ symmetric positive-definite, then there exists a nonsingular $X = [x_1, \ldots, x_s] \in \mathbb{R}^{(s,s)}$ such that both $X^\top \overline{P} X = \mathrm{diag}(a_1, \ldots, a_s)$ and $X^\top \overline{S} X = \mathrm{diag}(b_1, \ldots, b_s)$ are diagonal and real-valued. Moreover,*

$$\overline{P} x_i = \lambda_i \overline{S} x_i$$

for $i = 1, \ldots, s$ and $\lambda_i = a_i/b_i$.

Because of Lemma 4.2, the numerical approximation $\overline{P}_{appr.}$ of $\overline{P}$ should be replaced by a symmetric one, e.g. $\overline{P}_{appr.} \to 0.5\,(\overline{P}_{appr.} + \overline{P}_{appr.}^\top)$, before applying further computations. Recall that $\overline{P}$ is symmetric due to self-adjointness of P^τ. If $\overline{S}$ is a diagonally dominant matrix, then Lemma 4.2 implies the following

Corollary 4.3 *In the above situation let $\xi_1, \ldots, \xi_s : \Omega \to [0,1]$ such that $\overline{P}$ and $\overline{S}$ are symmetric and $\overline{S}$ is diagonally dominant, then the generalized eigenvalue problem (36) has a real-valued spectrum. If, additionally, $\overline{P}$ is diagonally dominant, then the spectrum is positive.*

Proof: In the corollary, $\overline{S}$ is a nonnegative diagonally dominant matrix. From Gerschgorin estimation for its eigenvalues it follows that $\overline{S}$ is positive-definite, i.e. Lemma 4.2 holds and $b_1, \ldots, b_s > 0$. If $\overline{P}$ is also positive-definite, then $a_1, \ldots, a_s > 0$ and $\lambda_i = a_i/b_i$ is positive. $\square$

Corollary 4.3 means that the generalized eigenvalue problem has a real-valued spectrum, if $\xi_1, \ldots, \xi_s$ is a membership basis, for which

$$\langle \xi_i, \xi_i \rangle_\pi > 0.5 \langle \xi_i \rangle_\pi, \quad i = 1, \ldots, s,$$

and it has a positive spectrum, if additionally

$$\langle \xi_i, P^\tau \xi_i \rangle_\pi > 0.5 \langle \xi_i \rangle_\pi, \quad i = 1, \ldots, s.$$

$\overline{S}$ **strictly positive definite.** The condition of the eigenvalue problem (36) may be poor, although the numerical approximation of $\overline{P}$ is symmetrized. The reason is that $\overline{S}$ may have rows i, which almost vanish: The i-th row-sum of $\overline{S}$ equals the weight $\langle \xi_i \rangle_\pi$ of the i-th membership function ξ_i, which may be low. Almost vanishing rows also means that $\overline{S}$ is almost singular. For the generalized eigenvalue problem this may lead to a poor condition number w.r.t. the eigenvalues. Defining symmetric $\mathcal{O}(\epsilon)$-error matrices $\Delta \overline{P}, \Delta \overline{S} \in \mathbb{R}^{s \times s}$, for an eigenvalue λ of the generalized eigenvalue problem (36) with left eigenvector $x \in \mathbb{R}^s$ and right eigenvector $y \in \mathbb{R}^s$, where $y^\top x = 1$, the absolute error $\Delta \lambda$ of the eigenvalue (presumed that the perturbed eigenvalue problem is solvable) can be determined via

$$(\overline{P} + \Delta\overline{P})(x + \Delta x) = (\lambda + \Delta\lambda)(\overline{S} + \Delta\overline{S})(x + \Delta x)$$

$$\Rightarrow \Delta\lambda = \frac{y^\top \Delta\overline{P}\, x - \lambda y^\top \Delta\overline{S}\, x + y^\top \Delta\overline{P}\, \Delta x - \lambda y^\top \Delta\overline{S}\, \Delta x}{y^\top \overline{S}\, x + y^\top \overline{S}\, \Delta x + y^\top \Delta\overline{S}\, x + y^\top \Delta\overline{S}\, \Delta x}$$

$$= \frac{y^\top \Delta\overline{P}\, x - \lambda y^\top \Delta\overline{S}\, x}{y^\top \overline{S}\, x} + \mathcal{O}(\epsilon^2). \tag{37}$$

From this equation it is clear that the denominator may be small, i.e. the absolute error of the eigenvalue λ may be high, if $\overline{S}$ is almost singular. Numerical problems do not occur if we assume that $\overline{S}$ is diagonally dominant, such that

$$\langle \xi_i, \xi_i \rangle_\pi > (0.5 + \delta) \langle \xi_i \rangle_\pi, \quad i = 1, \ldots, s, \tag{38}$$

with $\delta > 0$. A Gerschgorin estimation yields

$$\lambda(S) \geq \Big(\sum_{j=1}^s \overline{S}(i,j) \Big)^{-1} \Big(\overline{S}(i,i) - \sum_{j \neq i} \overline{S}(i,j) \Big)$$

$$
\begin{aligned}
&= \left(\sum_{j=1}^{s} \overline{S}(i,j) \right)^{-1} \left(2\overline{S}(i,i) \right) - 1 \\
&> \left(\sum_{j=1}^{s} \overline{S}(i,j) \right)^{-1} \left(2(0.5+\delta) \sum_{j=1}^{s} \overline{S}(i,j) \right) - 1 \\
&= 2\delta.
\end{aligned}
\tag{39}
$$

Condition (38) avoids singularity of $\overline{S}$ and leads to a better error dependence in equation (37).

Detour: Objective function. Like in Definition 3.2, the coupling matrix $\mathcal{P} \in \mathbb{R}^{n_C \times n_C}$ for a clustering $\chi_{disc} \in \mathbb{R}^{s \times n_C}$ is given as

$$
\mathcal{P} = \left(\frac{\langle \chi_i, P^\tau \chi_j \rangle_\pi}{\langle \chi_i \rangle_\pi} \right)_{i,j=1,\dots,n_C} = \mathrm{diag}(\widetilde{w}_1, \dots, \widetilde{w}_{n_C})^{-1} \chi_{disc}^\top \overline{P} \chi_{disc}.
$$

In this equation, $\widetilde{w}_i$ are the partition functions defined in (17). Let $\overline{D} := \mathrm{diag}(\widetilde{w}_1, \dots, \widetilde{w}_{n_C})$. The solution X of the eigenvalue problem (36) can be rescaled, such that $X^T \overline{S} X = I_{n_C}$, see Lemma 4.2. With the generalized eigenvalue problem (36) and $\mathcal{A}^\top \mathcal{A} = \chi_{disc}^\top \overline{S} \chi_{disc}$ we get:

$$
\begin{aligned}
\mathrm{trace}(\mathcal{P}) &= \mathrm{trace}(\overline{D}^{-1} \chi_{disc}^\top \overline{P} \chi_{disc}) \\
&= \mathrm{trace}(\overline{D}^{-1} \mathcal{A}^\top X^\top \overline{P} X \mathcal{A}) \\
&= \mathrm{trace}(\overline{D}^{-1} \mathcal{A}^\top X^\top \overline{S} X \Lambda \mathcal{A}) \\
&= \mathrm{trace}(\overline{D}^{-1} \mathcal{A}^\top \Lambda \mathcal{A}) \\
&= \mathrm{trace}(\underbrace{\overline{D}^{-1} \chi_{disc}^\top \overline{S} \chi_{disc}}_{=\mathcal{S}} \mathcal{A}^{-1} \Lambda \mathcal{A}) \\
&= \mathrm{trace}(\mathcal{S} \mathcal{A}^{-1} \Lambda \mathcal{A}).
\end{aligned}
\tag{40}
$$

In this equation, the expression $\mathrm{trace}(\mathcal{P}) = \mathrm{trace}(\overline{D}^{-1} \mathcal{A}^\top \Lambda \mathcal{A})$ shows that the computation of I_2 is exactly the same as in Section 3.4.2. Note that also Lemma 3.6 holds. The last line in (40) shows that like in Theorem 3.4 the trace of $\mathcal{P}$ is bounded by the trace of $\Lambda \in \mathbb{R}^{n_C \times n_C}$.

$\overline{S}$ positive semi-definite. If $\overline{S}$ is positive semi-definite, then computation of the eigenvector data according to (36) may be ill-conditioned. We have to look for an alternative computation of X. Like in the proof of Theorem 3.4 equation (24), the reason for the difference $\mathrm{trace}(\mathcal{P}) \neq \mathrm{trace}(\Lambda)$ in (40) is the stochastic matrix $\mathcal{S} \in \mathbb{R}^{n_C \times n_C}$:

$$
\mathcal{S} = \left(\frac{\langle \chi_i, \chi_j \rangle_\pi}{\langle \chi_i \rangle_\pi} \right)_{i,j=1,\dots,n_C}.
\tag{41}
$$

The diagonal of this stochastic matrix can be seen as a *measure of crispness* of the conformations χ_i for $i = 1, \ldots, n_C$, see also equation (43). Due to the fact that the spectrum of P^τ has an upper bound $\lambda_1 = 1$, metastability is not only bounded by the spectrum Λ but also by the crispness of the corresponding conformations $\chi_i, i = 1, \ldots, n_C$:

$$
\begin{aligned}
\text{trace}(\mathcal{P}) = \sum_{i=1}^{n_C} \frac{\langle \chi_i, P^\tau \chi_i \rangle_\pi}{\langle \chi_i \rangle_\pi} &= \sum_{i=1}^{n_C} \frac{\langle \chi_i, P^\tau \chi_i \rangle_\pi}{\langle \chi_i \rangle_\pi} \frac{\langle \chi_i, \chi_i \rangle_\pi}{\langle \chi_i, \chi_i \rangle_\pi} \\
&= \sum_{i=1}^{n_C} \frac{\langle \chi_i, P^\tau \chi_i \rangle_\pi}{\langle \chi_i, \chi_i \rangle_\pi} \frac{\langle \chi_i, \chi_i \rangle_\pi}{\langle \chi_i \rangle_\pi} \\
&\leq \sum_{i=1}^{n_C} \frac{\langle \chi_i, \chi_i \rangle_\pi}{\langle \chi_i \rangle_\pi} = \sum_{i=1}^{n_C} \mathcal{S}(i,i) = \text{trace}(\mathcal{S}). \qquad (42)
\end{aligned}
$$

This means that metastability can only be high, if

$$
\frac{\langle \chi_l, \chi_l \rangle_\pi}{\langle \chi_l, e \rangle_\pi} \approx 1 \quad \Leftrightarrow \quad \langle \chi_l, \chi_l \rangle_\pi - \langle \chi_l, e \rangle_\pi \approx 0, \qquad (43)
$$

for all conformations $\chi_l, l = 1, \ldots, n_C$.

Geometrical clustering. In *geometrical clustering* one computes the eigenvector data X for the Robust Perron Cluster Analysis via

$$
SX = X\Lambda \qquad (44)
$$

instead of solving the generalized eigenvalue problem (36). The spectrum of (44) is real-valued, because $\overline{S}$ is symmetric by construction. There is an advantage of using the geometrical cluster method. If one is only interested in partition functions and not in transition probabilities[12], see (17), then one can avoid the numerical computation of $\overline{P}$, because all information about statistical weights is contained in $\overline{S}$ also. This approach halves the numerical effort, which will be seen later in Figure 4 on page 62 below.

Dynamical clustering. In the case of an ill-conditioned generalized eigenvalue problem, for a so-called *dynamical clustering* the stochastic eigenvalue problem

$$
PX = X\Lambda \qquad (45)
$$

---

[12]Theoretically, partition functions provide all thermodynamically important informations about the molecular system. There is also a concept for the calculation of transition rates on the basis of partition functions via the famous *Eyring formula*, which can be found in textbooks on physical chemistry. Note, however, that $\widetilde{w}_j = \int_\Omega \chi_j(q)\,\pi(q)\,dq$ is only the spatial part of the partition function with normalized Boltzmann factor π.

with $P = D^{-1}\overline{P}$ and the symmetrized matrix $\overline{P}$ is solved instead of (36). The whole toolbox of Section 3 is available in this case. The following equations show that the difference between (36) and (45) is low in terms of almost characteristic functions. Assume that χ_{disc} is computed via Robust Perron Cluster Analysis for X of (45), i.e. $DX\Lambda = \overline{P}X$, with restricted matrices $X \in \mathbb{R}^{s \times n_C}$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_C})$, then

$$
\begin{aligned}
X^\top \overline{S} X \Lambda &= X^\top \overline{S} X \Lambda - X^\top D X \Lambda + X^\top \overline{P} X \\
(1^*) &= X^\top \overline{S} X \Lambda - X^\top \mathrm{diag}(\overline{S}e) X \Lambda + X^\top \overline{P} X \\
&= X^\top (\overline{S} - \mathrm{diag}(\overline{S}e)) X \Lambda + X^\top \overline{P} X \\
&= \mathcal{A}^{-1\top} \underbrace{\chi_{disc}^\top (\overline{S} - \mathrm{diag}(\overline{S}e)) \chi_{disc}}_{\approx 0} \mathcal{A}^{-1} \Lambda + X^\top \overline{P} X \\
(2^*) &\approx X^\top \overline{P} X.
\end{aligned}
\tag{46}
$$

In (1^*) it is used that the diagonal matrix $D = \mathrm{diag}(\overline{S}e)$ can be computed either from the row-sums of $\overline{P}$ or from $\overline{S}$ due to Lemma 4.1. For the approximation in (2^*) equation (43) is used, which is necessary if χ_l is metastable. Solving equation (45) in order to get input data for the Robust Perron Cluster Analysis is therefore a useful alternative to (36).

4.3 Metastability vs. Holding Probabilities

4.3.1 Infinitesimal Generator

This subsection includes some introductory remarks for Section 4.3. In equation (13) the propagator P^τ is constructed according to a short-time Hamiltonian dynamics (MD) simulation, which leads to $L^2(\pi)$-self-adjointness of P^τ. The time dependence of P^τ is not very realistic, however. In a realistic model of molecular transition probabilities in a heat bath, one would expect that the underlying Markov process is homogenous and that the famous Chapman-Kolmogorov equation

$$
P^{(\tau_1 + \tau_2)} = P^{\tau_2} \circ P^{\tau_1}
\tag{47}
$$

for different simulation lengths $\tau_1, \tau_2 > 0$ is satisfied. This is not the case for Hamiltonian dynamics with randomized start momenta, cf. [60]. Briefly stated, the disadvantage is that in the definition (13) of P^τ the momenta are only initial values for the dynamics equations, whereas in reality, the heat bath contact of the molecule exchanges kinetic energy all the time, which leads to the Chapman-Kolmogorov equation. In order to circumvent this problem, one has to replace the MD-part with a dynamics having an $L^2(\pi)$-self-adjoint propagator P^τ which meets (47) and which has the invariant density μ_{can}. An example for this dynamics is the so-called Smoluchowski equation of motion [60], which is the high-friction limit solution of the well-known Langevin equation. For this kind of dynamics, the Chapman-Kolmogorov property (47) implies semigroup properties for the propagator P^τ [71, 60] and the existence of an infinitesimal generator $\mathcal{L}$ with

$$
P^\tau f = \exp(\tau \mathcal{L}) f
\tag{48}
$$

for $f \in L^1(\mu)$. By this equation, for a transition matrix $\mathcal{P}$, which is a discretization of P^τ, there exists a transition rate matrix $\overline{\mathcal{L}}$, which is a discretization of $\mathcal{L}$. With Hamiltonian dynamics we do not have an infinitesimal generator in the strict sense, but we can simply *define* $\overline{\mathcal{L}}$ via $\mathcal{P} = \exp(\tau \overline{\mathcal{L}})$ as corresponding rate matrix. This is possible, because we do not take different time-steps τ into account, i.e. we do not need the correctness of (48) for arbitrary τ.

4.3.2   Explicit Upper and Lower Bounds for Metastability

A problem for finding explicit lower bounds for trace$(\mathcal{P})$ in equation (40) is the fact that the trace-function is not multiplicative. For explicit bounds in this case see Huisinga and Schmidt [62]. The following theorem is based upon (40), but instead of metastability in terms of trace$(\mathcal{P})$ a similar measure of quality is used: If metastability is high, then the determinant $\det(\mathcal{P})$ is nearly 1.

Theorem 4.4 *If $\mathcal{S}$ in (41) is diagonally dominant*[13]*, then for the coupling matrix $\mathcal{P}$ the following estimation holds*

$$a \prod_{i=1}^{n_C} \lambda_i \leq \det(\mathcal{P}) \leq \prod_{i=1}^{n_C} \lambda_i,$$

where

$$a = (\min_{i=1,\dots,n_C} \{2\mathcal{S}(i,i) - 1\})^{n_C}.$$

Proof: From equation (40) we get

$$
\begin{aligned}
\det(\mathcal{P}) &= \det(\mathcal{S}\,\mathcal{A}^{-1}\,\Lambda\,\mathcal{A}) \\
&= \det(\mathcal{S})\,\det(\mathcal{A}^{-1})\,\det(\Lambda)\,\det(\mathcal{A}) \\
&= \det(\mathcal{S})\,\det(\Lambda) \\
&= \det(\mathcal{S})\prod_{i=1}^{n_C} \lambda_i.
\end{aligned}
\tag{49}
$$

As the determinant of $\mathcal{S}$ is the product of its eigenvalues, a simple Gerschgorin estimation implies

$$(\min_{i=1,\dots,n_C} \{2\mathcal{S}(i,i) - 1\})^{n_C} \leq \det(\mathcal{S}) \leq 1.$$

$\square$

Remark 4.5 *Equation (49) clarifies the necessary and sufficient conditions for metastability of a set of almost characteristic functions. Since $\det(\mathcal{P}) = \det(\mathcal{S})\det(\Lambda)$, we need*

---

[13]This is a natural condition, because the trace of $\mathcal{S}$ is an upper bound for the metastability, see equation (42).

a good discretization X of the eigenfunctions of P^τ in order to maximize $\det(\Lambda)$ and a linear combination $\chi = X\mathcal{A}$, which is as crisp as possible in order to maximize $\det(\mathcal{S})$. The term $\det(\Lambda)$ motivates the spectral gap indicator for the number of clusters in Section 3.5.2, whereas $\det(\mathcal{S})$ motivates the minChi indicator for the simplex structure in Section 3.5.3.

Perturbation analysis. In contrast to Theorem 3.4, Theorem 4.4 allows for a direct computation of upper and lower bounds for a metastability measure in terms of the conformations χ_i, but it can also be used for a perturbation analysis: If $\mathcal{S} = I + \mathcal{O}(\epsilon^2)$ like in Theorem 3.4, then $\det(\mathcal{S}) = 1 + \mathcal{O}(\epsilon^2)$ and therefore the results of perturbation analysis can directly be inserted into equation (49) of the above proof. For an application of Theorem 4.4 in practice, a *physical interpretation* for $\det(\mathcal{P})$ is missing, which will be motivated now.

Monomolecular transformations. Conformational changes can be modeled as mono-molecular reactions. For all pairs of conformations $\chi_1, \ldots, \chi_{n_C}$ we have a reaction of the type

$$\chi_i \rightharpoonup \chi_j, \quad i, j = 1, \ldots, n_C,$$
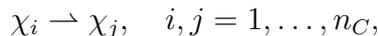
with a corresponding time-independent transition rate $\overline{\mathcal{L}}(i, j) > 0$ for $i \neq j$ and $\overline{\mathcal{L}} \in \mathbb{R}^{n_C \times n_C}$. The diagonal elements of $\overline{\mathcal{L}}$ are the negative exit rates, i.e.

$$\overline{\mathcal{L}}(i, i) = -\sum_{j \neq i} \overline{\mathcal{L}}(i, j).$$

For a concentration vector $x(t) \in \mathbb{R}^{n_C}$, where the i-th component of $x(t)$ is the concentration of conformation χ_i in the ensemble at time t, the following differential equation describes the reaction process:

$$\frac{\partial x(t)}{\partial t} = \overline{\mathcal{L}}^\top x(t). \tag{50}$$

For $t \to \infty$ the equilibrium concentration is $x(\infty) \propto (\widetilde{w}_1, \ldots, \widetilde{w}_{n_C})^\top$. Due to metastability equation (50) is a slow process.

Markov process. Let $\bar{x}$ be a realization of equation (50) for one single molecule. This molecule has a certain state $\overline{x}(t) \in \{1, \ldots, n_C\}$ at time t. The matrix $\overline{\mathcal{L}}$ generates the corresponding continuous-time Markov jump process, which is metastable if $\overline{\mathcal{L}} \approx 0$, i.e. $\mathrm{trace}(\overline{\mathcal{L}}) \approx 0$. From theory of Markov processes, see eq. (4.16) in [71], the matrix $\mathcal{P}$ of transition probabilities according to τ time units can be calculated as

$$\mathcal{P} = \exp(\overline{\mathcal{L}}\tau), \tag{51}$$

where $\overline{\mathcal{L}}$ is called the infinitesimal generator of $\mathcal{P}$. With Theorem 4.4 one can transfer the point of view from metastability and transition matrices to exit rates and transition rate matrices. Inserting equation (51) into Theorem 4.4 yields:

$$a \prod_{i=1}^{n_C} \lambda_i \leq \prod_{i=1}^{n_C} \exp(\overline{\mathcal{L}}(i,i)\tau) \leq \prod_{i=1}^{n_C} \lambda_i,$$

using $\det(\exp(M)) = \exp(\text{trace}(M))$ for matrices $M \in \mathbb{R}^{n_C \times n_C}$. In this equation, $\exp(\overline{\mathcal{L}}(i,i)\tau) \in [0,1]$ is the *holding probability* $h_i(\tau)$ of conformation i corresponding to time span τ:

Theorem 4.6 *(see also Theorem 4.6 in [71]) Let $\overline{\mathcal{L}}$ be a generator of a Markov process with $\overline{\mathcal{L}}(i,i) < 0$ for every $i \in \{1, \ldots, n_C\}$. If $\bar{x}(t) \in \{1, \ldots, n_C\}$ is a realization of the Markov process (50), then*

$$h_i(\tau) := \mathbf{P}(\underbrace{\inf\{t \geq 0 : \bar{x}(t) \neq i\}}_{\text{exit time}} > \tau \mid \bar{x}(0) = i) = \exp(\overline{\mathcal{L}}(i,i)\tau)$$

is the holding probability of this process according to state i. The transition probability after leaving state i is given by

$$p_{ij} := \mathbf{P}(\bar{x}(\inf\{t \geq 0 : \bar{x}(t) \neq i\}) = j \mid \bar{x}(0) = i) = -\frac{\overline{\mathcal{L}}(i,j)}{\overline{\mathcal{L}}(i,i)}.$$

By Theorem 4.6, the product of holding probabilities of the conformations is bounded by the product of the Perron cluster eigenvalues. If $h_i \approx 1$, then the conformation χ_i is metastable. In terms of $L^1(\pi)$- and $L^2(\pi)$-norms of χ_i the result is the following

Corollary 4.7 *With the above notations and in the situation of Theorem 4.4 the following estimation holds:*

$$a \prod_{i=1}^{n_C} \lambda_i \leq \prod_{i=1}^{n_C} h_i \leq \prod_{i=1}^{n_C} \lambda_i,$$

where the crispness factor *is*

$$a = (\min_{i=1,\ldots,n_C} \{2\frac{\|\chi_i\|_{L^2(\pi)}}{\|\chi_i\|_{L^1(\pi)}} - 1\})^{n_C}.$$

Holding probability and metastability. For a realization $\bar{x}(t)$ of a continuous-time Markov jump process the holding time $h_i(\tau)$ is defined as the probability that a process starting in $\bar{x}(0) = i$ stays in i during the whole simulation length τ. The metastability $\mathcal{P}(i,i)$ of a conformation χ_i is defined as the probability that a process starting in χ_i is again in conformation χ_i after time τ, which means, the process can also leave χ_i

during simulation and return to it again. Therefore, $h_i \leq \mathcal{P}(i, i)$. But due to metastability, returning to conformation χ_i after entering another conformation is a rare event, i.e. $h_i \approx \mathcal{P}(i, i)$. For further investigations of the correspondence between exit rates and eigenvectors of $\mathcal{P}$ see also Huisinga et al. [61]. For the application of rate matrices out of molecular simulations, see Weber and Kube [132].

4.4   Uncoupling in Function Space

Due to equation (36), for a meshless identification of molecular conformations the numerical approximations of $\overline{S}$ and $\overline{P}$ are necessary. Assume, there is a given set $\mathcal{Q}$ of position states $\mathcal{Q} = \{q_1, \ldots, q_N\} \subset \Omega$ sampled according to the Boltzmann distribution, then the matrices can be approximated via Monte Carlo importance sampling like in equation (4). For $i, k = 1, \ldots, s$:

$$\overline{S}(i, k) = \langle \xi_i, \xi_k \rangle_\pi \quad \approx \quad \frac{1}{N} \sum_{l=1}^{N} \xi_i(q_l)\xi_k(q_l),$$

$$\overline{P}(i, k) = \langle \xi_i, P^\tau \xi_k \rangle_\pi \quad \approx \quad \frac{1}{N} \sum_{l=1}^{N} \xi_i(q_l)\xi_j(\overline{q}_l),$$

where $\overline{q}_l \in \Omega$ is a propagation of q_l w.r.t. the equation of motion, where the start momenta are distributed according to the Maxwell-Boltzmann distribution at the observation temperature. At this stage, the task of conformation dynamics seems to be solved. One can choose any method[14] to generate an adequate importance sampling for the computation of $\overline{S}$ and $\overline{P}$. Robust Perron Cluster Analysis from this point of view is a separated post-processing routine. In practice, however, the generation of a sampling $\mathcal{Q}$ is a severe problem. One possible sampling method is Hybrid Monte Carlo. In Hybrid Monte Carlo sampling the proposal step is based on a short-time dynamics simulation of the molecule. This, however, means that energy barriers are hardly crossed, i.e. the sampling distribution is governed by rare events. This effect is called "critical slowing down". An enormous number of methods has been developed to tackle this problem, for an excellent introduction into molecular simulation see Frenkel and Smit [38] or Allen and Tildesley [2]. Additionally, the numerical error of Monte Carlo approximation mainly depends on the variance of the integrand. One can imagine that there are also many possibilities to accelerate Monte Carlo integration by variance reduction. In the following, a sampling method is presented, which can tackle both, variance reduction and critical slowing down. With this method it is possible to relate the sampling error of $\mathcal{Q}$ to a numerical error, which corresponds to the approximation of almost characteristic functions by a given membership basis, see Section 5.2.3.

---

[14]for example multi-canonical algorithm [54], adaptive temperature HMC [112], simulated tempering [84], or other methods [10]

4.4.1   Partitioning Methods

The well-known grid-based partitioning methods [52, 104, 103] have been developed in the early 1960s. These methods can be used in order to decrease the variance of a standard Monte Carlo estimator of an integral I

$$I = \int_\Omega h(q)\,\pi(q)\,dq$$

which has a probability density function π. In these methods the domain Ω is decomposed into a partition $\mathcal{X}_1, \ldots, \mathcal{X}_{\tilde{s}}$ and the integral I is computed for each region $\mathcal{X}_i$ separately, i.e.

$$I = \sum_{j=1}^{\tilde{s}} w_j\,I_j, \quad I_j = \int_{\mathcal{X}_j} h(q)\,\pi_j(q)\,dq, \tag{52}$$

where the probability density $\pi_j : \Omega \to \mathbb{R}$ is the normalized restriction of π to $\mathcal{X}_j$. w_j is the probability of partition $\mathcal{X}_j$, i.e.

$$w_j = \int_{\mathcal{X}_j} \pi(q)\,dq.$$

w_j is also called *coupling weight* or *statistical weight*. The aim of these methods is to find a partition, such that the variance factors

$$v_i := \int_{\mathcal{X}_i} (h(q) - I_i)^2\,\pi_i(q)\,dq \tag{53}$$

of the Monte Carlo integration estimator is similar for each region $\mathcal{X}_i$.

In our setting, the distribution π_j inside the partition $\mathcal{X}_j$ is unknown a priori. An importance sampling approach via Monte Carlo methods can be used for the evaluation of the integrals $I_j, j = 1, \ldots, \tilde{s}$. For this purpose, Fischer et al. [35, 34, 36] define a modified potential for each partition $\mathcal{X}_j$, which restricts the sampling accordingly, i.e. for $j = 1, \ldots, \tilde{s}$

$$V_j(q) := \begin{cases} V(q), & q \in \mathcal{X}_j, \\ \infty, & q \notin \mathcal{X}_j, \end{cases} \tag{54}$$

and use Hybrid Monte Carlo sampling. The disadvantage of this method is that the coupling weights w_j for equation (52) are unknown. In order to calculate them, Fischer [36] uses a hierarchical uncoupling/coupling method. In this method the partition $\mathcal{X}_1, \ldots, \mathcal{X}_{\tilde{s}}$ is constructed hierarchically using samplings at higher temperature for each stage of the hierarchy in order to overcome energy barriers. Re-weighting to lower temperatures can be done by a so-called bridge sampling strategy. For more details of set-based uncoupling/coupling methods see also Meyer [91] and Schütte et al. [113].

The disadvantage of these partitioning methods is that they need accurate integration for each stage of the hierarchy, i.e. convergence of the hybrid Monte Carlo Markov chain

at different temperatures. We will see that meshless partitioning methods overcome this disadvantage. Although the meshless partition should also be computed hierarchically, convergence of the samplings is only needed for the "finest" discretization at the lowest temperature, see Section 5.3.

In the meshless partitioning approach, the weights w_i are computed from the overlap integrals of the "softly separated regions" $\mathcal{X}_i$.

4.4.2   Meshless Partitioning Methods

In the meshless approach to conformation dynamics, the sets $\mathcal{X}_j$ are replaced by membership functions. An adequate extension of the above uncoupling method can be done via uncoupling functions. At first glance, concept consistency is the reason for this extension, but Theorem 4.11 will show that this special "soft" uncoupling routine provides a method for the computation of the unknowns w_j.

Definition 4.8 *A set of ansatz functions* $\Phi_1, \ldots, \Phi_{\tilde{s}} \in C^1(\Omega)$ *is called an* uncoupling basis *if it is a* partition of unity*, i.e. for all* $q \in \Omega$

$$\sum_{j=1}^{\tilde{s}} \Phi_j(q) = 1$$

holds and Φ_i *is positive*

$$\Phi_i(q) > 0 \tag{55}$$

for all $q \in \Omega$ *and* $i = 1, \ldots, \tilde{s}$.

This definition is quite the same as the Definition 3.1 for the membership basis, only the positivity property and the $C^1(\Omega)$ property are different. With these additional assumptions the uncoupling basis can be used in order to define modified potential energy functions V_j, i.e. a kind of continuation of (54) with $\exp(-\beta V_j) = \Phi_j \exp(-\beta V)$:

$$V_j(q) = V(q) - \frac{1}{\beta} \ln(\Phi_j(q)), \quad j = 1, \ldots, \tilde{s}. \tag{56}$$

These functions V_j are well-defined, because Φ_j meets the positivity condition (55) and $V_j \in C^1(\Omega)$, since $V, \Phi_j \in C^1(\Omega)$.

Importance sampling.   Each modified potential V_j has a corresponding probability density $\pi_j : \Omega \to \mathbb{R}$ given by its Boltzmann expression for the canonical ensemble. In order to calculate observables according to these probability densities one can use importance sampling like in equation (4) and HMC strategies. The only change in the corresponding algorithms is the replacement of the original potential energy function V with its modification V_j given in equation (56).

Ergodicity. From the positivity constraint (55) it directly follows that the molecular system associated with the modified potential V_j in equation (56) shares the same ergodicity properties like the original system according to V. This means that theoretically the importance sampling of the modified potential covers the whole domain Ω. In practice, however, the sampling is restricted to "windows", because the Markov chains do not reach high-energy regions in a low temperature simulation.

Umbrella sampling. Equation (56) is also the link between partitioning methods with the well-known umbrella sampling methods invented in 1977, see Torrie and Valleau [123, 124]. In the original version, umbrella sampling was used in order to re-weight a non-canonical ensemble sampling into the desired equilibrium distribution. Torrie and Valleau suggested performing independent samplings with overlapping distributions like in the above meshless partitioning method. In Theorem 4.11 it will be shown that in the present context the statistical weights $w_1, \ldots, w_{\tilde{s}}$ can easily be calculated due to the partition of unity property, which does not hold for general umbrella algorithms.

Since the improvement of umbrella sampling by Berg and Neuhaus [9] in 1992 called multi-canonical algorithm (MUCA), in most cases the potential energy V has been modified in order to eliminate barriers in the potential energy landscape, i.e. to get a flat histogram ideally. In contrast to that, the meshless partitioning method introduces new barriers into the potential and restricts the sampling to local "windows" of the configuration space Ω.

In the last few years, combinations of the traditionally N-way sampling and the so-called flat histogram method become more and more popular, see e.g. Virnau and Müller [127, 128] using successive umbrella or Schulz et al. [110] using the famous Landau-Wang method [129].

Out-of-cluster rejections. Due to the fact that for the *set-based* partitioning method the potential energy $V_j, j = 1, \ldots, \tilde{s}$, coincides with the original potential V for $q \in \mathcal{X}_j$, see equation (54), the MD steps in the Hybrid Monte Carlo routine are implemented with the original forces $-\nabla V$. After the MD proposal step, however, the acceptance step is not only based upon the total energy of the proposed molecular state $(\widetilde{p}, \widetilde{q})$. Additionally, if its position coordinates $\widetilde{q}$ meet $\widetilde{q} \notin \mathcal{X}_j$, the proposed molecule is rejected (this is called *out-of-cluster rejection*).

The decrease of the acceptance ratio due to this kind of rejection for less metastable subsets $\mathcal{X}_j \subset \Omega$ implies that the set-based uncoupling method needs a decomposition of Ω into almost invariant subsets $\mathcal{X}_j, j = 1, \ldots, \tilde{s}$.

In the meshless approach there is no out-of-cluster rejection. For this reason, also transition regions and saddle-point regions of the potential energy landscape can be focused via a suitable construction of an uncoupling basis. Moreover, the examination of these regions may be important for the transition probabilities of the conformations. On

page 101 the advantage of meshless partitioning methods is exemplified.

Restricted transfer operator.   Like in Section 2.4, the sampling process for the modified potential V_j can be associated with a transfer operator P_j^τ. The following lemma collects some properties of this operator.

Lemma 4.9 *For a given uncoupling basis $\Phi_1, \ldots, \Phi_{\tilde{s}}$ and for the uncoupled operators P_j^τ, for $j = 1, \ldots, \tilde{s}$, the following properties hold:*

i) $P_j^\tau e = e$ *for the constant function $e : \Omega \to \mathbb{R}$, $e(q) \equiv 1$.*

ii) $P_j^\tau : L^2(\pi_j) \to L^2(\pi_j)$ *is self adjoint w.r.t. the probability measure $\mu_j(dq) = \pi_j(q) \, dq$ defined via*

$$\pi_j(q) = \frac{\Phi_j(q) \, \pi(q)}{w_j}, \quad w_j = \int_\Omega \Phi_j(q) \, \pi(q) \, dq. \tag{57}$$

iii) *The probability density π is a convex combination of the partial densities π_j with*

$$\pi = \sum_{j=1}^{\tilde{s}} w_j \pi_j. \tag{58}$$

iv) $P_j^\tau : L^1(\pi_j) \to L^1(\pi_j)$ *is a Markov operator.*

Proof: The properties i),ii) and iv) are a consequence of Theorem 2.2, because P_j^τ is the transfer operator of a Hamiltonian differential equation with a modified potential energy function V_j.

For ii) it is only left to show that π_j meets equation (57). From the Boltzmann distribution we know that the following terms are proportional:

$$
\begin{aligned}
\pi_j(q) \ &\propto \ \exp(-\beta \, V_j(q)) \\[2mm]
&= \ \exp(-\beta \, V(q) + \ln(\Phi_j(q))) \\[2mm]
&= \ \Phi_j(q) \, \exp(-\beta \, V(q)). \\[2mm]
&\propto \ \Phi_j(q) \, \pi(q).
\end{aligned}
$$

Clearly the normalization constant for this relation is w_j defined in (57). The positivity of π_j as a probability density follows from the positivity of Φ_j. From (57) to (58) we use the fact that $\Phi_1, \ldots, \Phi_{\tilde{s}}$ is a partition of unity.

Finally, the linear combination factors w_j are convex combination factors, i.e. they are positive and the sum of them is 1, because π is a normalized probability density:

$$\sum_{j=1}^{\tilde{s}} w_j = \sum_{j=1}^{\tilde{s}} \langle \Phi_j \rangle_\pi = \langle e \rangle_\pi = 1,$$

where $e : \Omega \to \mathbb{R}$ is the constant function $e(q) \equiv 1$.   □

4.5 Coupling in Function Space

Coupling condition.   The calculation of an observable $\langle A \rangle_\pi$ can be done via a coupling of partial densities, like in equation (52),

$$\langle A \rangle_\pi = \sum_{j=1}^{\tilde{s}} w_j \langle A \rangle_{\pi_j}, \tag{59}$$

if the probability densities $\pi_1, \ldots, \pi_{\tilde{s}}$ meet the coupling condition

$$\pi = \sum_{j=1}^{\tilde{s}} w_j \pi_j.$$

Equation (58) in Lemma 4.9 has shown that the coupling condition holds for the partial densities obtained by the uncoupling method via modified potential energies V_j defined in (56).

Importance sampling for S and P.   In order to use importance sampling, one generates a Markov chain $q_1^j, \ldots, q_N^j \in \mathbb{R}^{3n}$ of position coordinates according to the distribution π_j via Hybrid Monte Carlo method (see Section 2.3) with the modified potential V_j. For
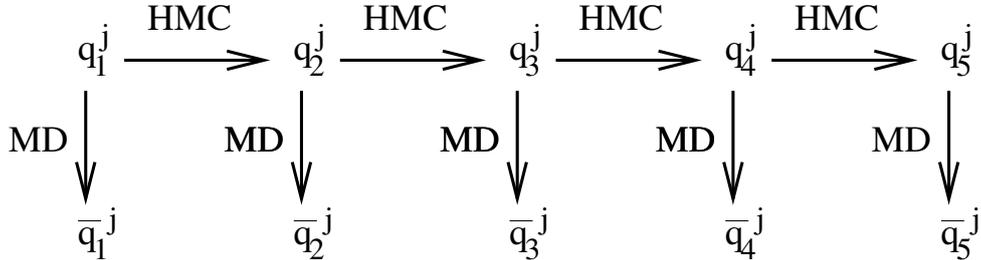
Figure 4: For each modified potential energy V_j a HMC trajectory is computed and for each point q_i of this trajectory a MD-simulation with the original potential energy and the randomly generated momenta is computed.

an approximation of P^τ, however, transitions with the original potential V are necessary. One therefore constructs a sampling in configuration space $\bar{q}_1^j, \dots, \bar{q}_N^j \in R^{3n}$, where the position coordinates $\bar{q}_i^j$ are obtained by a molecular dynamics simulation with the original potential energy V for a time span τ with initial position vector q_i^j and initial momenta coordinates p distributed according to η. In practice, i.e. in the Hybrid Monte Carlo setting, at each step of the Markov chain one randomly chooses a vector p according to η and propagates the vector $q_i^j \in \Omega$ with the original force field (to get $\bar{q}_i^j$) and with the modified force field (to get a proposal $\tilde{q}_i^j$ for the acceptance step) using p as initial momenta. Figure 4 shows the sampling scheme, which has to be applied to each modified potential.

Remark 4.10 *Sampling according to Figure 4 seems to double the numerical effort. In my opinion, there is no computationally efficient way to perform both steps (discretize P^τ and sample the right distribution π) together in only one Markov chain. The reason is that numerically efficient sampling of the Boltzmann distribution needs rapidly mixing Markov chains, whereas the discretization of P^τ needs the opposite – low transition rates between the conformations of the molecule. If one wants to obtain the distribution π and the discretization of P^τ in only one sampling, then either the chain converges very slowly or the transfer operator is not discretized correctly. From this point of view, separating these two tasks is necessary.*

Different integrators. The advantage of the splitting in Figure 4 is that one may choose different integrators for the HMC and the MD part. For the HMC part a reversible and area-preserving integration method can be used in order to satisfy the requirements for the detailed balance condition (6). For the MD part a different integrator can be chosen, e.g. the extrapolation method DIFEX2 [22, 23]. A motivation for a different integrator is that for the calculation of $P^\tau f$ for some function $f \in L^2(\pi)$ the error of the numerical integrator with respect to the position coordinates is important, because f is evaluated at a propagated point in configuration space, $f(\Pi_1 \Phi_h^\tau(q, p))$. One can also replace the MD equation of motion in Fig. 4 with a heat bath dynamics, a mathematical motivation for this change is given in Section 4.3.

$\bar{q}_i^j$ **is a Markov chain.** The computation of $\bar{q}_i^j$ out of q_i^j only depends on the choice of the corresponding momenta $p \propto \eta$. Due to Bayes' rule for the conditional probability $\mathbf{P}(\bar{q}_i^j | q_i^j) = \mathbf{P}(q_i^j \rightarrow \bar{q}_i^j)$ there is a reverse probability $\mathbf{P}(q_i^j | \bar{q}_i^j)$. Taking the route

$$\bar{q}_i^j \rightarrow q_i^j \rightarrow q_{i+1}^j \rightarrow \bar{q}_{i+1}^j$$

in Figure 4 shows that $\bar{q}_1^j, \dots, \bar{q}_N^j$ is also a realization of a Markov chain by the Chapman-Kolmogorov equation:

$$\mathbf{P}(\bar{q}_{i+1}^j | \bar{q}_i^j) = \int_\Omega \int_\Omega \mathbf{P}(q_i^j | \bar{q}_i^j) \, \mathbf{P}(q_{i+1}^j | q_i^j) \, \mathbf{P}(\bar{q}_{i+1}^j | q_{i+1}^j) \, dq_i \, dq_{i+1}.$$

One can therefore apply Markov chain convergence results to the $\overline{q}$-chain, too.

Computation of S and P. After these preparations, we now compute S and P for the membership basis $\xi_1, \ldots, \xi_s$. With equation (59) the matrices $\overline{P}$ and $\overline{S}$ are given as

$$
\begin{aligned}
\overline{P}(i,k) &= \langle \xi_i, P^\tau \xi_k \rangle_\pi = \sum_{j=1}^{\tilde{s}} w_j \langle \xi_i, P^\tau \xi_k \rangle_{\pi_j} = \sum_{j=1}^{\tilde{s}} w_j \overline{P}_j(i,k), \\
\overline{S}(i,k) &= \langle \xi_i, \xi_k \rangle_\pi = \sum_{j=1}^{\tilde{s}} w_j \langle \xi_i, \xi_k \rangle_{\pi_j} = \sum_{j=1}^{\tilde{s}} w_j \overline{S}_j(i,k).
\end{aligned}
\tag{60}
$$

If we normalize the rows of these matrices $\overline{S}$ and $\overline{P}$, we get the desired stochastic matrices S and P. Like in equation (4), the matrices $\overline{S}_j$ and $\overline{P}_j$ can be approximated by the importance sampling, i.e. for $i, k = 1, \ldots, s$ and $j = 1, \ldots, \tilde{s}$

$$
\begin{aligned}
\overline{S}_j(i,k) = \langle \xi_i, \xi_k \rangle_{\pi_j} &\approx \frac{1}{N} \sum_{l=1}^{N} \xi_i(q_l^j) \xi_k(q_l^j), \\
\overline{P}_j(i,k) = \langle \xi_i, P^\tau \xi_k \rangle_{\pi_j} &\approx \frac{1}{N} \sum_{l=1}^{N} \xi_i(q_l^j) \xi_k(\overline{q}_l^j).
\end{aligned}
\tag{61}
$$

Computation of w_j. It is only left to compute w_j for $j = 1, \ldots, \tilde{s}$, if we want to apply equation (60). From equation (57) we know

$$
w_j = \langle \Phi_j, e \rangle_\pi = \langle \Phi_j \rangle_\pi.
\tag{62}
$$

This means, one might also approximate w_j as an observable. But it cannot be computed with the same methods used for S and P, because this would already require the knowledge of w_j. Here is a trick, how to compute w_j out of the data collected for S and P: If the uncoupling basis and the membership basis are identical, i.e. $\tilde{s} = s$ and

$$
\xi_j = \Phi_j, \quad j = 1, \ldots, s,
\tag{63}
$$

then the vector $w = (w_1, \ldots, w_s)^\top \in \mathbb{R}^s$ is the unique normalized positive left eigenvector of S and of P for the Perron eigenvalue $\lambda_1 = 1$, see Lemma 4.1. For both cases, either $M = S$ and $\overline{M} = \overline{S}$ or $M = P$ and $\overline{M} = \overline{P}$, the following equation holds

$$
w^\top = w^\top M = w^\top \operatorname{diag}(w_1, \ldots, w_s)^{-1} \overline{M} = \sum_{i=1}^{s} \overline{M}(i, :).
$$

Furthermore, we get from equation (60)

$$
w^\top = \sum_{i=1}^{s} \sum_{j=1}^{s} w_j \overline{M}_j(i, :) = \sum_{j=1}^{s} w_j \left(\sum_{i=1}^{s} \overline{M}_j(i, :) \right) = w^\top \mathcal{M},
\tag{64}
$$

where $\mathcal{M}$ is a positive matrix with

$$\mathcal{M}(j,k) := \sum_{i=1}^{s} \overline{S}_j(i,k) \quad \text{or} \quad \mathcal{M}(j,k) := \sum_{i=1}^{s} \overline{P}_j(i,k), \qquad (65)$$

which can be computed approximately by equation (61). Because of Perron's Theorem $\mathcal{M}$ has a unique positive left eigenvector. This eigenvector is obviously w and the eigenvalue is 1, see (64). We now collect the main propositions about the relationship between π and the partial densities π_j made so far in the following

Theorem 4.11 *The probability density function π is a convex combination of the partial probability density functions π_j, $j = 1, \ldots, s$, with the convex combination factors $w = (w_1, \ldots, w_s)^\top$. Herein, w is the unique positive normalized left eigenvector of the stochastic matrix $\mathcal{M} \in \mathbb{R}^{(s,s)}$ w.r.t. its Perron eigenvalue $\lambda_1(\mathcal{M}) = 1$, where*

$$\mathcal{M}(j,k) := \langle \Phi_k \rangle_{\pi_j},$$

or respectively

$$\mathcal{M}(j,k) := \langle P^\tau \Phi_k \rangle_{\pi_j}.$$

These matrices[15] can be approximated via equation (61) and (65), if the membership basis ξ and the uncoupling basis Φ are identical.

Proof: A simple calculation shows that $\mathcal{M}$ defined in (65) meets

$$\mathcal{M}(j,k) = \sum_{i=1}^{s} \langle \Phi_i, \Phi_k \rangle_{\pi_j} = \langle \Phi_k \rangle_{\pi_j},$$

or respectively

$$\mathcal{M}(j,k) = \sum_{i=1}^{s} \langle \Phi_i, P^\tau \Phi_k \rangle_{\pi_j} = \langle P^\tau \Phi_k \rangle_{\pi_j},$$

and that $\mathcal{M}$ is stochastic, because π_j is normalized. As $\mathcal{M}$ is positive and stochastic, its Perron eigenvalue is 1. From Perron's Theorem, see Theorem 1.2.2. in [7], we conclude that $\lambda_1 = 1$ is geometrically simple and that any positive left eigenvector is a scalar multiple of w, which completes the proof. $\qquad\square$

Alternative proof. Theorem 4.11 becomes clear due to Lemma 4.1, if we show that $\mathcal{M}$ equals S or P respectively. For the case $\overline{M} = \overline{S}$:

$$\mathcal{M}(j,k) = \langle \Phi_k \rangle_{\pi_j} = \int_\Omega \Phi_k(q) \frac{\Phi_j(q)\,\pi(q)}{w_j}\,dq = \frac{1}{w_j}\langle \Phi_j, \Phi_k \rangle_\pi = S(j,k).$$

The result for $\overline{M} = \overline{P}$ follows analogously.

---

[15]In general, $\langle \Phi_k \rangle_{\pi_j} \neq \langle P^\tau \Phi_k \rangle_{\pi_j}$, because P^τ is not self-adjoint w.r.t. the probability measure $\mu_j(dq) = \pi_j(q)\,dq$.

Theorem 4.11 in practice.   Although, $\mathcal{M}$ is an approximation of S or P and can be used for a Perron Cluster Analysis, in practice we use $\mathcal{M}$ only for the computation of w according to the above theorem. Then the calculation of $\overline{S}$ and $\overline{P}$ is done via (60) and (61). The reason for this two step computation is that because of numerical errors $\mathcal{M}$ might have complex eigenvalues, which is inconvenient for the Perron Cluster Analysis. Whereas within the two step calculation, the approximation of $\overline{S}$ in (60) is symmetric by construction. After normalization of the rows, the stochastic counterpart S of this matrix $\overline{S}$ has a real-valued spectrum. On the other hand, the approximation of $\overline{P}$ in (60) should be symmetric by theory. If $\overline{P}$ is not symmetric, one can use this information for an identification of badly resolved transition samplings, see Section 5.3.

Using parallel computers.   Theorem 4.11 provides a method that allows a *parallel* sampling of the modified potentials on different computers and a concluding computation of w_j afterwards. As $\mathcal{M}$ is a stochastic matrix, Markov chain sensitivity analysis can be applied to this kind of weight computation, see Section 5.2.2.

Well-conditioned Robust Perron Cluster Analysis.   On page 60 it has been shown that meshless partitioning cannot have out-of-cluster rejections, which means that the uncoupling basis functions need not be metastable. This still holds, if membership basis and uncoupling basis are identical. For a well-conditioned Robust Perron Cluster Analysis according to a generalized eigenvalue problem, however, condition (38) should be satisfied, which means $S(i,i) > 0.5$ for $i = 1, \ldots, s$. The interpretation of this condition is: The softly restricted sampling should *mainly* take place "inside" the corresponding uncoupling basis function. Or: The potential modification V_i in (56) should "really" restrict the sampling to a region, where Φ_i is approximately 1.

5 Design of Algorithmic Details

5.1 Construction of Shape Functions

5.1.1 Partition of Unity Methods

In order to apply Theorem 4.11 for the computation of the vector $w \in \mathbb{R}^s$ of weights, the uncoupling basis and the membership basis must be identical. The basis functions $\Phi_1, \ldots, \Phi_s : \Omega \to [0, 1]$ are also referred to as *shape functions* in meshless methods. The creation of shape functions is the central issue in meshless methods. Only shape functions, which are created according to the partition of unity (PU) methods, meet the requirements of Definition 4.8. The PU methods are further developments of the partition of unity finite element methods of Babuška and Melenk [4, 5] in 1996 and *hp*-clouds of Duarte and Oden [31] in 1995. The main idea of this approach is the construction of a partition of unity in function space using Shepard's method [117], i.e. for given weight functions $W_j : \Omega \to \mathbb{R}, j = 1, \ldots, s$, define the shape functions $\Phi_j : \Omega \to \mathbb{R}$ via

$$\Phi_j(q) = \frac{W_j(q)}{\sum_{i=1}^{s} W_i(q)}. \tag{66}$$

Obviously the shape functions Φ_j meet the partition of unity property and, if each W_j is a positive and continuously differentiable function, the shape functions are also positive and continuously differentiable and meet Definition 4.8.

Zero-th order of consistency. For the solution of partial differential equations the *order of consistency* of the function basis is important. An order k of consistency denotes the capability of the function basis to reproduce complete polynomials of order k at any point of Ω. Shape functions, which meet the partition of unity property, are therefore zero-th order consistent, see Liu [80]. Theorem 5.10 will show that zero-th order consistency is sufficient for the approximation of membership functions. For a detailed approximation error analysis of meshless interpolants with p-th order of (differential) consistency see Section 4.2 and 4.3 in [79].

Higher order of consistency. Griebel and Schweitzer [45, 115] increase the order of consistency by multiplying the shape functions Φ_j in equation (66) with multi-dimensional polynomials to get a higher order function basis. The curse of dimensionality enters the method, because the number of basis functions for an n-dimensional polynomial representation of order k increases exponentially with $(k + 1)^n$. The idea of higher-order Shepard methods of Fasshauer [32] leads to similar shape functions and also suffers from the curse of dimensionality. For this reason, approximation of membership functions is done with a zero-th order consistency function basis given by equation (66).

Choice of the weight functions. In the algorithmic realization, each W_j is a *radial basis function* (RBF). This kind of shape functions fits into the HMC particle sampling concept. A radial basis function is associated with a particle $q_j \in \Omega$, which is called a *node*. The function value $W_j(q)$ of an RBF depends on the distance $d(q, q_j)$ between its node q_j and the evaluation point $q \in \Omega$.

Definition 5.1 *For a subset $\mathcal{Q}^* = \{q_1, \ldots, q_s\} \subset \Omega$ of* nodes *and a monotone decreasing continuously differentiable function $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $\omega(0) = 1$ define a set of* radial basis functions $W_1, \ldots, W_s : \Omega \to \mathbb{R}_{\geq 0}$ *via*

$$W_j(q) := \omega(d(q, q_j)).$$

Macrostate dissection. Shalloway et al. [99, 116, 15] in 1994 also use this kind of meshless partitioning into *macrostates*, which are defined according to the modified potential in (56). They apply Shepard's method to Gaussian type weight functions. However, the aim of their work is different. Shalloway et al. try to fit the Boltzmann distribution (i.e. the canonical ensemble) with a set of Gaussians in order to calculate thermodynamical properties analytically. For the identification of the corresponding Gaussian parameters a self-consistent iteration method is provided in [99]. This requires some knowledge about the number of macrostates and the Boltzmann distribution, which in our case is not given a priori.

5.1.2 Unimodal and Strictly Quasi-Concave Functions

For Definition 5.1 any decreasing C^1-function ω is suitable. There is a detailed overview of radial basis functions used in practice in [80]. Among all possible functions ω, the Gaussian type radial basis functions play the most important role. There is a mathematical reason for this kind of function basis, which will be derived now:

Shepard's method destroys some properties of the weight function, e.g. the local maximizer of Φ_j is in general different from the maximizer of W_j, i.e. different from the node q_j. A priori one cannot even assume that Φ_j has a unique local maximizer. However, this is an important property. A function Φ_j of the uncoupling basis is supposed to stamp out a certain "window" of the configuration space Ω, i.e. it should at least not have different local maximizers. The class of *strictly quasi-concave* functions meet this propery which is called *unimodality*. For an introduction into fundamental results on concave functions and maximization see e.g. Horst et al. [58, 59]. In this section, it will be shown that special shape functions meet unimodality.

Definition 5.2 *A function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega \subset \mathbb{R}^m$ is called* strictly quasi-concave *if it meets*

$$f(\lambda x + (1 - \lambda)y) > \min\{f(x), f(y)\}$$

for all $x, y \in \Omega$ with $x \neq y$ and $0 < \lambda < 1$, see Fig. 5.
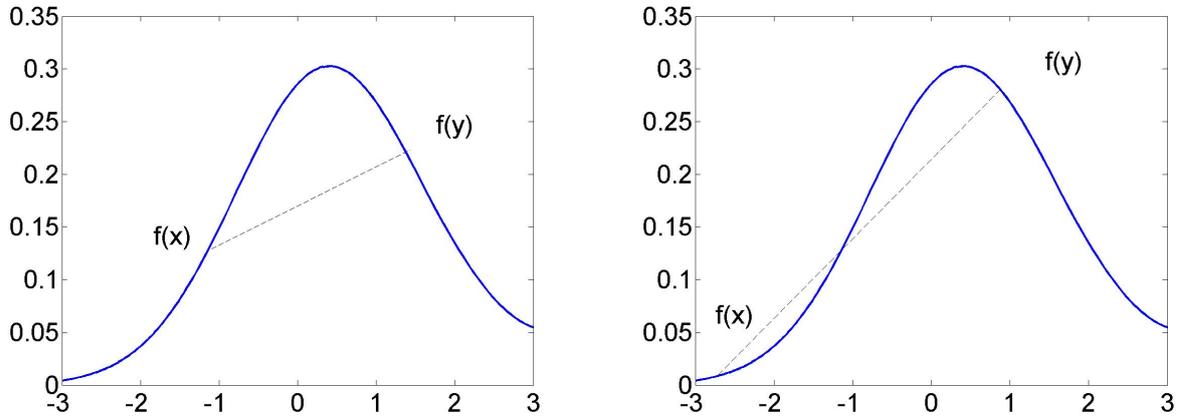
Figure 5: A strictly quasi-concave function f meets $f(\lambda x + (1-\lambda)y) > \min\{f(x), f(y)\}$ for $\lambda \in (0,1)$.

Lemma 5.3 *If $x, y \in \Omega$ are local maximizers of a strictly quasi-concave continuous function $f : \Omega \to \mathbb{R}$, then $x = y$.*

Proof: Obviously, a strictly quasi-concave function f cannot have different local maximizers along *any* arbitrary line segment in Ω, which is the proof for the above Lemma. Moreover, unimodality is not the only consequence of this fact, e.g. f cannot have saddle points or local minima inside Ω either. This is only possible at the boundaries $\partial\Omega$. And even more, the level sets of f are convex in Ω, see Lowen [81]. $\qquad\square$

Eliminating trivial degrees of freedom. For Gaussian type weight functions we need a distance measure between different position states of a molecule. In the molecular dynamics part of the sampling algorithm the position states are described by Cartesian coordinates. For the calculation of the distance $d : \Omega \times \Omega \to \mathbb{R}$ between two position states in Ω, however, the Cartesian coordinates are not appropriate. Rotational and translatory degrees of freedom of the molecule have to be eliminated first, only the geometry of the molecule itself is important. For this purpose, the configuration space of the molecule is projected to a (sub)space $\overline{\Omega}$ via a continuously differentiable projection function $\varphi : \Omega \to \overline{\Omega}$. For a continuously differentiable metric $\bar{d} : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$ in $\overline{\Omega}$ define the distance function d via

$$d(q, \tilde{q}) = \bar{d}(\varphi(q), \varphi(\tilde{q})). \tag{67}$$

This distance function is nonnegative and symmetric, because $\bar{d}$ is a metric. Furthermore, a triangular inequality holds for the distance function d:

Lemma 5.4 *For all $q_0, q_1, \tilde{q} \in \Omega$*

$$d(q_0, q_1) \leq d(q_0, \tilde{q}) + d(q_1, \tilde{q}).$$

Proof:

$$
\begin{aligned}
d(q_0, q_1) &= \bar{d}(\varphi(q_0), \varphi(q_1)) \\[2mm]
&\leq \bar{d}(\varphi(q_0), \varphi(\tilde{q})) + \bar{d}(\varphi(\tilde{q}), \varphi(q_1)) \\[2mm]
&= d(q_0, \tilde{q}) + d(q_1, \tilde{q}).
\end{aligned}
$$

$\square$

With these preparations, we consider Φ_j as a mapping $\Phi_j : \Omega \to \mathbb{R}$ and show in the following theorem that Φ_j is strictly quasi-concave for Gaussian type radial basis functions.

Theorem 5.5 *Let $\overline{\Omega}$ be an m-dimensional vector space and $\bar{d}$ the euclidian metric, i.e. for the components $x(k), y(k)$, for $k = 1, \ldots, m$, of the vectors $x, y \in \overline{\Omega}$*

$$\bar{d}(x, y) = \sqrt{\sum_{k=1}^{m} (x(k) - y(k))^2}.$$

Define ω as a Gaussian type function $\omega(\bar{d}) := \exp(-\alpha\, \bar{d}^2)$ with $\alpha > 0$. Furthermore, for a set of linearly independent nodes $q_1, \ldots, q_s \in \Omega$ and $i = 1, \ldots, s$ ($s > m$) define the projections $y_i := \varphi(q_i) \in \overline{\Omega}$. With the above notations, the basis functions $\Phi_j : \overline{\Omega} \to [0, 1]$, which are given by

$$\Phi_j(x) = \frac{\omega(\bar{d}(x, y_j))}{\displaystyle\sum_{i=1}^{s} \omega(\bar{d}(x, y_i))},$$

are strictly quasi-concave for $j = 1, \ldots, s$.

Proof: Note that the functions $l_{ij}(x) := \bar{d}^2(x, y_i) - \bar{d}^2(x, y_j)$ are linear for all $i, j = 1, \ldots, s$ due to cancellation of the quadratic terms and that ω is strictly convex for $\alpha > 0$. Φ_j is positive and therefore its reciprocal exists. With these preparations it is shown that Φ_j^{-1} is strictly convex. For $x, y \in \overline{\Omega}$ and $0 < \lambda < 1$:

$$\Phi_j^{-1}(\lambda x + (1 - \lambda)y) = \sum_{i=1}^{s} \frac{\exp(-\alpha\, \bar{d}^2(\lambda x + (1 - \lambda)y, y_i))}{\exp(-\alpha\, \bar{d}^2(\lambda x + (1 - \lambda)y, y_j))}$$

$$
\begin{aligned}
&= \sum_{i=1}^{s} \exp(-\alpha\, l_{ij}(\lambda x + (1-\lambda)y)) \\
&= 1 + \sum_{i \neq j} \exp(-\alpha\, [\lambda l_{ij}(x) + (1-\lambda)l_{ij}(y)]) \\
&< 1 + \sum_{i \neq j} \lambda \exp(-\alpha\, l_{ij}(x)) + (1-\lambda)\exp(-\alpha\, l_{ij}(y)) \\
&= \lambda \Phi_j^{-1}(x) + (1-\lambda)\Phi_j^{-1}(y). \tag{68}
\end{aligned}
$$

The above inequality is implied by the fact that $s > m$ and that the nodes are linearly independent, because in this case for every j there exists an index $i \in \{1, \ldots, s\}$ such that for the corresponding linear function $l_{ij}(x) \neq l_{ij}(y)$. In order to show that Φ_j is strictly quasi-concave, we use that Φ_j^{-1} is strictly convex and $f(r) = 1/r$ is a strictly monotone decreasing function for $r > 0$:

$$
\begin{aligned}
\Phi_j(\lambda x + (1-\lambda)y) &= \frac{1}{\Phi_j^{-1}(\lambda x + (1-\lambda)y)} \\
&> \frac{1}{\lambda \Phi_j^{-1}(x) + (1-\lambda)\Phi_j^{-1}(y)} \\
&> \min\left\{ \frac{1}{\Phi_j^{-1}(x)}, \frac{1}{\Phi_j^{-1}(y)} \right\} \\
&= \min\{\Phi_j(x), \Phi_j(y)\}.
\end{aligned}
$$

$\square$

For the proof it is essential that ω is an exponential function and that $\bar{d}$ is the euclidian metric. Therefore Gaussian type functions ω play an important role in the following sections.

Kabsch projection.   A possible projection φ preserving all intramolecular informations is the well-known Kabsch algorithm [66, 67] to align molecules. In this method one fixes a reference molecule first and aligns each molecule to this structure. For the distance between two aligned molecules the euclidian metric can be used to meet the assumptions of Theorem 5.5, i.e. this method yields the unimodality.

Cyclic coordinates.   In practice, however, the Kabsch algorithm is computationally expensive and the gradient of this projection φ can only be computed numerically. Another possibility is a description of the molecule via internal coordinates like bond lengths, angles and dihedrals. In this context it is a frequently-used working hypothesis that the

dihedrals[16] of a molecule comprise all conformational information, e.g. see Cordes et al. [17]. Therefore one can use a distance measure based only on the dihedrals of the molecule.

Assume a molecule described by m dihedrals, then the configuration space Ω of this molecule is mapped into $\Omega^* := (-\pi, \pi]^m$. Its elements are m-tuples $\delta \in \Omega^*$ of the dihedrals $\delta(1), \ldots, \delta(m) \in (-\pi, \pi]$ of the molecule. Dihedrals are cyclic coordinates, which lead to different distance functions in Ω^* and to a different definition of unimodality.
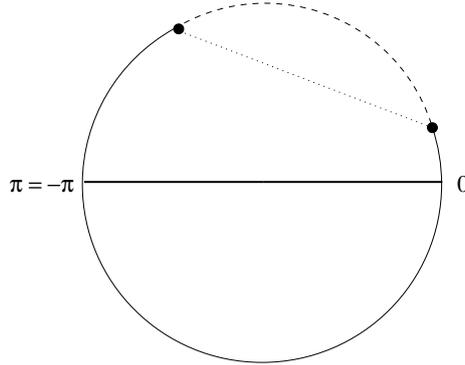
Figure 6: Different distance functions of cyclic coordinates. *Dotted line:* Mapping of a cyclic coordinate to $\mathbb{R}^2$ and using the euclidian metric, see (69). *Dashed line:* Using the shortest arc length measured in radians between the two cyclic data points, see (70).

Distance functions for cyclic coordinates. One possible distance function is based on a proposal of Galliat et al. in [41], see dotted line in Figure 6. For two points $\delta_1, \delta_2 \in \Omega^*$ define $\bar{d} : \Omega^* \times \Omega^* \to \mathbb{R}_{\geq 0}$ as

$$\bar{d}(\delta_1, \delta_2) = \sqrt{\sum_{i=1}^m \left(\sin[\delta_1(i)] - \sin[\delta_2(i)] \right)^2 + \left(\cos[\delta_1(i)] - \cos[\delta_2(i)] \right)^2}. \qquad (69)$$

$\bar{d}$ is the euclidian metric in $\overline{\Omega}$, where $\overline{\Omega} \subset \mathbb{R}^{2m}$ and the projection $\varphi : \Omega^* \to \overline{\Omega}$ is defined via

$$\varphi(\delta) := (\sin[\delta(1)], \cos[\delta(1)], \ldots, \sin[\delta(m)], \cos[\delta(m)])^\top \in \mathbb{R}^{2m}.$$

However, for the Definition 5.2 of strictly quasi-concave functions the domain has to be convex, which is not the case for $\overline{\Omega} = \varphi(\Omega^*)$. This means that the distance function (69) meets the assumptions of Theorem 5.5 and that the corresponding shape functions Φ_i are

---

[16]Four atoms in a molecule which are connected to one another end-to-end form a dihedral angle. The first three atoms span one plane, the last three span another one. The angle between these two planes is defined as dihedral angle.

strictly quasi-concave in $\mathbb{R}^{2m}$, but they might have different local maximizers in a subset $\overline{\Omega} \subset \mathbb{R}^{2m}$. The problem is that the $2m$ "coordinates" are not independent.

For the numerical routines, a more natural distance function for cyclic coordinates is used, see dashed line in Fig. 6: The convex domain in this case is $\Omega^* = (-\pi, \pi]^m$ itself. For cyclic coordinates $z, x \in \Omega^*$ one defines the shortest path euclidian distance

$$\bar{d}(z, x) := \sqrt{\min_{\sigma \in \{-2\pi, 0, +2\pi\}^m} \sum_{l=1}^{m} (z(l) - x(l) + \sigma(l))^2}. \tag{70}$$

Remark 5.6 *The distance function $\bar{d}$ in (70) is independent of the choice of the basis interval, i.e. instead of $(-\pi, \pi]$ one can choose any 2π-interval for representation of the dihedrals. One can also choose different intervals for different dihedrals.*

Unimodality for cyclic coordinates. Unimodality in the presence of periodic coordinates is defined as unimodality per period. For cyclic coordinates, unimodality is satisfied for a function Φ_j if there is a basis interval representation of Ω^* (see Remark 5.6) for which the function is strictly quasi-concave. It will be shown that the representation of Ω^*, for which the node y_j is the central point, meets this property. The convexity assumption in Definition 5.2 is satisfied for Ω^*. For the estimation (68) in the proof of Theorem 5.5 it is sufficient that the functions l_{ij} are concave. For the special choice of Ω^* we get

$$
\begin{aligned}
l_{ij}(z) &= \bar{d}^2(z, y_i) - \bar{d}^2(z, y_j) \\
&= \left(\min_{\sigma \in \{-2\pi, 0, +2\pi\}^m} \sum_{l=1}^{m} (z(l) - y_i(l) + \sigma(l))^2 \right) - \sum_{l=1}^{m} (z(l) - y_j(l))^2 \\
&= \min_{\sigma \in \{-2\pi, 0, +2\pi\}^m} \left(\sum_{l=1}^{m} a(l)\, z(l) + b(l) \right),
\end{aligned}
$$

where

$$a(l) = 2(y_j(l) - y_i(l) + \sigma(l))$$

and

$$b(l) = y_i^2(l) - y_j^2(l) + \sigma^2(l) - 2\sigma(l)\, y_i(l).$$

l_{ij} is concave because it is the minimum of a set of linear (concave) functions in z. Therefore, the distance function $\bar{d}$ defined via (70) is unimodal. This distance function is used for the numerical routines.

Gradient of the modified potential. If we are not as strict as in Theorem 5.5 and take a look at the potential modification connected to the shape function Φ_j from a molecular dynamics point of view, then the following observations can be made. With the special choice of ω the gradient (force) of the modified part of the potential functions

V_j defined in (56) is always "acting", because due to Lemma 5.3 minima and saddle-points of Φ_j only occur at the boundary of the special representation Ω^*. The gradient is (see Appendix A.1):

$$\nabla(-\beta^{-1}\ln(\Phi_j(q))) = \frac{\alpha}{\beta}\left(-\left[\sum_{i=1}^{s}\Phi_i(q)\,\nabla d^2(q, q_i)\right] + \nabla d^2(q, q_j)\right). \tag{71}$$

For *arbitrary* distance functions the *qualitative* behavior of the gradient in equation (71) for $\alpha, \beta > 0$ again goes well with the claimed property to stamp out a certain window of the configuration space: For $i \neq j$ the gradient $\nabla d^2(\cdot, q_i)$ has a contribution to (71) with a negative sign. In the dynamics steps of the HMC this "pushes" the molecule away from the nodes q_i, for $i \neq j$. Whereas $\nabla d^2(\cdot, q_j)$ in (71) has a positive sign tending to minimize the distance $d(q, q_j)$. Note that for the calculation of the gradient it is important that ω is an exponential function, whereas the properties of d are less important.

5.1.3   Sensitivity Analysis of Shape Functions

Unimodality has been a motivation for the use of Gaussian type radial basis functions. In view of clustering and restricted sampling, the interfaces between different membership functions are important. In this section these interfaces are studied. It will be explained why Gaussian type radial basis functions together with Shepard's method can also be seen as a *generalized Voronoi Tessellation* of the configuration space Ω. And it is shown that α is a soft parameter with regard to cluster analysis.

Definition 5.7 *(Voronoi Tessellation, see e.g. Kohonen [73]) For a given distance function $d : \Omega \times \Omega \to \mathbb{R}_{\geq 0}$ and a set of nodes $\mathcal{Q} = \{q_1, \dots, q_s\} \subset \Omega$ define the following characteristic functions $\Phi_i : \Omega \to \{0, 1\}, i = 1, \dots, s$,*

$$\Phi_i(q) := \begin{cases} 1, & \text{if } \forall_{j \neq i}\, d(q, q_i) < d(q, q_j) \\ 0, & \text{otherwise,} \end{cases}$$

The partition $\mathcal{X}_1, \dots, \mathcal{X}_s$ with $\mathcal{X}_i = \Phi_i^{-1}(\{1\})$ is a Voronoi Tessellation of Ω.

Convexity of generalized Voronoi Tessellation.   In a Voronoi Tessellation a point $q \in \Omega$ is assigned to its most proximate node. Interfaces are subsets of hyperplanes having the same distance to different nodes, like perpendicular bisectors in two dimensional linear spaces. Convexity of the partition sets $\mathcal{X}_i$ is not so easy to show in the presence of cyclic coordinates. It will be shown by using the level sets of the shape functions Φ_i defined according to the last section. The α-dependent shape functions $\Phi_j^{[\alpha]} : \Omega \to \mathbb{R}_{\geq 0}, i =$

$1, \ldots, s$, are

$$\Phi_i^{[\alpha]}(q) = \frac{\exp(-\alpha d^2(q, q_i))}{\sum\limits_{j=1}^{s} \exp(-\alpha d^2(q, q_j))}. \tag{72}$$

In the notations of the last section, the *projected level sets* $L(\Phi_i^{[\alpha]}, \gamma) \subset \Omega^*$ of these functions are

$$L(\Phi_i^{[\alpha]}, \gamma) := \{y \in \Omega^*; \ \Phi_i^{[\alpha]}(q) \geq \gamma, q \in \Omega, y = \varphi(q)\}$$

for $\gamma \in [0, 1]$. Equation (72) defines a *generalized Voronoi Tessellation*. This is motivated by the following

Lemma 5.8 *For all $q \in \Omega$ with $d(q, q_i) < d(q, q_j)$, $j \neq i$, the following equation holds:*

$$\lim_{\alpha \to \infty} \Phi_i^{[\alpha]}(q) = 1.$$

The level sets of $\Phi_i^{[\alpha]}$ for the special choice of Ω^* having $y_i = \varphi(q_i)$ as central point (see Remark 5.6) are convex w.r.t. Ω^*. Especially, the convexity of level sets together with the pointwise convergence of $\Phi_i^{[\alpha]}$ implies convexity of the Voronoi Tessellation for a suitable choice of basis intervals Ω^* for each set $\mathcal{X}_i$.

Proof: Let $q \in \Omega$ with $d(q, q_i) < d(q, q_j)$, $j \neq i$. After a suitable permutation of indices one can assume that $d(q, q_l) \leq d(q, q_j)$ for $l \leq j$. Define γ as $\gamma = d(q, q_{i+1})^2 - d(q, q_i)^2$. Then the following inequality holds

$$
\begin{aligned}
1 \ \geq \ & \Phi_i^{[\alpha]}(q) \\
= \ & \frac{\exp(-\alpha d^2(q, q_i))}{\sum\limits_{j=1}^{s} \exp(-\alpha d^2(q, q_j))} \\
\geq \ & \frac{\exp(-\alpha d^2(q, q_i))}{\exp(-\alpha d^2(q, q_i)) + (s-1)\exp(-\alpha d^2(q, q_{i+1}))} \\
= \ & \frac{\exp(-\alpha d^2(q, q_i))}{\exp(-\alpha d^2(q, q_i)) + (s-1)\exp(-\alpha d^2(q, q_i))\exp(-\alpha\gamma)} \\
= \ & \frac{1}{(s-1)\exp(-\alpha\gamma) + 1}.
\end{aligned}
$$

Hence,

$$1 \geq \lim_{\alpha \to \infty} \Phi^{[\alpha]}(q) \geq \lim_{\alpha \to \infty} \frac{1}{(s-1)\exp(-\alpha\gamma) + 1} = 1.$$

For the special choice of Ω^* the shape function $\Phi_i^{[\alpha]}$ is strictly quasi-concave, see Theorem 5.5. This directly implies convexity of its level sets. $\square$

Structure of meshless partitioning.   Lemma 5.8 gives an idea of the structure of the function basis and its interfaces. The shape functions (72) define a generalized Voronoi Tessellation of the configuration space Ω. The corresponding shape functions can be seen as *convex fuzzy sets*, see [81][17].

α-dependence of the shape functions.   The dependence of the basis $\Phi^{[\alpha]}$ on a single parameter α opens access to results of perturbation theory from Section 3.3. For every basis function $\Phi_i^{[\alpha]}$, $i = 1, \ldots, s$ we have (see Appendix A.2)

$$\frac{\partial}{\partial \alpha} \Phi_i^{[\alpha]} \Big|_{\alpha = \alpha^*} = \sum_{j=1, j \neq i}^{s} \left(d^2(q, q_j) - d^2(q, q_i) \right) \Phi_i^{[\alpha^*]}(q) \, \Phi_j^{[\alpha^*]}(q). \tag{73}$$

This equation shows that α-dependence mainly arises at the interfaces between different basis functions, because only in those regions the factor $\Phi_i^{[\alpha^*]}(q) \, \Phi_j^{[\alpha^*]}(q)$ has significant values. α-dependence of $\overline{P}^{[\alpha]}$ and $\overline{S}^{[\alpha]}$ can be estimated by the following inequality, where the self-adjoint operator $M : L^2(\pi) \to L^2(\pi)$ either equals the propagator P^τ or the identity operator; furthermore, define $d_k := d^2(\cdot, q_k)$ and $d_{max} := \max_{i,k} |d_i(q) - d_k(q)|$:

$$
\begin{aligned}
\left| \frac{\partial}{\partial \alpha} \langle \Phi_i^{[\alpha]}, M \Phi_j^{[\alpha]} \rangle_\pi \right| &= \left| \langle \frac{\partial}{\partial \alpha} \Phi_i^{[\alpha]}, M \Phi_j^{[\alpha]} \rangle_\pi + \langle \frac{\partial}{\partial \alpha} \Phi_j^{[\alpha]}, M \Phi_i^{[\alpha]} \rangle_\pi \right| \\
&= \left| \langle \sum_{k=1}^{s} (d_k - d_i) \, \Phi_i^{[\alpha]} \Phi_k^{[\alpha]}, M \Phi_j^{[\alpha]} \rangle_\pi + \right. \\
&\qquad \left. \langle \sum_{k=1}^{s} (d_k - d_j) \, \Phi_j^{[\alpha]} \Phi_k^{[\alpha]}, M \Phi_i^{[\alpha]} \rangle_\pi \right| \\
&\leq d_{max} \big(\langle \Phi_i^{[\alpha]} \sum_{k=1}^{s} \Phi_k^{[\alpha]}, M \Phi_j^{[\alpha]} \rangle_\pi + \\
&\qquad \langle \Phi_j^{[\alpha]} \sum_{k=1}^{s} \Phi_k^{[\alpha]}, M \Phi_i^{[\alpha]} \rangle_\pi \big) \\
&= 2 d_{max} \langle \Phi_i^{[\alpha]}, M \Phi_j^{[\alpha]} \rangle_\pi. \tag{74}
\end{aligned}
$$

Although (74) is only a rough estimation, it shows that a given block-like structure of $\overline{P}^{[\alpha]}$ and $\overline{S}^{[\alpha]}$ is less affected by the α-parameter, because small entries stay small when α changes. Keeping the block structure also preserves the Perron cluster eigenvalues and the space of Perron cluster eigenvectors. In other words: α is a "soft parameter". Changes only take place inside the blocks of $\overline{P}^{[\alpha]}$ and $\overline{S}^{[\alpha]}$.

---

[17]Fuzzy sets $\Phi : \Omega^* \to [0, 1]$ are called *convex* iff Φ is (strictly) quasi-concave.

$$\widetilde{M}^{[\alpha_1]} \xrightarrow{\quad \alpha \quad} \widetilde{M}^{[\alpha_2]}$$
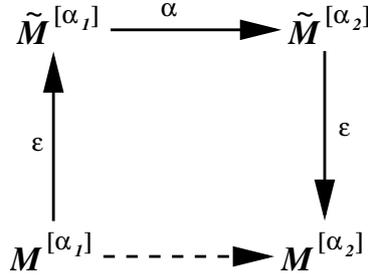
Figure 7: Perturbation scheme for the stochastic matrices S and P. Via ϵ-perturbation the matrices are transformed into uncoupled, block structured matrices $\widetilde{S}$ and $\widetilde{P}$. In this case, α-perturbation has no effect on the first n_C eigenvectors and eigenvalues, because zero-entries are kept.

α-dependence of spectra of S and P.   Figure 7 shows the associated perturbation scheme. The block structures of the unperturbed matrices $\widetilde{S}^{[\alpha]}$ and $\widetilde{P}^{[\alpha]}$ are not affected by α and therefore these matrices have the n_C-fold Perron eigenvalue $\lambda = 1$. In contrast to that, further eigenvalues and eigenvectors are affected, because α changes the entries of $\widetilde{S}^{[\alpha]}$ and $\widetilde{P}^{[\alpha]}$ inside the blocks. For an example of the α-dependence of shape functions also see Section 6 on page 96.

5.2   Error Analysis

Via Definition 5.1 the shape functions $\Phi_1, \ldots, \Phi_s : \Omega \to \mathbb{R}$ are determined by the choice of the nodes $\mathcal{Q}^* \subset \Omega$, the distance function, and the decreasing function $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. In order to construct shape functions explicitly, they have to meet some conditions:

- $\Phi_1, \ldots, \Phi_s$ should have good approximation properties w.r.t. the desired membership functions $\chi_1, \ldots, \chi_{n_C}$, see Section 5.2.1.

- The calculation of the weights w_j should be well-conditioned, see Section 5.2.2.

- The restricted samplings with the modified potentials V_j should converge in reasonable time, see Section 5.2.3.

5.2.1   Approximation Error of Membership Functions

Using Shepard's method (66), the quasi-interpolant

$$\chi(q) \approx \sum_{i=1}^{N} \chi(q_i) \frac{W_i(q)}{\sum_{j=1}^{s} W_j(q)}$$

of a function $\chi : \Omega \to \mathbb{R}$ converges to χ linearly according to a decreasing maximal "diameter" θ of the compactly supported weight functions W_i [117]. However, the compactness

argument becomes invalid in the presence of the positivity constraint (55). Therefore, some changes are necessary for the approximation theory of Shepard's method in this situation. First, a kind of Lipschitz continuity is defined for the approximated membership functions with respect to the new distance function d:

Definition 5.9 *A function $\chi : \Omega \to \mathbb{R}$ is said to be d-Lipschitz continuous if there is an $L \in \mathbb{R}_{>0}$ such that*

$$|\chi(q_1) - \chi(q_2)| \leq L\,d(q_1, q_2)$$

for all $q_1, q_2 \in \Omega$, where d is a distance function defined above.
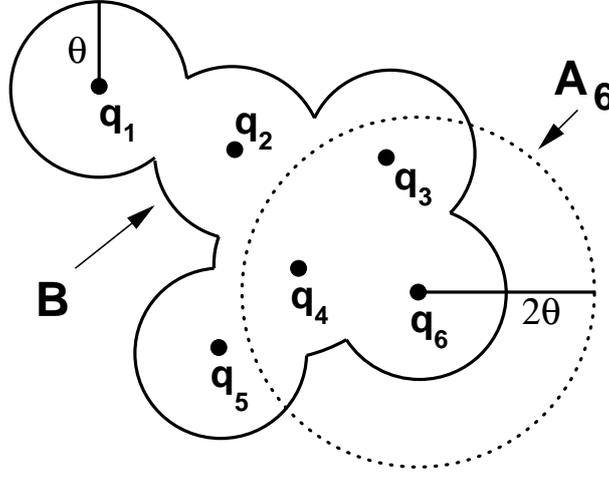
Figure 8: Sketch of the sets B and A_i defined in Theorem 5.10 for $s = 6$.

With these preparations the following theorem can be shown, which provides a similar approximation order $\mathcal{O}(\theta)$ for a kind of discretization constant θ as in the original method of Shepard.

Theorem 5.10 *Let $\chi : \Omega \to [0,1]$ be a d-Lipschitz continuous membership function with Lipschitz constant $L > 0$. Further, let $\{q_1, \ldots, q_s\} = \mathcal{Q}^* \subset \Omega$ be the set of nodes, where*

$$B := \{q \in \Omega;\, \exists_{\bar{q} \in \mathcal{Q}^*}\, d(q, \bar{q}) \leq \theta\} \tag{75}$$

and for $i = 1, \ldots, s$,

$$A_i := \{q \in \Omega;\, d(q, q_i) \leq 2\theta\}$$

with a discretization constant $\theta > 0$. If

$$\mu(B^c) = 1 - \int_B \pi(q)\,dq = \epsilon_1 \tag{76}$$

and

$$\Phi_i(q) \le s^{-1}\epsilon_2, \quad \text{for} \quad q \in B, q \notin A_i, \tag{77}$$

for some $\epsilon_1, \epsilon_2 > 0$ and Φ_i defined via (66), see Fig. 8. Then the following error estimation holds for the best approximating membership function $\widetilde{\chi} \in L^1(\pi)$ with $\widetilde{\chi} \in$ span$\{\Phi_1, \dots, \Phi_s\}$:

$$\|\chi - \widetilde{\chi}\|_{L^1(\pi)} \le \epsilon_1 + \epsilon_2 + 4\theta L. \tag{78}$$

In contrast to Shepard's method, in Theorem 5.10 the compactness of the support of the shape functions is replaced by a condition forcing the shape functions to decrease rapidly, see (77). The fact that it is numerically impossible to "cover" the whole domain Ω with nodes is expressed by a condition that measures the uncovered part with a parameter ϵ_1, see (76).

Proof: Define $C_i := A_i \cap B$ and $\langle f \rangle_{C_i} := \int_{C_i} f(q)\pi(q)\,dq$. With these preparations the following estimation holds

$$
\begin{aligned}
\|\chi - \widetilde{\chi}\|_{L^1(\pi)} \quad &= \quad \int_\Omega |\chi(q) - \widetilde{\chi}(q)|\,\pi(q)\,dq \\[2mm]
(1^*) \quad &\le \quad \epsilon_1 + \int_B |\chi(q) - \widetilde{\chi}(q)|\,\pi(q)\,dq \\[2mm]
&\le \quad \epsilon_1 + \int_B |\chi(q) - \sum_{i=1}^s \frac{\langle \chi, \Phi_i \rangle_{C_i}}{\langle \Phi_i \rangle_{C_i}} \Phi_i(q)|\,\pi(q)\,dq \\[2mm]
&= \quad \epsilon_1 + \int_B |\sum_{i=1}^s (\chi(q) - \frac{\langle \chi, \Phi_i \rangle_{C_i}}{\langle \Phi_i \rangle_{C_i}})\,\Phi_i(q)|\,\pi(q)\,dq \\[2mm]
&\le \quad \epsilon_1 + \int_B \sum_{i=1}^s |\chi(q) - \frac{\langle \chi, \Phi_i \rangle_{C_i}}{\langle \Phi_i \rangle_{C_i}}|\,\Phi_i(q)\,\pi(q)\,dq \\[2mm]
&= \quad \epsilon_1 + \sum_{i=1}^s \int_B |\chi(q) - \frac{\langle \chi, \Phi_i \rangle_{C_i}}{\langle \Phi_i \rangle_{C_i}}|\,\Phi_i(q)\,\pi(q)\,dq \\[2mm]
(2^*) \quad &\le \quad \epsilon_1 + \epsilon_2 + \sum_{i=1}^s \int_{C_i} |\chi(q) - \frac{\langle \chi, \Phi_i \rangle_{C_i}}{\langle \Phi_i \rangle_{C_i}}|\,\Phi_i(q)\,\pi(q)\,dq \\[2mm]
&= \quad \epsilon_1 + \epsilon_2 + \sum_{i=1}^s \langle \Phi_i \rangle_{C_i}^{-1} \int_{C_i} |\chi(q)\langle \Phi_i \rangle_{C_i} - \langle \chi, \Phi_i \rangle_{C_i}|\,\Phi_i(q)\,\pi(q)\,dq \\[2mm]
(3^*) \quad &\le \quad \epsilon_1 + \epsilon_2 + \sum_{i=1}^s \max_{q \in C_i} |\chi(q)\langle \Phi_i \rangle_{C_i} - \langle \chi, \Phi_i \rangle_{C_i}| \\[2mm]
(4^*) \quad &\le \quad \epsilon_1 + \epsilon_2 + 4\theta L. \tag{79}
\end{aligned}
$$

More detailed:

(1^*): The inequality holds, because $\max_{q \in \Omega} |\chi - \tilde{\chi}| \leq 1$. And for the complement B^c of B: $\mu(B^c) = \epsilon_1$.

(2^*): The value of the nominator $\langle \chi, \Phi_i \rangle_{C_i}$ is always less than $\langle \Phi_i \rangle_{C_i}$. Therefore, the whole integrand can be estimated by Φ_i. For $q \notin A_i$ the condition $\Phi_i \leq s^{-1}\epsilon_2$ holds.

(3^*): This inequality holds due to the definition

$$\langle \Phi_i \rangle_{C_i} = \int_{C_i} \Phi_i(q)\, \pi(q)\, dq.$$

(4^*): Let $q^a \in C_i$ with $\chi(q^a) = \max_{q \in C_i} \chi(q)$ and $q^b \in C_i$ with $\chi(q^b) = \min_{q \in C_i} \chi(q)$, then

$$\max_{q \in C_i} |\chi(q)\langle \Phi_i \rangle_{C_i} - \langle \chi, \Phi_i \rangle_{C_i}| \leq \langle \Phi_i \rangle_{C_i} |\chi(q^a) - \chi(q^b)|.$$

Furthermore, $\langle \Phi_i \rangle_{C_i} \leq \langle \Phi_i \rangle_{\Omega}$ and the partition of unity property holds. Because of Lemma 5.4 the maximal distance of two points in A_i is 4θ.

$\square$

Theorem 5.10 directly implies some requirements w.r.t. the choice of the shape functions. These requirements are described in the following.

Restriction to $\mu(B^c) = \epsilon_1$.   For the selection of nodes $\mathcal{Q}^* = \{q_1, \ldots, q_s\} \subset \Omega$ for the radial basis functions $W_1, \ldots, W_s$ it is important to have some insights into the structure of the potential energy landscape $V(q)$ of the molecule. For this purpose, an HMC trajectory $\mathcal{Q} = \{Q_1, \ldots, Q_N\} \subset \Omega, N \gg s$, of position states is generated at a high temperature to overcome energy barriers. After this *pre-sampling* one examines the set $\mathcal{Q}$. Ergodicity of the pre-sampling implies that one can select nodes as a subset $\mathcal{Q}^* \subset \mathcal{Q}$, because for $N \to \infty$ the pre-sampling comprises all position states of the molecule. The aim, however, is not to "cover" the whole domain Ω, but to construct nodes $\mathcal{Q}^*$, such that $\mu(B) = \epsilon_1$ is small with B defined in equation (75). The advantage of a pre-sampling is that within this procedure the points $Q_i, i = 1, \ldots, N$, are collected with a low probability for physically nonsensical position states. One can therefore assume that the pre-sampling covers the *relevant* part of Ω.

Node selection using θ.   If we assume that $\mathcal{Q}$ covers the relevant part of Ω sufficiently, then the selection of $\mathcal{Q}^* \subset \mathcal{Q}$ has to ensure that $\mathcal{Q} \subset B$ using the construction rule (75) for B, i.e. for every $Q \in \mathcal{Q}$ there is a $q \in \mathcal{Q}^*$ such that $d(Q, q) \leq \theta$. This directly implies a possible selection method:

Algorithm 5.11 Node selection - initial discretization

0. Select an arbitrary node $q \in \mathcal{Q}$. Add q to $\mathcal{Q}^*$.

1. Out of the set

$$\overline{B} = \{Q \in \mathcal{Q}; \ \forall_{q \in \mathcal{Q}^*} \ d(Q, q) > \theta\}$$

select $q \in \overline{B}$ having a minimum distance to $\mathcal{Q}^*$. If $\overline{B}$ is empty STOP, else

2. Add q to $\mathcal{Q}^*$. Go on with 1).

Determination of θ.   For a good approximation result due to the term $4\theta L$ in the error estimation (78), θ has to be as small as possible. But for decreasing θ the value $\mu(B^c) = \epsilon_1$ in (78) increases via (75).

For an estimation of θ, note that the dynamics part of the pre-sampling is a continuous process, i.e. $\mathcal{Q}$ is a set of "snapshots" of some dynamically coherent superset. A reasonable choice for θ is therefore the smallest θ, which assures the coherence of $B \supset \mathcal{Q}$.

Rapidly decreasing ω-function.   Good approximation properties of the shape functions correspond to a rapidly decreasing function ω in Theorem 5.10. This can be shown as follows.

Lemma 5.12 *In the situation of Theorem 5.10 the inequality*

$$\Phi_i(q) \leq \frac{\omega(2\theta)}{\omega(\theta)}$$

holds for $q \in B \cap A_i^c$.

Proof: Let $q \in B \cap A_i^c$, then

$$\Phi_i(q) \ = \ \frac{\omega(d(q, q_i))}{\sum_{j=1}^{s} \omega(d(q, q_j))}$$

$$\leq \ \frac{\omega(2\theta)}{\sum_{j=1}^{s} \omega(d(q, q_j))}$$

$$\leq \ \frac{\omega(2\theta)}{\omega(\theta)}.$$

The first inequality holds, because $d(q, q_i) > 2\theta$ and ω is a decreasing function. The second inequality holds, because there is a $q_j \in \mathcal{Q}^*$ with $d(q, q_j) \leq \theta$ and ω is nonnegative. $\square$

The condition $\Phi_i(q) \leq s^{-1}\epsilon_2$ for $q \in B \cap A_i^c$ in Theorem 5.10 follows from Lemma 5.12, if ω meets the rough sufficient condition

$$\frac{\omega(2\theta)}{\omega(\theta)} = s^{-1}\epsilon_2.$$

A possible choice for ω is an exponential function $\omega(d) = \exp(-\alpha\, d^2)$ with

$$\alpha = -\frac{1}{3\theta^2} \ln(s^{-1}\epsilon_2). \tag{80}$$

This special choice of a rapidly decreasing ω is also reasonable from the viewpoint of Theorem 5.5. Note that Lemma 5.12 is only a rough estimation. In practice, a much smaller α is possible, because in Lemma 5.12 it is assumed that each $q \in B$ has at least one neighbor in $\mathcal{Q}^*$ with distance θ. Whereas in high dimensional spaces the number of neighbors is much more than one.

Summarizing the algorithmic issues of Theorem 5.10. Several algorithmic issues are implied by Theorem 5.10. The algorithm used for computer experiments in Section 6 is based on a distance function d using the dihedrals of the molecule, see (70). The radial basis functions are rapidly decreasing exponential functions $\omega(d) = \exp(-\alpha\, d^2)$ with an input parameter $\epsilon_2 > 0$ and α according to (80). The node selection method uses Algorithm 5.11, where θ is an input parameter.

Computation of the discretization error. For the computation of the discretization error $4\theta L + \epsilon_1 + \epsilon_2$ some problems remain.

- The Lipschitz constant $L > 0$ is unknown.

- The computation of $\epsilon_1 = \mu(B^C)$ is not possible or at least as difficult as the computation of the partition function Z_q in equation (2).

For these reasons, Theorem 5.10 is only qualitative and not quantitative. It only shows that it is possible to approximate arbitrary membership functions with the given set of basis functions. In conformation dynamics a defining equation for χ does not exist. Thus, not only approximation properties are important which will be shown in the next section.

5.2.2 Truncation Analysis of Statistical Weight Computation

The approximation of the almost characteristic functions is only a subproblem of conformation analysis. The problem of a correct calculation of the statistical weights $w_1, \ldots, w_s$ is harder and eventually needs more hierarchical refinements of the uncoupling basis. In this section a quantitative error estimation for the statistical weights is presented.

Stationary distribution of the matrix $\mathcal{M}$.   In Theorem 4.11 the calculation of the weights $w_j, j = 1, \ldots, s$, can be seen as a calculation of the stationary distribution of a stochastic, positive matrix $\mathcal{M}$. The quality of this solution is determined by some condition number $\kappa \in \mathbb{R}$ and the estimated error of the Monte Carlo integration for the elements of $\mathcal{M}$.

Condition number of the stationary distribution.   We now derive an error estimation for the weights w_j from the condition number of the stationary distribution of $\mathcal{M}$. This error estimation is similar to the sensitivity analysis for finite Markov chains. The stationary distribution of $\mathcal{M} \in \mathbb{R}^{s \times s}$ is the unique positive vector $w \in \mathbb{R}^s$ with

$$w^\top \mathcal{M} = w^\top, \quad \sum_{i=1}^{s} w_i = 1. \tag{81}$$

Numerically, we can only compute an approximation $\mathcal{M}_{app.}$ of $\mathcal{M}$, which defines an error matrix $E = \mathcal{M}_{app.} - \mathcal{M}$ and a stationary distribution $w_{app.}$ via

$$w_{app.}^\top \mathcal{M}_{app.} = w_{app.}^\top, \quad \sum_{i=1}^{s} w_{app.,i} = 1. \tag{82}$$

The corresponding condition number $\kappa \in \mathbb{R}$ depending on $\mathcal{M}_{app.}$ is the smallest number for which

$$\|w - w_{app.}\|_\infty \leq \kappa \, \|E\|_\infty \tag{83}$$

holds for all $\mathcal{M} \in \mathbb{R}^{s \times s}$ and $w, w_{app.} \in \mathbb{R}^s$ defined via (81) and (82) respectively.

Estimation of κ.   In 2001 Cho and Meyer [13] give an overview of eight different possibilities for an upper bound of κ for various norms in (83). It turns out, that two of these possibilities, the condition number κ_{IM} of Ipsen and Meyer [63, 72] and the condition number κ_{HH} of Haviv and van der Heyden [56, 72], provide sharper bounds than the others. Note that in the present context the roles of $\mathcal{M}$ and $\mathcal{M}_{app.}$ are exchanged with regard to [13].

The condition number κ_{IM}. For simplicity we assume that the Perron eigenvalue $\lambda_1 = 1$ of $\mathcal{M}_{app.}$ is simple and therefore $A := I - \mathcal{M}_{app.}$ for the unit matrix $I \in \mathbb{R}^{s \times s}$ has rank $s - 1$. Further, let $A_{(j)} \in \mathbb{R}^{(s-1) \times (s-1)}$ be the principal submatrix of A obtained by deleting the j-th row and column from A, then the condition number κ_{IM} is defined as

$$\kappa_{IM} = \frac{\min\limits_{j=1,\ldots,s} \|A_{(j)}^{-1}\|_\infty}{2}. \tag{84}$$

The best error bounds in (83) are often given by this condition number in the numerical experiments of Section 6.

The condition number κ_{HH}. The second condition number κ_{HH} in its original definition is obtained from the group inverse of A. But in [13] it has been shown that it can also be computed in terms of the principal submatrices $A_{(j)}$. In this context κ_{HH} is defined as

$$\kappa_{HH} = \frac{\max\limits_{j=1,\ldots,s} w_{app.,j} \|A_{(j)}^{-1}\|_\infty}{2}. \tag{85}$$

Sometimes κ_{HH} provides better error bounds than κ_{IM}.

Best synthesis of condition numbers. The condition number κ_{HH} in (85) for an error estimation of the statistical weights $w_i, i = 1, \ldots, s$, can be written as a component-wise error bound given by Ipsen and Meyer [63] and Cho and Meyer [13]:

$$|w_i - w_{app.,i}| \le w_{app.,i} \|A_{(i)}^{-1}\|_\infty \|E\|_\infty. \tag{86}$$

Summarizing (84),(85), and (86), we use the following synthesis of these condition numbers for the numerical routines, $i = 1, \ldots, s$:

$$|w_i - w_{app.,i}| \le \min\{\kappa_{IM}, \ \kappa_{HH}, \ w_{app.,i} \|A_{(i)}^{-1}\|_\infty\} \|E\|_\infty.$$

Error of the partition functions. More interesting than the detailed statistical weights are the partition functions defined in (17) as a result of the conformation analysis. The partition functions $\widetilde{w}_1, \ldots, \widetilde{w}_{n_C}$ of the conformations $\chi_1, \ldots, \chi_{n_C} : \Omega \to \mathbb{R}$ can be transformed via (63) and (62) into

$$\begin{aligned}
\widetilde{w}_i &= \langle \chi_i \rangle_\pi = \sum_{j=1}^s \langle \chi_{discr.}(j,i) \, \xi_j \rangle_\pi \\
&= \sum_{j=1}^s \chi_{discr.}(j,i) \, \langle \xi_j \rangle_\pi = \sum_{j=1}^s \chi_{discr.}(j,i) \, \langle \Phi_j \rangle_\pi
\end{aligned}$$

$$= \sum_{j=1}^{s} \chi_{discr.}(j,i)\, w_j. \tag{87}$$

By equation (87) the error estimation can be extended to the sensitivity analysis of $\widetilde{w}_i$. Let $\widetilde{w}_{app.,i}$ denote the numerical approximation of $\widetilde{w}_i$ and assume $\chi_{discr.}$ to be calculated without numerical error, then

$$
\begin{aligned}
|\widetilde{w}_i - \widetilde{w}_{app.,i}| &= \left| \sum_{j=1}^{s} \chi_{discr.}(j,i)\,(w_j - w_{app.,j}) \right| \\
&\leq \sum_{j=1}^{s} \chi_{discr.}(j,i)\, |w_j - w_{app.,j}| \\
&\leq \sum_{j=1}^{s} \chi_{discr.}(j,i)\, \min\{\kappa_{IM},\ \kappa_{HH},\ w_{app.,j}\, \|A_{(j)}^{-1}\|_\infty\}\, \|E\|_\infty.
\end{aligned}
$$

Estimation of $\|E\|_\infty$.   For an error analysis via (83) an estimation of $\|E\|_\infty$ is left to compute. From equation (11) in Theorem 2.1 we can conclude with a probability of 84% that for a finite HMC simulation of "large" length N_i and variance $\sigma^2(i,j)$ of the element (i,j) of the matrix $\mathcal{M}$

$$\left| \mathcal{M}_{app.}(i,j) - \mathcal{M}(i,j) \right| \leq \frac{\sigma(i,j)}{\sqrt{N_i}}.$$

For the norm $\|E\|_\infty$ this means

$$\|E\|_\infty \leq \max_{i=1,\ldots,s} \frac{\sum_{j=1}^{s} \sigma(i,j)}{\sqrt{N_i}}. \tag{88}$$

The value of $\sigma(i,j)$ can be estimated by the *sample standard deviation*, i.e.

$$\sigma(i,j) \approx \sqrt{ \frac{1}{N_i - 1} \sum_{k=1}^{N_i} (\Phi_j(q_k^i) - \mathcal{M}_{app.}(i,j))^2 }, \tag{89}$$

for $\mathcal{M} = S$, see Section 4.5, or

$$\sigma(i,j) \approx \sqrt{ \frac{1}{N_i - 1} \sum_{k=1}^{N_i} (\Phi_j(\overline{q}_k^i) - \mathcal{M}_{app.}(i,j))^2 }, \tag{90}$$

for $\mathcal{M} = P$.

Equation (89) can be simplified. Because N_i is large, one can replace $(N_i - 1)$ by N_i. A further simplification can be done, if we recall equation (61) and use

$$\overline{S}_i(j, j) \approx \frac{1}{N_i} \sum_{k=1}^{N_i} \Phi_j(q_k^i)^2.$$

Summarizing, for $\mathcal{M} = S$ we can estimate the error $\|E\|_\infty$ directly from the sampling data:

$$\|E\|_\infty \overset{0.84}{=} \max_{i=1,\dots,s} \frac{\sum\limits_{j=1}^{s} \sqrt{\overline{S}_i(j, j) - \mathcal{M}_{app}^2(i, j)}}{\sqrt{N_i}}, \tag{91}$$

where " $\overset{0.84}{=}$ " means that this estimation can only be understood from a probabilistic point of view. For a convergence check during the sampling algorithm, $\|E\|_\infty$ can better be approximated by using different random numbers and starting points and by comparing these results, see Section 5.2.3.

Better error bounds for S or P? The condition number κ in (83) for $\mathcal{M} = P$ might be better than for $\mathcal{M} = S$, because the block structure of $\mathcal{M} = P$ is more perturbed. But whereas the sampling points q_j^i are restricted to "windows" by the modification of the potential, the sampling points $\bar{q}_j^i$ of the pure MD part are not restricted. That means that the error estimation $\|E\|_\infty$ is much worse for $\mathcal{M} = P$ than for $\mathcal{M} = S$. Therefore the statistical weights are calculated from the approximation of $\mathcal{M} = S$ in the numerical examples. If one uses S for a geometrical clustering $\chi_{discr.}$, one can determine the statistical weights and the partition functions $\widetilde{w}_j$ in equation (17) without any pure MD simulation, see Fig. 4 on page 62.

The importance of correct sub-samplings. The condition numbers κ_{IM} and κ_{HH} are optimal in the sense that they are always smaller than a set of condition numbers examined in [13]. For one of the sub-optimal condition numbers κ_{ME}, Meyer [92] provided the following bounds

$$\frac{1}{s\,|1 - \lambda_2|} \le \kappa_{ME} \le \frac{2(s-1)}{\Pi_{i=1}^s(1 - \lambda_i)}, \tag{92}$$

see also (3.3) in [13]. Here is a consequence of this estimation:

1. $\lambda_2 \approx 1$ is a necessary condition for high metastability inside a molecular system, see upper bounds for $\det(\mathcal{P})$ in Theorem 4.4 and for $\operatorname{trace}(\mathcal{P})$ in Theorem 3.4. ("high metastability" $\Rightarrow \lambda_2 \approx 1$)

2. On the other hand the estimates (92) of Meyer have shown that if $\lambda_2 \approx 1$, then the eigenvalue problem for the weight computation in terms of (83) is ill-conditioned. ($\lambda_2 \approx 1 \Rightarrow$ "ill-conditioned weight computation")

Even small entries in $\mathcal{M}$ are very important for the computation of w and cannot be neglected. Therefore, it is necessary to focus on transition regions during the sampling, too, in order to get enough information about the rare event of conformational changes (i.e. in order to get the small entries right). Due to out-of-cluster rejections in algorithms for which the basis functions $\Phi_1, \ldots, \Phi_s$ are strict (characteristic functions of some sets $\mathcal{X}_j$), it is impossible to focus the sampling on transition regions, because after the proposal step of the HMC one has to check if the propagated position coordinates are still within the range of the observed basis function. In transition regions this mostly will not be the case causing the acceptance ratio of the HMC method to decrease. See also equation (54) and page 60. In meshless partitioning methods with a general uncoupling basis out-of-cluster rejections do not occur. In our numerical examples the improvement gained by the use of meshless methods can be seen on page 101, where the acceptance ratio of a transition region sampling is still about 80%, although the dynamics tend to leave the transition region in most cases.

5.2.3   Convergence Indicator

In short, this section shows that a restricted sampling converges badly, if there is a barrier inside the corresponding modified potential. And if all restricted samplings converge fast[18], then the corresponding shape functions are suitable for approximating almost invariant membership functions χ_l.

Convergence diagnostics.   From equation (83) arise two major sources of error for the weight computation. One is the condition κ of $\mathcal{M}$, which mainly depends on the number and the location of nodes and the α-parameter. The other source is the error of the Monte Carlo integration approach. This error mainly depends on the variance of the elements of $\mathcal{M}$ and the number of Monte Carlo steps, see (88). As the variance for $\mathcal{M} = P$ is assumed to be worse than the variance for $\mathcal{M} = S$ one may restrict the test of convergence to $\mathcal{M} = P$. The norm $\|\mathcal{M} - \mathcal{M}_{app}\|_\infty = \|E\|_\infty \leq \epsilon_E$ in (88) is small, if the sum of the absolute elements of each row of E is less than ϵ_E, because $\|E\|_\infty = \max_{i=1,\ldots,s} \|E(i,:)\|_1$. Each row $E(i,:)$ is related to one sampling of a modified potential V_i, see index j in equation (65) for the calculation of $\mathcal{M}$. This fact can be used to check for convergence. The sampling procedure based on the modified potential V_i can be stopped if

$$\|E(i,:)\|_1 \leq \epsilon_E, \tag{93}$$

where $\|\cdot\|_1$ is the sum of the absolute elements of the vector $E(i,:)$. In order to estimate this error during the sampling one starts c well dispersed Markov chains for each modified

---

[18]It is not sufficient that the samplings only converge. It is important that convergence is *fast*. In the sense of eq. (88), this means that the corresponding variance is small. Or: The sampling converges in a predefined "small" number of steps. If the number of steps until convergence is similar for each modified potential, then the partitioning was successful in terms of similar variance factors, see equation (53).

potential V_i. Then after a predefined number of steps one computes $\mathcal{M}_{app,k}(i,:)$ for all of these chains $k = 1, \ldots, c$, and estimates

$$\|E(i,:)\|_1 \approx \max_{k,l=1,\ldots,c} \|\mathcal{M}_{app,k}(i,:) - \mathcal{M}_{app,l}(i,:)\|_1 \tag{94}$$

for the mean value $\mathcal{M}_{app} = \frac{1}{c} \sum_{k=1}^c \mathcal{M}_{app,k}$.

Metastability of χ_l.   If χ_l is a metastable membership function, then

$$P^\tau \chi_l \approx \chi_l.$$

This means for the pure MD-part of the sampling algorithm (see Fig. 4 on page 62) with a high probability q_k^i stays in conformation l. In other words:

$$\chi_l(q_k^i) \approx \chi_l(\overline{q}_k^i)$$

for every sampling point $k = 1, \ldots, N_i$ and every sampling $i = 1, \ldots, s$ performed with a modified potential V_i. In Fig. 9 this situation is exemplified. If a point $q_k^i \in \Omega$ is in a region where $\chi_l(q_k^i) \approx 1$, then the probability that $\chi_l(\overline{q}_k^i) \approx 1$ is high. Also low χ_l values should stay low after a short dynamics simulation. If the node of a shape function Φ_i is located in the transition region of χ_l, then this leads to a *water shed* behavior. Recall that the shape functions can be seen as *convex* fuzzy sets, see Lemma 5.8. Therefore, taking different "directions" from within Φ_i "leads into" different neighboring shape functions Φ_j, i.e. different Markov chains have different Φ_j-patterns, which causes bad convergence according to (94). Only if the HMC-part of the algorithm for the restricted sampling according to V_i easily overcomes the probability barrier between high and low χ_l-values, convergence can be fast. As this barrier crossing is a rare event for the original potential function, this situation only takes place in an overall unprobable region, i.e. w_i would be small.

L^1-definition of variance.   In summary, convergence of the restricted sampling in the situation of Fig. 9 is bad, if the statistical weight w_i of Φ_i and the variance of χ_l w.r.t. Φ_i are not small, i.e. if the integral

$$I_{il} := \int_\Omega |\chi_l(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi}| \, \Phi_i(q) \, \pi(q) \, dq \tag{95}$$

is not small. Note that $w_i \geq I_{il}$ for $i = 1, \ldots, s$ and $l = 1, \ldots, n_C$. The integral I_{il} in (95) is denoted as L^1-*variance of χ_l inside Φ_i.*
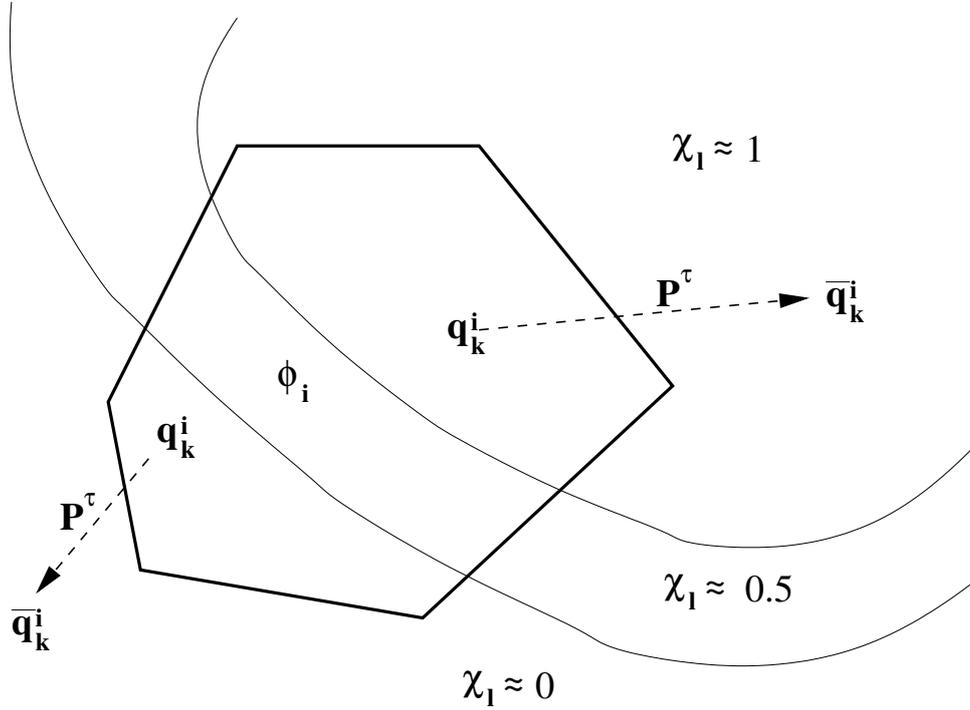
Figure 9: The variance of χ_l "inside" a shape function Φ_i is high (the shape function is illustrated by a convex polyeder similar to a Voronoi set, see Lemma 5.8). The pure MD part of the sampling algorithm (i.e. application of P^τ) hardly crosses the barrier between high and low membership values of metastable conformations χ_l in the configuration space Ω. This can be seen as a kind of *water shed* behaviour.

Good approximation results.   High metastability of χ_l together with high L^1-variance of χ_l inside a particular restricted sampling window implies bad convergence properties for the Monte Carlo integration method. On the other hand, if convergence of the restricted samplings is fast, although there are metastable membership functions $\chi_l : \Omega \to [0,1]$, we can therefore conclude that the L^1-variance of χ_l inside the sampling windows is low, i.e. there is a small $\epsilon > 0$ with $\epsilon > I_{il}$.

Lemma 5.13 *In the situation of Theorem 5.10 let $\epsilon > 0$ such that $\epsilon > I_{il}$, for some $l \in \{1, \ldots, n_C\}$ and for all $i = 1, \ldots, s$. Then $s\epsilon > \|\chi_l - \widetilde{\chi}_l\|_{L^1(\pi)}$.*

Proof: From the assumption

$$\epsilon > \int_\Omega |\chi_l(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi}| \, \Phi_i(q) \, \pi(q) \, dq, \quad \text{for all } i = 1, \ldots, s,$$

we arrive at

$$s\epsilon \;\; > \;\; \sum_{i=1}^s \int_\Omega |\chi_l(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi}| \, \Phi_i(q) \, \pi(q) \, dq$$

$$
\begin{aligned}
&= \int_\Omega \sum_{i=1}^s |\chi_l(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi}| \, \Phi_i(q) \, \pi(q) \, dq \\
&= \int_\Omega \sum_{i=1}^s |\chi_l(q)\Phi_i(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi} \Phi_i(q)| \, \pi(q) \, dq \\
&\geq \int_\Omega |\sum_{i=1}^s (\chi_l(q)\Phi_i(q) - \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi} \Phi_i(q))| \, \pi(q) \, dq \\
&= \int_\Omega |\chi_l(q) - \sum_{i=1}^s \frac{\langle \chi_l, \Phi_i \rangle_\pi}{\langle \Phi_i \rangle_\pi} \Phi_i(q)| \, \pi(q) \, dq \\
&\geq \|\chi_l - \widetilde{\chi}_l\|_{L^1(\pi)}.
\end{aligned}
\tag{96}
$$

For the last transformation compare (96) with the proof of Theorem 5.10. ☐

Lemma 5.13 shows that in the case of low L^1-variance the membership function is also well approximated by the given shape functions. In words, the important result is: *The approximation of the almost invariant membership functions is good, if each restricted sampling converges fast.*
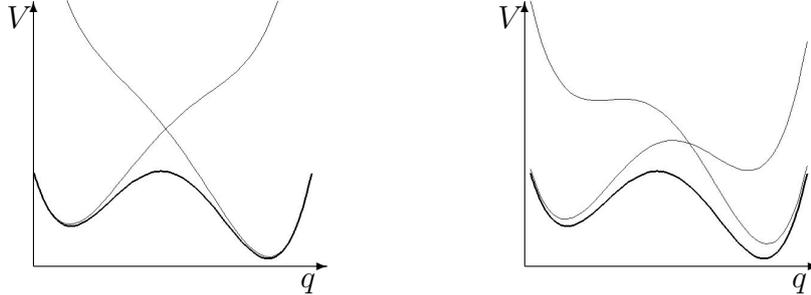
Figure 10: The original potential function $V(q)$ (thick line) with two modified potentials $V(q) - \beta^{-1}\ln(\Phi_i(q)), i = 1, 2$ (thin lines). The intersection point x of the modified potentials meets $0.5 = \Phi_1(x) = \Phi_2(x)$. *Left:* If Φ_1 and Φ_2 approximate the corresponding almost characteristic functions well, then the modified potentials do not include metastable substructures. *Right:* If the approximation is bad (see intersection point of thin lines), then a modified potential may include metastable substructures, which leads to bad convergence properties of the HMC.

Variance reduction. Convergence of the HMC part of the algorithm according to (89) or (91) is a prerequisite for good convergence properties of the MD part according to (90). If the MD part has good convergence properties, then the modified potentials cannot include almost invariant substructures, see Fig. 10. With Lemma 5.13 two kinds of result errors, i.e. the error w.r.t. the calculation of the statistical weights (see Section 5.2.2) and the error w.r.t. the approximation of almost invariant membership functions

(see Section 5.2.1), are summed up in only one indicator, namely that for the convergence properties of the restricted samplings according to (94). The speed of convergence can be determined by the variance of the elements of $\mathcal{M}$, see (89) and (90). In other words, variance reduction by partitioning methods is the key for both: good approximation of statistical weights and good approximation of almost invariant membership functions.

5.3 Hierarchical Refinement

Algorithm 5.11 is only an initial guess for a function-based decomposition of Ω. In this section the results of a meshless partitioning method are improved by further refinements of the uncoupling basis. In this section, a hierarchical refinement method is presented which allows a detailed reconstruction of P and S like in Section 4.5.

Performing a hierarchical refinement. For a given uncoupling basis

$$\{\Phi_1, \ldots, \Phi_s\} : \Omega \to [0, 1]$$

we want to replace a basis function (without loss of generality Φ_1) with a set $\{\Phi_{11}, \ldots, \Phi_{1\tilde{s}}\}$ of $\tilde{s}$ new basis functions, such that the basis expansion $\{\Phi_{11}, \ldots, \Phi_{1\tilde{s}}, \Phi_2, \ldots, \Phi_s\}$ is an uncoupling basis again. This can be done via the following method. First, find an uncoupling basis $\widetilde{\Phi}_{11}, \ldots, \widetilde{\Phi}_{1\tilde{s}} : \Omega \to [0, 1]$, i.e. basis functions which meet the partition of unity and the positivity constraint. Then, construct the basis expansion via:

$$\Phi_{1i}(q) := \Phi_1(q)\, \widetilde{\Phi}_{1i}(q), \quad q \in \Omega.$$

The coupling methods of Section 4.5 can be applied to the basis expansion, too, such that the expanded Galerkin discretization $P, S \in \mathbb{R}^{(s-1+\tilde{s}) \times (s-1+\tilde{s})}$ can be constructed. For the restricted sampling in equation (56) we have to replace Φ_1 with its hierarchical refinement, such that the new modified potentials $V_{1i} : \Omega \to \mathbb{R}$ for $i = 1, \ldots, \tilde{s}$ are given as[19]

$$
\begin{aligned}
V_{1i} &:= V - \frac{1}{\beta} \log(\Phi_{1i}) \\
&= V - \frac{1}{\beta} \log(\Phi_1 \widetilde{\Phi}_{1i}) \\
&= V - \frac{1}{\beta} \log(\Phi_1) - \frac{1}{\beta} \log(\widetilde{\Phi}_{1i}) \\
&= V_1 - \frac{1}{\beta} \log(\widetilde{\Phi}_{1i}).
\end{aligned}
\tag{97}
$$

Comparing equation (97) with (56) shows that a hierarchical refinement is nothing else but applying all above theory and methods to V_1 as "original potential" instead of V. E.g. for the selection of nodes this means that one can interpret the HMC sampling for V_1 as pre-sampling for V_{1i} and construct the partition of unity $\widetilde{\Phi}_{11}, \ldots, \widetilde{\Phi}_{1\tilde{s}}$ accordingly.

---

[19]Recall the definition of the modified potential V_1 in (56).

Advantage of meshless partitioning.   With the above hierarchical refinement method it is possible to get P and S for the "finest" uncoupling basis without using information of the preceding hierarchy steps. That means convergence of the samplings restricted according to the modified potentials is only necessary for the "leaves" of the discretization tree. If convergence of a restricted sampling is not reached in a predefined number of steps, then hierarchical refinement of the corresponding modified potential is necessary due to Section 5.2.3 for a better approximation of the conformations *and* for better convergence properties. Even if convergence is reached for every modified potential within the predefined number of steps, hierarchical refinement may also be desirable. Here are some refinement indicators.

Equilibrium state meets the detailed balance condition.   The statistical weights of the modified potentials are computed incorrectly, if the transition probabilities in P do not meet the detailed balance condition. Recall that $\overline{P}$ is a symmetric matrix due to momentum-reversibility of Hamiltonian dynamics. Calculation of the approximated weights $w_{app.,1}, \ldots, w_{app.,s}$ according to Theorem 4.11 yields an approximation $\overline{P}_{app.}$ for $\overline{P}$ via equations (60) and (61). If this matrix is not symmetric, i.e. for some $i, j \in \{1, \ldots, s\}$ we have

$$|\overline{P}_{app.}(i,j) - \overline{P}_{app.}(j,i)| > \gamma,$$

for a predefined $\gamma \in \mathbb{R}_{>0}$, then the transition probabilitie $P(i,j)$ and $P(j,i)$ are not well calculated. This means, one has to discretize the space between node i and node j in a better way, i.e. replace Φ_i and Φ_j by hierarchical refinements.

Condition of weight computation.   A direct refinement indicator which is based on the error of the weight computation can be given by equation (86). If we assume that the contribution to the deviation $\|E\|_\infty$ is almost equal for every sampling of a modified potential due to the convergence check in Section 5.2.3, then the error of weight computation $|w_i - w_{app,i}|$ is governed by the expression $w_{app,i} \|A_{(i)}^{-1}\|_\infty$. The sampling run with index i which maximizes this expression is also "responsible" for a poor condition number κ_{HH}, see equation (85). The corresponding shape function Φ_i should be subjected to further refinements.

6 Numerical Experiments

The following results are based upon an HMC sampling method according to Section 4.4. The construction of the uncoupling basis is described in Section 5.1 and Section 5.2.1, Algorithm 5.11. As a result of this sampling one gets the two matrices $\overline{S}$ and $\overline{P}$, see equation (35). The HMC sampling method is part of a software package called ZIBgridfree [94]. Further results in the present section are computed with Matlab [43] on the basis of $\overline{S}$ and $\overline{P}$. These results may differ from those obtained from the analysis routines of ZIBgridfree[20].

6.1 Interpretation of Function-Based Conformations

In this section we will investigate the molecules n-pentane and cyclohexane, which are shown in Fig. 11. The conventional view of conformations is a point concept. In most cases critical points, such as minima and saddle points, of the potential energy surface of the molecules are called "conformations", which does not include their flexibility and thermodynamically correct weighting according to entropical effects. In order to get a mathematically rigorous definition of conformations, Deuflhard et al. [25] introduced conformations as almost invariant *sets* in conformational space. From this point of view conformations are given by partial densities in configuration space, where these densities are restricted to separate sets. In terms of this set-based concept, the interpretation and visualization of "transition states" is difficult, because each spatial state is assigned to exactly one conformation. In the function-based concept each position state q is assigned to different conformations with a different degree of membership, such that transition regions can be extracted, if we define them as sets of points in configuration space which have a significant degree of membership w.r.t. more than one conformation.

Example: n-Pentane. The conformational space of n-pentane can mainly be described by two central dihedrals, see Fig. 11 left. For each of the dihedrals the values $180°$, $-120°$ and $+120°$, denoted as t, g- and g+, are preferred. Therefore, in chemical literature the nine combinations

$$(t/t),(g+/t),(g-/t),(t/g+),(g+/g+),(g-/g+),(t/g-),(g+/g-),(g-/g-)$$

of the two dihedrals are said to be "the conformations" of n-pentane. In contrast to this point concept, ZIBgridfree computes metastable conformations together with their flexibility and weights in conformational space. For the current example, a $300K$ computation of n-pentane with 60 basis functions and $\alpha = 11.3712$ has been taken from [93]. Since the numerical solution of the generalized eigenvalue problem (36) does not give $\lambda_1 = 1$ ($\bar{S}$ is

---

[20]In ZIBgridfree the matrix $\overline{S}$ is used for a geometrical clustering according to (44) with Robust Perron Cluster Analysis as described in Section 3 (local optimization of $I_2(\mathcal{A})$).
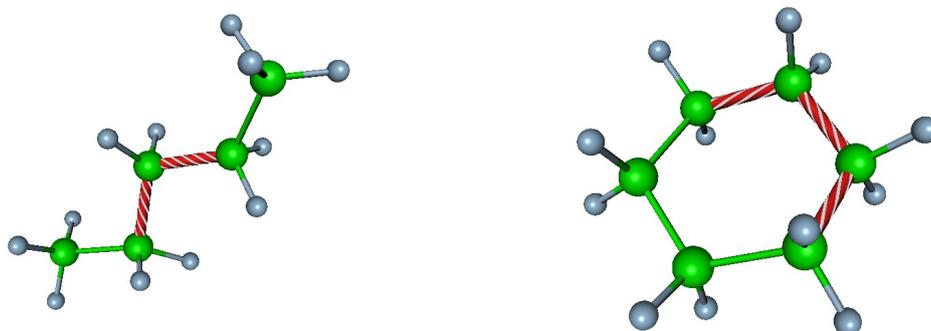
Figure 11: *Left:* n-Pentane with two dihedrals spanning the conformational space. *Right:* Cyclohexane with three dihedrals spanning the conformational space.

positive semi-definite), the dynamical clustering (45) is used for Robust Perron Cluster Analysis as derived in Section 3. In order to identify $n_C = 9$ as the number of conformations, the spectrum of the transition matrix P (see also Fig. 14 on page 97) has been examined like in Section 3.5.2. The result is a significant gap between the 9th and 10th eigenvalue: $\lambda_9 = 0.8948, \lambda_{10} = 0.6630$. An example for the visualization of the main conformation is given in Fig. 12 left. The weights and the density plots of the conformations agree[21] with the results of Schütte et al. [112], which are based on metastable sets instead of membership functions. The difference between our function-based approach and the set-based one is that a function-based description of conformations allows an overlap between different conformations. In this case, an assignment of position states to conformations based on membership functions provides a simple method for the visualization of transition states, too: E.g. in order to compute transition states for a $(t/t) \rightarrow (g+/t)$ transition of n-pentane one simply extracts the states in configuration space, which are assigned to both of the conformations with a degree of membership which is about 0.5. An example of such a transition state is shown in Figure 12 right. Obviously the clustering in terms of almost characteristic functions not only asigns membership values to position states, but also provides dynamical information about the molecule.

Example: Cyclohexane. Cylcohexan in Fig. 11 right is a 3-dimensional example. The values of three consecutive dihedral angles of the ring determine the conformation of cyclohexane. There are many differences between the n-pentane and the cyclohexane example.

---

[21]For a detailed numerical evaluation and justification of ZIBgridfree see Meyer [93]

Figure 12: *Example:* n-pentane. *Left:* 300K volume rendering in *amira/amiraMol*[122, 109] of the main (t/t)-conformation of n-pentane, which covers about 45% of the conformational space. *Right:* Transition state. 300K volume rendering of spatial states of n-pentane, which are assigned to the (t/t)- and to the (g+/t)-conformation with a degree of membership between 0.4 and 0.6.

- Taking a 2-dimensional domain, the density functions for the different conformations of n-pentane based on the two dihedral angles in Fig. 11 are bell-shaped and can sufficiently be approximated with Gaussians. Due to the ring structure, the density functions of the conformations of cyclohexane according to the three dihedrals in Fig. 11 are not bell-shaped. The conformational space includes much more complexity and higher energy barriers as n-pentane.

- The *chair position state* as local minimum of the energy surface of cyclohexane is characterized by an alternating dihedral angle value of $+55.9°$ and $-55.9°$ for dihedrals defined via 4 consecutive carbon atoms of the ring. With $\rho := 55.9°$, there are two different possibilities for this pattern

$$(+\rho, -\rho, +\rho, -\rho, +\rho, -\rho)$$

and

$$(-\rho, +\rho, -\rho, +\rho, -\rho, +\rho).$$

According to [133], these two minima and their "attraction basin" cover about 99.99% of the conformational space. Cyclohexane is an example for a molecule having two deep local minima with an high energy barrier between them, which also consists of some more local minima. I.e. in contrast to n-pentane, cyclohexane is an example for a conformation analysis in the presence of a transition region, see Fig. 1 (b) on page 5.
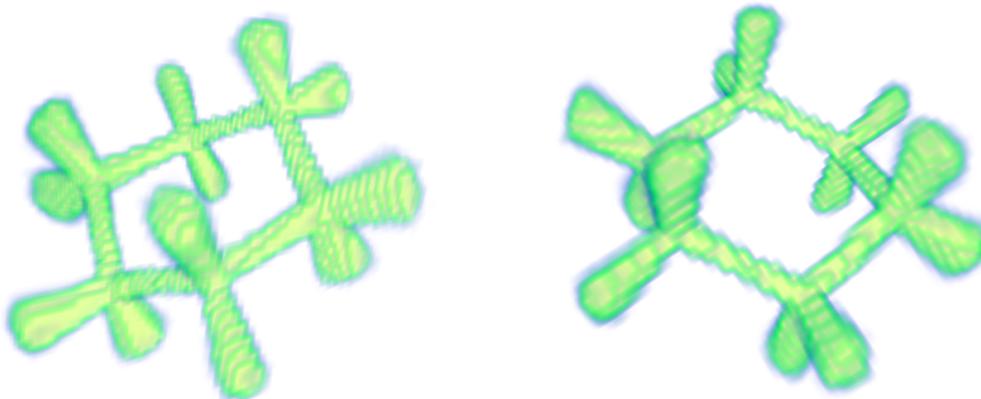
Figure 13: Volume rendering of the two cyclohexane conformations of a 300K sampling with
ZIBgridfree in *amira/amiraMol*[122, 109].

The result of ZIBgridfree with $s = 10$ is indeed a 2-clustering. In Fig. 13 the density plot of
the two conformations is shown. These conformations also look like the position state of
minimal energy of the chair conformation, but with the data provided by ZIBgridfree and
amiramol the inflexibility of these chairs can be visualized. Due to symmetry, the weights
of the two chair conformations should be equal. The robustness of ZIBgridfree with regard
to the weight computation of $\widetilde{w}_1$ and $\widetilde{w}_2$ using different random number sequences will be
examined in Section 6.4 on page 102. As mentioned above, the conformational space Ω
of cyclohexane is discretized into 10 shape functions. In comparison with n-pentane, this
means that not the dimension of the problem guides the number of basis functions, im-
portant is the number of conformations and their orientation in Ω. Complexity reduction
may avoid the curse of dimensionality.

6.2   Parameter Sensitivity of the Spectrum of P

The spectrum of P depends on the properties of the shape functions $\Phi_i^{[\alpha]}$, especially on the
parameter α, which controls the overlap between these basis functions. In Section 5.1.3 it
has been shown that α is a "soft parameter". In order to exemplify the dependence of the
spectra of $P^{[\alpha]}$ and $S^{[\alpha]}$ on the parameter α described in Section 5.1.3, one computes $P^{[\alpha]}$
and $S^{[\alpha]}$ for different α. Plotting the eigenvalues λ of these matrices for different α shows,
which eigenvalues are sensitive w.r.t. α. In Figure 14, an example is given which is taken
from a $300K$ sampling of pentane [93] with 60 shape functions for 3 different α-values[22]
$(11.3712, 12.8332, 14.2951)$. One can clearly see that the 10-th eigenvalue is the first one,

---

[22]These strange values arise from equation (80) for diffrent ϵ_2.

which changes significantly with α. The first 9 eigenvalues are nearly constant, i.e. for $n_C \leq 9$ the parameter α is soft w.r.t. the spectrum of P.
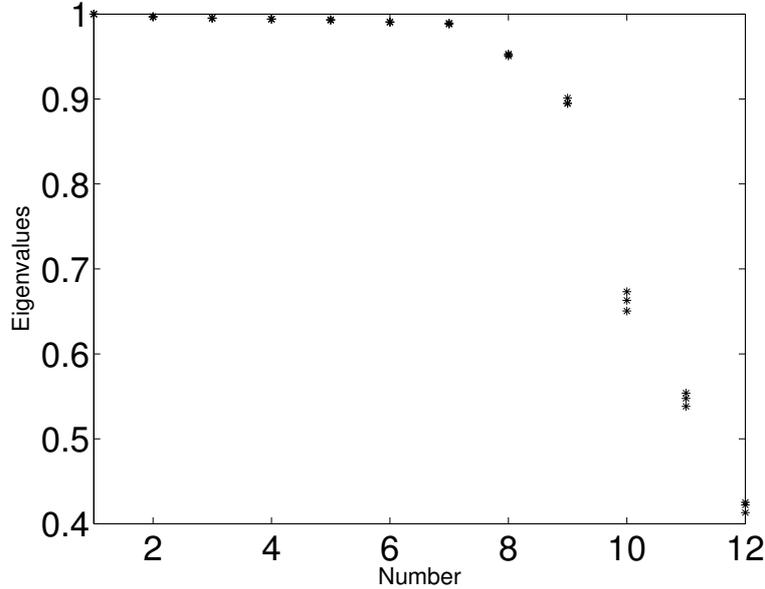
Figure 14: The first 12 eigenvalues of $P^{[\alpha]}$ for three different α. Whereas, the first 9 eigenvalues do not depend so much on α, the 10-th eigenvalue changes with α.

Example for different degrees of α-dependence. Example (33) has shown that some rows of the transition matrix P might not fit into the "almost block-structure" concept in Section 3.3. These rows represent transition states. The perturbation scheme in Figure 7 may be invalid for them, because deviation from block structure is not small for the corresponding rows of P, see Section 3.5.1. The following example shows that ,in fact, transition states may be affected by the α parameter. This is not alarming, because in Section 5.2.3 it has been shown that this kind of states has a small statistical weight or leads to bad convergence properties of the restricted samplings and, therefore, needs further refinements of the shape function basis anyway. In ϵ-perturbation analysis an $\mathcal{O}(\epsilon^2)$ perturbation result has been shown for almost characteristic functions, whereas eigenvectors are perturbed of order $\mathcal{O}(\epsilon)$. In the above context one can assume

$$\frac{\partial}{\partial \alpha} \chi_i^{[\alpha]} \approx 0, \quad i = 1, \ldots, n_C. \tag{98}$$

To find out, which shape function mainly affects a change of the block structure of $P^{[\alpha]}$ or $S^{[\alpha]}$ w.r.t. α, one can e.g. evaluate the following finite difference approach $\Delta \in \mathbb{R}^{s \times n_C}$:

$$\Delta := \frac{1}{h} \left| (M^{[\alpha^* + h]} - M^{[\alpha^* - h]}) \chi_{disc}^{[\alpha^*]} \right|. \tag{99}$$

In this expression $M^{[\alpha]}$ is either $S^{[\alpha]}$ or $P^{[\alpha]}$ for the corresponding parameter α and $\chi_{disc}^{[\alpha^*]}$ is the result of a Robust Perron Cluster Analysis, which should be almost independent from α due to (98). Δ can be seen as an approximation of

$$\Delta \approx \frac{\partial}{\partial \alpha} M^{[\alpha]} \chi_{disc}^{[\alpha]}|_{\alpha=\alpha^*}.$$

The key is: Δ should almost vanish, because $M^{[\alpha]}$ is a stochastic matrix for every α, i.e. the row-sums of the difference matrix $(M^{[\alpha^*+h]} - M^{[\alpha^*-h]})$ are zero. Furthermore, due to the assumption that α does not effect outer block elements, row-sums can be restricted to the particular blocks, which is done by multiplication with $\chi_{disc}^{[\alpha^*]}$. The interpretation of Δ is the following: If $\Delta(i,j)$ is high, then in the i-th row of M for the j-th block there is a significant change of block structure. Therefore, one should e.g. refine the shape function Φ_i. In the example of Figure 14 Δ has been evaluated for $M = P$. In this example, the partition function of the 9-th cluster is $\widetilde{w}_9 = 44\%$, i.e. $j = 9$ is the main conformation. The maximum entry in the column $\Delta(:,9)$ is $\Delta(46,9) = 0.0164$ followed by $\Delta(12,9) = 0.0142$ and some minor entries. In Figure 15 this situation is exemplified[23]. The selected shape functions Φ_{12} and Φ_{46} obviously represent transition states.
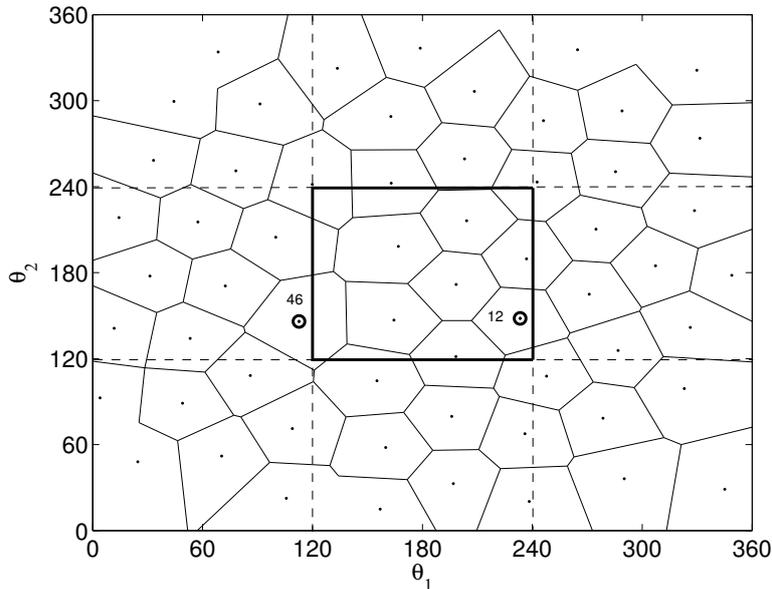
Figure 15: Voronoi Tessellation in dihedral space of pentane with 60 nodes. The central square indicates the "boundaries" of the main conformation. Two nodes are marked, which tend to effect the block structure of P, when α changes.

---

[23]In the spirit of Lemma 5.8 and for the reason of simplicity, a "real" Voronoi Tessellation for the 60 nodes has been generated in Matlab [43] to illustrate the situation $\alpha \to \infty$.

6.3   Epigallocatechin(-3-gallate)
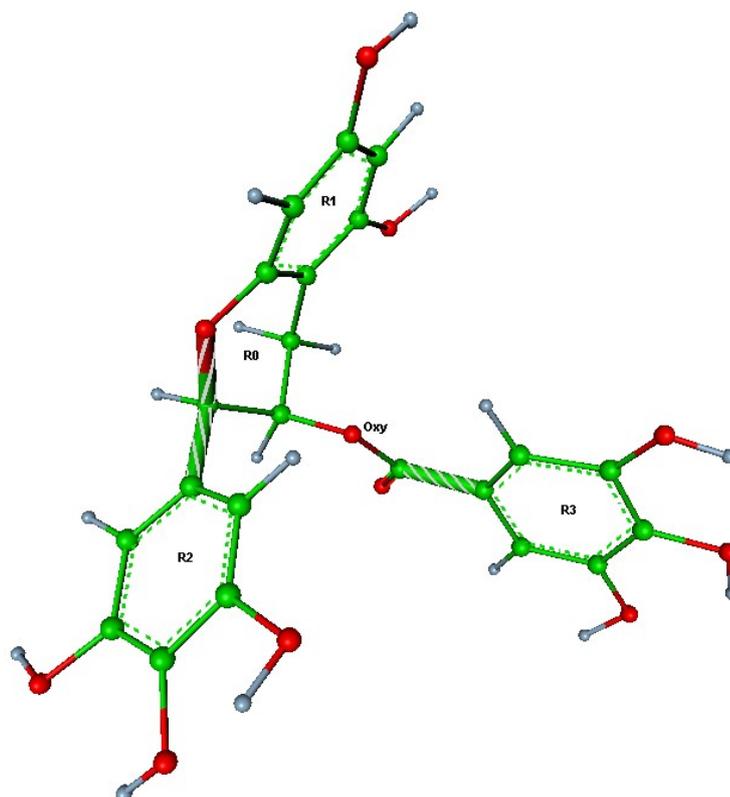
Figure 16: Epigallocatechin with three aromatic rings (R1, R2, R3) and one non-aromatic ring (R0). Its conformational space can be described by 7 dihedral angles. Like in cyclohexane, three dihedrals determine the form of the ring R0, one of these dihedrals is marked. Three dihedrals at the oxygen atom (Oxy) determine the position of the ring R3, one of these is marked. One dihedral determines the orientation of the ring R2.

For a line formula and for some remarks about the mode of action of epigallocatechin as a drug see Appendix A.4. Epigallocatechin in Fig. 16 has got different rings, three aromatic rings R1, R2, and R3, and one non-aromatic ring R0. This central non-aromatic ring consists of five carbon atoms and one oxygen (in Fig. 16 on the left side of the ring). ZIBgridfree has been applied to this molecule with $s = 158$ and $s = 398$ basis functions. With a total number of about 1.5 million resp. 1 million sampling points. Both approaches lead to the same results, therefore we only give the results for $s = 158$. In order to justify the correctness of the weight computation $w \in \mathbb{R}^s$ according to Theorem 4.11, note that a rotation of the rings R3 and R2 by an angle of $180°$ in Figure 16 leads to a chemically equivalent molecule. In Fig. 17 one can see a w-weighted histogram plot of the
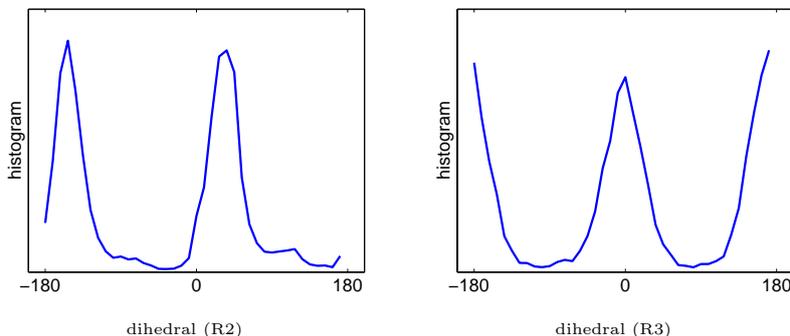
Figure 17: In Fig. 16, the two marked dihedrals at the aromatic rings R2 (left histogram) and R3 (right histogram) have a two-fold symmetry, which is almost preserved in a histogram plot of the 300K sampling result.

corresponding dihedral angles at R3 and R2 as they have been computed via ZIBgridfree. The 180°-symmetry is indeed preserved.

After weight computation, $\overline{S}$ and $\overline{P}$ can be determined via formula (60). Due to numerical errors, the evaluation of the generalized eigenvalue problem (36) does not lead to a Perron eigenvalue $\lambda_1 = 1$. Therefore, the Robust Perron Cluster Analysis with the dynamical clustering (45) has been applied to P. Like for cyclohexane, we get $n_C = 2$ metastable conformations. In this case the partition functions for the conformations are $\widetilde{w}_1 = 0.6$ and $\widetilde{w}_2 = 0.4$ for both experiments, $s = 158$ and $s = 398$. In Figure 18 left one can see that these conformations can be described by the dihedral angle at the oxygen atom of the ring R0 in Fig. 16. The dihedral angle either prefers values at about $+50°$ (for conformation 1) or at about $-50°$ (for conformation 2). This means that the conformations differ in the geometry of the non-aromatic ring R0. But this observation does not explain the different weights of the conformations. A closer look at the molecule shows that there is another indicator separating the two conformations, the distance between the oxygen atom marked as "oxy" in Fig. 16 and the oxygen atom of the ring $R0$. In Fig. 18 right one can see that for conformation 2 the oxygen atoms are closer than for conformation 1. Since both oxygen atoms carry negative partial charges, there is a repulsive force preferring a higher distance between them, which explains the different weights.

Figure 18 also shows that for conformation analysis a point concept is not adequate:

- In order to respond to the steric stress of two negative charges at a low distance of about 3Å in conformation 2, epigallocatechin compensates the unfavorable geometry by means of more flexibility than in conformation 1, where the partial negative charges are more separated from each other. In other words, in Figure 18 the histogram for conformation 2 is low but broad, whereas the histogram for conformation 1 is high and narrow. This "natural entropical response" leading to more flexibility cannot be modeled with conformation analysis based upon a point concept. The
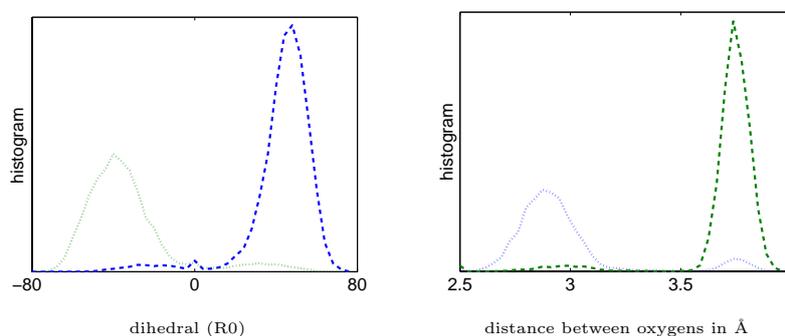
Figure 18: The separation of the two conformations of epigallocatechin are at best indicated by the different values of the marked dihedral in the ring R0 in Fig. 16, see left histogram, or by the distance of the oxygen atom of the ring R0 and the oxygen atom marked as "Oxy", see right histogram. *Dashed line:* Distribution of the values for conformation one, covering 60% of the overall histogram. *Dotted line:* Distribution for conformation two (40%).

difference of the global potential energy minima of the two conformations in Figure 18 is about 14 kJ/mol, with $\exp(-\beta\,14) \approx 0.004$ for $300K$, which does not provide any information about the correct ratio of thermodynamical weights.

- The coarseness of the clustering of epigallocatechin into two dynamically separated conformations cannot be achieved with geometry based cluster algorithms. The euclidian distance between two local minima of the energy surface does not include informations about the similarity of the corresponding position states or about the number, height and expansion of the energy barriers between them.

6.4 Acceptance and Convergence Properties

Out-of-cluster rejections. In ZIBgridfree the acceptance ratio of the HMC routine is always high. The so-called out-of-cluster rejections of a set-based approach like the one described on page 60 do not occur. In order to exemplify the advantage of meshless partitioning, remember that in the matrix $\overline{P}$, element $\overline{P}(i,i)$ denotes the overlap of a distribution sampled according to basis function Φ_i with its propagated distribution. If propagation did not "take place", then the entry $\overline{P}(i,i)$ would equal the entry $\overline{S}(i,i)$ of the mass matrix, which is the discretization of the identity operator. The relative difference between $\overline{P}(i,i)$ and $\overline{S}(i,i)$ can be seen as the expected out-of-cluster rejection ratio of a set-based approach. For the $12th$ basis function in the example of Figure 15 this relative difference is 50% for the $46th$ it is 33%, which means that a restricted sampling according to these basis functions has got a high tendency to leave the corresponding regions. The acceptance ratio of a set-based HMC-approach would decrease accordingly and convergence of the HMC method would be worse. In the given examples for n-pentane and cyclohexane, the acceptance ratio of the meshless approach was always above 80%.

"normE" convergence indicator (93) vs. Gelman-Rubin indicator. The next numerical example motivates the actual choice of the convergence indicator, which differs from indicators given in literature. In [133] Meyer and Weber computed the statistical weights for a 2-clustering of cyclohexane via ZIBgridfree with four different random number sequences. The two clusters mainly consist of the different chair conformations of cyclohexane and should be equally weighted because of the symmetry of the molecule. ZIBgridfree has been started with a constant step number (namely 40000) for each of the $s = 10$ modified potential samplings. The result is a mean deviation from symmetry of about 5%. Details are shown in Table 2.

random seed	% weight 1	% weight 2	% deviation
12	49.1	50.8	1.7
123	54.3	45.7	8.6
1234	49.4	50.6	1.2
12345	54.2	45.8	8.4

Table 2: ZIBgridfree experiment for cyclohexane with 10 basis functions and two clusters. ZIBgridfree has been started with different random seeds and each modified potential has been sampled with 40000 points.

random seed	% weight 1	% weight 2	% deviation
12	52.2	47.8	4.4
123	47.5	52.5	5.0
1234	53.7	46.3	6.4
12345	49.1	50.9	1.8

Table 3: ZIBgridfree experiment for cyclohexane with 10 basis functions and two clusters. ZIBgridfree has been started with different random seeds. The convergence of the modified potential samplings has been tested with the criterion (93).

random seed	% weight 1	% weight 2	% deviation
123	43.6	56.4	12.8
1234	33.9	66.1	32.2
12345	62.5	37.5	25.0
123456	60.1	39.9	20.2

Table 4: ZIBgridfree experiment for cyclohexane with 10 basis functions and two clusters. ZIBgridfree has been started with different random seeds. The convergence of the modified potential samplings has been tested with the standard Gelman-Rubin criterion.

For a second experiment, in the normE-criterion (93) the value ϵ_E has been chosen, such that the mean number of steps for the four different random sequences was $399500 \approx s \cdot 40000$ and therefore comparable to the constant step number experiment. The result is shown in Table 3, which provides a slightly lower mean deviation of the weights as in the first experiment.

The third experiment was performed with the standard Gelman-Rubin convergence indicator [42] instead of the normE-criterion. The result in Table 4 shows that the convergence diagnostics of Gelman and Rubin does not fit into the meshless partitioning concept. The mean deviation from symmetry is more than 20%, although the mean number of steps was 545500 and therefore much more than in Table 3.

6.5   Error Analysis

In this section one problem connected to the equilibrium sampling of high dimensional conformational spaces Ω is exemplified.

Small overlap.   In the calculation of pentane and cyclohexane, the normE convergence indicator (93) has been used. As shown in Section 5.2.3, the restricted sub-samplings only converge, if there is no metastability inside the proposed uncoupling basis functions. In order to meet this condition for the samplings of pentane and cyclohexane, the α-value has been chosen high enough to separate the sub-samplings well. If the overlap between the sub-samplings is small, the condition numbers κ_{IM} in (84) and κ_{HH} in (85) become poor, since $\mathcal{M}$ is nearly the identity matrix. E.g. the minimal condition number for pentane was higher than $\kappa = 200$. Combined with an error estimation of $\|E\|_\infty \approx 0.01$ this means that the error of weight computation (83) is higher than the weights themselves.

High overlap.   In the example of epigallocatechin, a small α-value has been chosen[24]. In this case, the overlap between the sub-samplings increases, which leads to a condition number of $\min\{\kappa_{HH}, \kappa_{IM}\} = 2.03$ and therefore to a meaningful weight computation. But, these results have also a drawback. Although, the total samplings of epigallocatechin comprise of about 1.5 million (for $s = 158$) or 1 million (for $s = 398$) sampling points, it has not converged according to the normE-criterion (93) with $\epsilon_E = 0.3$, which means that the error estimation of $\|E\|_\infty \approx 0.028$ according to (91) may be too optimistic.

Systematic problem.   The systematic problem of equilibrium sampling is the excluding relation between a well-conditioned weight computation and a fast converging sampling method. Meshless methods do not circumvent this problem, as the examples have shown.

---

[24]First experiment for epigallocatechin: $\alpha = 1.05$ with $\epsilon_2 = 0.001$, $\theta = 1.948$, $s = 158$ via equation (80), see Section 5.2.1. Second experiment for epigallocatechin: $\alpha = 1.62, \epsilon_2 = 0.001, \theta = 1.63, s = 398$. Whereas for the experiments with pentane and cyclohexane $\alpha > 10$ has been used.

Hierarchical function refinements like in Section 5.3 or more sophisticated sampling methods for the restricted sub-samplings can be further steps for improving the results. An advantage of the meshless approach is that there is a parameter α which can be used in order to either increase accuracy with worse convergence properties or in order to accelerate convergence getting a less accurate weight computation. Here it is important to mention that α has been shown to be a soft parameter corresponding to the computation of the conformations, see Section 5.1.3 equation (74).

7 Conclusion

In this thesis the conformation dynamics approach has been extended to a cluster analysis in terms of almost characteristic functions. This "soft" clustering concept can be applied whenever objects[25] occur that have to be assigned to different clusters with a non-vanishing degree of membership. For this purpose, we derived Robust Perron Cluster Analysis (PCCA+) from perturbation theory of stochastic matrices, which are computed as discretizations of some transfer operator here. For other applications, however, these matrices can also be based on any pairwise similarity measure. In contrast to other spectral clustering methods, PCCA+ exploits the simplex structure of the eigenvector data of the stochastic matrix, which leads to a simple, fast, and robust algorithm. Robustness has been shown by an $\mathcal{O}(\epsilon^2)$ perturbation result. PCCA+ is not only useful for molecular conformation analysis, but it is a multi-purpose algorithm used for various further applications in life sciences, see also [132] and [135].

An advantage of PCCA+ is that it defines *transition states*, i.e. objects which are assigned to different clusters with a significant degree of membership. In molecular dynamics these are states $q \in \Omega$ which have a low Boltzmann factor $\exp(-\beta V(q))$. Although, a change of the assignment of transition states w.r.t. the conformations does not have a significant influence in the thermodynamical weight computation, the memership values $\chi_i(q), i = 1, \ldots, n_C$, provide important information about the dynamical structure of the molecular system. In Section 3.4.5 it has been shown that this soft kind of clustering can be unique w.r.t. the degrees of membership.

The computation of membership functions implies a new interpretation of conformations as macrostates having their own modified potential energy functions which are defined for the complete set Ω of spatial states. Since a molecular system is uniquely defined by its potential energy function, one can compute any thermodynamical property for conformations, too, like observables, dynamics, and partition functions. From this point of view, conformations are more than only points or subsets of Ω.

The change of the point of view for conformation dynamics from a set-based concept to a function-based concept allows the use of meshless methods with overlapping basis functions. Only this kind of methods can break the curse of dimensionality in high-dimensional spaces in future.

Geometrical conformation analysis algorithms focus on minima of the potential energy landscape. However, from error analysis we derived that a balanced sampling of local minima and transition regions is necessary for a correct computation of thermodynamical weights, which led to an improved error-related convergence indicator in Section 5.2.3. This insight also implies that sampling of Ω and its analysis are two parts of the algorithm which cannot be separated from each other.

---

[25]In our case these objects are spatial states $q \in \Omega$.

A   Appendix

A.1   Gradient of a Modified Potential

Equation (71) is shown. For the reason of simplicity, define the nodal square distance functions $d_j : \Omega \to \mathbb{R}$ as $d_j(q) := d^2(q, q_j)$ for the nodes $q_1, \ldots, q_s \in \Omega$. The following equations hold:

$$\nabla(-\beta^{-1} \ln(\Phi_j(q)))$$

$$= -\frac{1}{\beta \, \Phi_j(q)} \nabla \Phi_j(q)$$

$$= -\frac{1}{\beta \, \Phi_j(q)} \left(\frac{\nabla \exp(-\alpha \, d_j(q))}{\sum\limits_{k=1}^{s} \exp(-\alpha \, d_k(q))} - \Phi_j(q) \frac{\sum\limits_{i=1}^{s} \nabla \exp(-\alpha \, d_i(q))}{\sum\limits_{k=1}^{s} \exp(-\alpha \, d_k(q))} \right)$$

$$= -\frac{1}{\beta \Phi_j(q)} \left(\frac{-\alpha \, \exp(-\alpha \, d_j(q)) \, \nabla d_j(q)}{\sum\limits_{k=1}^{s} \exp(-\alpha \, d_k(q))} - \Phi_j(q) \frac{\sum\limits_{i=1}^{s} -\alpha \, \exp(-\alpha \, d_i(q)) \, \nabla d_i(q)}{\sum\limits_{k=1}^{s} \exp(-\alpha \, d_k(q))} \right)$$

$$= -\frac{1}{\beta \Phi_j(q)} \left(-\alpha \, \Phi_j(q) \, \nabla d_j(q) - \Phi_j(q) [\sum\limits_{i=1}^{s} -\alpha \, \Phi_i(q) \, \nabla d_i(q)] \right)$$

$$= \frac{\alpha}{\beta} \left(-[\sum\limits_{i=1}^{s} \Phi_i(q) \, \nabla d_i(q)] + \nabla d_j(q) \right).$$

A.2   Parameter Dependence of a Membership Basis

Equation (73) is shown.

$$\frac{\partial}{\partial \alpha} \Phi_i^{[\alpha]}(q) \Big|_{\alpha = \alpha^*} = \frac{\partial}{\partial \alpha} \frac{\exp(-\alpha d^2(q, q_i))}{\sum\limits_{j=1}^{s} \exp(-\alpha d^2(q, q_j))} \Big|_{\alpha = \alpha^*}$$

$$= \frac{\frac{\partial}{\partial \alpha} \exp(-\alpha d^2(q, q_i))|_{\alpha = \alpha^*}}{\sum\limits_{j=1}^{s} \exp(-\alpha^* d^2(q, q_j))} - \Phi_i^{[\alpha^*]}(q) \frac{\frac{\partial}{\partial \alpha} \sum\limits_{j=1}^{s} \exp(-\alpha d^2(q, q_j))|_{\alpha = \alpha^*}}{\sum\limits_{k=1}^{s} \exp(-\alpha^* d^2(q, q_k))}$$

$$= -d^2(q, q_i) \, \Phi_i^{[\alpha^*]}(q) + \sum\limits_{j=1}^{s} \Phi_i^{[\alpha^*]}(q) \, d^2(q, q_j) \, \Phi_i^{[\alpha^*]}(q)$$

$$= -\sum\limits_{j=1}^{s} \Phi_j^{[\alpha^*]}(q) \, d^2(q, q_i) \, \Phi_i^{[\alpha^*]}(q) + \sum\limits_{j=1}^{s} \Phi_i^{[\alpha^*]}(q) \, d^2(q, q_j) \, \Phi_j^{[\alpha^*]}(q)$$

$$= \sum_{j=1}^{s} (d^2(q, q_j) - d^2(q, q_i)) \, \Phi_i^{[\alpha^*]}(q) \, \Phi_j^{[\alpha^*]}(q).$$

A.3 Monte Carlo Integration vs. Randomized Function Approximation

An example for breaking the curse of dimensionality is the Monte Carlo integration method, where the numerical effort does not depend on the dimension of the problem but on the variance of some integrand. Assume a sample set $\mathcal{Q} = \{q_1, \ldots, q_s\}$ of points $q_i \in \Omega$, which are distributed according to the Boltzmann distribution π, and uncoupling basis functions $\Phi_i(q) : \Omega \to [0, 1]$ associated with these sampling points via Shepard's method applied to radial basis functions. Then $\langle \Phi_i \rangle_\pi = 1/s$ for all $i = 1, \ldots, s$. For a sufficiently smooth function $f : \Omega \to \mathbb{R}$ and for this function basis $\{\Phi_i\}$ the error of Monte Carlo integration is:

$$
\begin{aligned}
e_{int} &= \left| \int_\Omega f(q) \, \pi(q) \, dq - \frac{1}{s} \sum_{i=1}^{s} f(q_i) \right| \\
&= \left| \int_\Omega f(q) \, \pi(q) \, dq - \sum_{i=1}^{s} \langle \Phi_i \rangle_\pi \, f(q_i) \right| \\
&= \left| \int_\Omega \left(f(q) - \sum_{i=1}^{s} f(q_i) \, \Phi_i(q) \right) \pi(q) \, dq \right|.
\end{aligned}
\tag{100}
$$

e_{int} does not suffer from the curse of dimensionality due to Monte Carlo integration theory. That the use of randomization is not only sufficient but also necessary for breaking the curse in numerical quadrature has been shown by Novak [98], where an exponential (w.r.t. the dimension of the function space) lower bound is derived for the integration error via positive quadrature formulas.

The weighted $L^1(\pi)$-error of randomized function approximation is:

$$
e_{approx} = \int_\Omega \left| f(q) - \sum_{i=1}^{s} f(q_i) \Phi_i(q) \right| \pi(q) \, dq.
\tag{101}
$$

Note that the equations (101) and (100) only differ in a permutation of integration and absolute value calculation. Therefore, the integration error e_{int} is always smaller than the approximation error e_{approx}. This leads to the conjecture that deterministic function approximation also suffers from the curse of dimensionality and only randomization can be the key to solve this problem in future. Here is the motivation for the conjecture

$$
e_{int}(\text{deterministic}) \leq e_{approx}(\text{deterministic})
$$

"curse of dimensionality″   "surely curse of dimensionality″

$$e_{int}(\text{randomized}) \quad \leq \quad e_{approx}(\text{randomized})$$
"no curse″   "possibly no curse″

A.4   Epigallocatechin

Epigallocatechin is an important substance of content of green tea. Empirical observations have shown that an ingestion of epigallocatechin together with other polyphenolic agents boosts the anti-oxidative activity in humans and animals and reduces cell death, cf. Kurihara et al. [75] pp. 441-442 and BBC News "Green tea 'may protect the heart' " from February 28th 2005 (available via http://news.bbc.co.uk/2/hi/health/4298403.stm). Anti-oxidative activity is important in order to prevent from cancer, heard disease, Parkinson's disease, reumatoid arthritis and Alzheimer's. For more detailed informations I recommend an introductory textbook of Montagnier et al. [96] on the relation between oxidative stress and diseases.
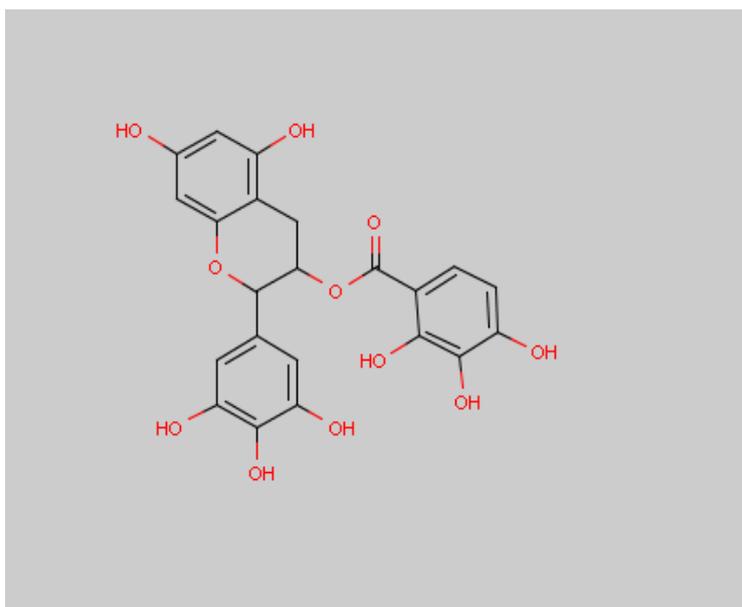
Figure 19: Line formula of epigallocatechin-3-gallate. For a 3D representation see Fig. 16.

B   Lebenslauf von Marcus Weber

20.03.1972	Geboren in Emsdetten. Nationalität: deutsch. Eltern: Bernhard Weber, Techniker, Birgitt Weber, geb. Watermeyer, Dipl. Sozialpädagogin (FH)
Schulbildung	kath. Grundschule in Emsdetten-Sinningen, städt. Gymnasium Martinum in Emsdetten
1991	Allgem. Hochschulreife mit Notendurchschnitt 1,3. Auszeichnung als bester Chemieabiturient der Jahrgangsstufe. Beginn des Zivildienstes in der Jugendherberge auf Juist.
1992	Westfälische-Wilhelms-Universität Münster. Immatrikulation mit Studienziel Diplom Chemiker.
1993-1999	Immatrikulation für das zusätzliche Studienfach Mathematik mit Nebenfach Chemie, Abschlussziel Diplom. März 1999: Abschluss als Diplom-Mathematiker, mit Auszeichnung, interdisziplinäres Thema der Arbeit: "Die Gaußtransformation als Homotopiemethode zur globalen Optimierung von Molekülpotentialfunktionen".
1999-2001	Zunächst: Betriebspraktikum in der Arbeitsvorbereitung bei der Firma Saertex in Saerbeck. Ab Mai 1999: Systementwickler und Consultant für ORACLE-Datenbanken bei der Firma TEAM GmbH in Paderborn.
seit April 2001	Wissenschaftlicher Angestellter am Zuse Institut Berlin (ZIB) im Bereich Computational Drug Design.

C Zusammenfassung

Die vorliegende Arbeit untersucht den Einsatz gitterfreier Methoden zur Beschreibung und zur Analyse von molekularen Konformationen. Einen Vorteil dieser Methoden hinsichtlich der Beschreibung von Konformationen stellt die Behandlung von Übergangszuständen dar. Hier kann der Zugehörigkeitsgrad eines Übergangszustandes zu verschiedenen Konformationen analysiert werden. Die Zugehörigkeitsgrade werden dabei mit Hilfe der Robusten Perron Cluster Analyse (PCCA+) ermittelt, die in dieser Arbeit detailliert untersucht wird.

Neben der Clusterung von Simulationsdaten widmet sich die vorliegende Arbeit auch der Durchführung von so genannten Monte-Carlo-Simulationen unter Einhaltung des gitterfreien Konzeptes und zeigt eine gitterfreie Partitionierungsmethode auf, mit deren Hilfe thermodynamische Gewichte der einzelnen Konformationen adaptiv ermittelt werden können.

Eine entsprechende Fehleranalyse hinsichtlich der thermodynamischen Gewichtsberechnung und der Approximation der Zugehörigkeitsfunktionen führt dabei zu einem ergebnisorientierten Abbruchkriterium für die Monte-Carlo-Simulation.

References

[1] E. Akhmatskaya and S. Reich. The targeted shadowing hybrid Monte-Carlo (tshmc) method. Technical report, Institut für Mathematik, Uni Potsdam, December 17, 2004.

[2] M. P. Allen and D. J. Tildesley. *Computer Simulations of Liquids*. Clarendon, Oxford, 1987.

[3] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J.Chem.Phys.*, 72:2384–2393, 1980.

[4] I. Babuška and J. M. Melenk. The partition of unity finite element method: Basic theory and applications. *Comput. Meths. Appl. Mech. Eng., Special Issue on Meshless Methods*, 139:289–314, 1996.

[5] I. Babuška and J. M. Melenk. The partition of unity method. *Int. J. Numer. Meth. Eng.*, 40:727–758, 1997.

[6] J. Baker. A algorithm for the location of transition states. *J. Comput. Chem.*, 7:385–395, 1986.

[7] R.B. Bapat and T.E.S. Raghavan. *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.

[8] R. Bellmann. *Adaptive Control Processes:A Guided Tour*. Princeton University Press, 1961.

[9] B. A. Berg and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order transitions. *Physical Review Letters*, 68(1):9–12, 1992.

[10] B. Berne and J. Straub. Novel methods of sampling phase space in simulation of biological systems. *Curr. Opinion Struct. Biol.*, 7:181, 1997.

[11] L. J. Billera and P. Diaconis. A geometric interpretation of the Metropolis-Hastings algorithm. *Statistical Science*, 16(4):335–339, 2001.

[12] P. G. Bolhuis, D. Chandler, Ch. Dellago, and P. L. Geissler. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Ann. Rev. Phys. Chem.*, 53:291–318, 2002.

[13] G. E. Cho and C. D. Meyer. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Lin. Alg. App.*, 335(1–3):137–150, 2001.

[14] K. L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer, New York, 2nd edition, 1967.

[15] B. W. Church, A. Ulitsky, and D. Shalloway. Macrostate dissection of thermodynamic monte carlo integrals. In D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, editors, *Advances in Chemical Physics*, volume 105 of *Monte Carlo Methods in Chemical Physics*, pages 273–310. John Wiley and Sons, Inc., 1999. Series Editors I. Prigognine and S. A. Rice.

[16] M. E. Clamp, P. G. Baker, C. J. Stirling, and A. Brass. Hybrid Monte Carlo: An efficient algorithm for condensed matter simulation. *J.Comput.Chem.*, 15(8):838–846, 1994.

[17] F. Cordes, M. Weber, and J. Schmidt-Ehrenberg. Metastable conformations via successive perron-cluster cluster analysis of dihedrals. Technical Report ZR-02-40, Zuse Institute Berlin, 2002.

[18] M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numer. Math.*, 75:293–317, 1997.

[19] M. Dellnitz and O. Junge. An adaptive subdivision technique for the approximation of attractors and invariant measures. *Comput. Visual. Sci.*, 1:63–68, 1998.

[20] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statis. Soc. B*, 39:1–38, 1977.

[22] P. Deuflhard. Order and stepsize control in extrapolation methods. *Numer. Math.*, pages 399–422, 1983.

[23] P. Deuflhard. Recent progress in extrapolation methods for ordinary differential equations. *SIAM Review*, 27:505–535, 1985.

[24] P. Deuflhard. From molecular dynamics to conformational dynamics in drug design. In M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, editors, *Trends in Nonlinear Analysis*, pages 269–287. Springer, 2003.

[25] P. Deuflhard, M. Dellnitz, O. Junge, and Ch. Schütte. Computation of Essential Molecular Dynamics by Subdivision Techniques. In P. Deuflhard, J. Hermans, B. Leimkuhler, A.E. Mark, S. Reich, and R.D. Skeel editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas. Lecture Notes in Computational Science and Engineering*, 4:98–115, Springer, 1999.

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

[27] P. Deuflhard and Ch. Schütte. Molecular conformation dynamics and computational drug design. In J. M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21th Century*. ICIAM 2003, Sydney Australia, 2004.

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

[29] G. M. Downs and J. M. Barnard. Clustering Methods and Their Uses in Computational Chemistry. *Rev. Comp. Chem.*, 18:12–13, 2002.

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

[31] C. A. M. Duarte and J. T. Oden. Hp clouds - a meshless method to solve boundary value problems. Technical Report TICAM Report 95-05, TICAM University of Texas, 1995.

[32] G. E. Fasshauer. Approximate moving least-squares approximation with compactly supported radial weights. In M. Griebel and M. A. Schweitzer, editors, *Meshfree Methods for Partial Differential Equations*, volume 26 of *Lecture Notes in Computational Science and Engineering*, pages 105–116. Springer, Berlin, Heidelberg, New York, 2003.

[33] A. Fischer. Die Hybride Monte–Carlo Methode in der Molekülphysik. Diploma thesis, Department of Mathematics and Computer Science, Free University Berlin, (in German), 1997.

[34] A. Fischer. An uncoupling-coupling technique for Markov chain Monte Carlo methods. Technical Report ZR-00-04, Zuse Institute Berlin, 2000.

[35] A. Fischer, F. Cordes, and C. Schütte. Hybrid Monte Carlo with adaptive temperature in a mixed–canonical ensemble: Efficient conformational analysis of RNA. *J. Comput. Chem.*, 19:1689–1697, 1998.

[36] A. Fischer, Ch. Schütte, P. Deuflhard, and F. Cordes. Hierarchial uncoupling–coupling of metastable conformations. In *[107]*, pages 235–259, 2002.

[37] B. M. Forrest and U. W. Suter. Hybrid Monte Carlo simulations of dense polymer systems. *J.Chem.Phys.*, 101, 1994.

[38] D. Frenkel and B. Smit. *Understanding Molecular Simulation – From Algorithms to Applications*, volume 1 of *Computational Science Series*. Academic Press, A Division of Harcourt, Inc., www.academicpress.com/computationalscience, 2nd edition, 2002.

[39] T. Galliat. *Adaptive Multilevel Cluster Analysis by Self-Organizing Box Maps.* PhD thesis, FU Berlin, March 2002.

[40] T. Galliat and P. Deuflhard. Adaptive hierarchical cluster analysis by self-organizing box maps. Technical Report SC-00-13, Zuse Institute Berlin, 2000.

[41] T. Galliat, P. Deuflhard, R. Roitzsch, and F. Cordes. Automatic identification of metastable conformations via self–organized neural networks. In *[107]*, pages 260–284, 2002.

[42] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–511, 1992.

[43] TheMathWorks Inc. Germany. Matlab(R) 6.5.0, 1994–2005.

[44] G.H. Golub and C.F. van Loan. *Matrix Computations.* Johns Hopkins University Press, 3rd edition, 1996.

[45] M. Griebel and M. A. Schweitzer. A particle-partition of unity method for the solution of elliptic, parabolic and hyperbolic pde. *SIAM J. Sci. Comp.*, 22(3):853–890, 2000.

[46] D. G. Gromov and J. J. de Pablo. Structure of binary polymer blends - multiple time step hybrid Monte Carlo simulations and self-consistent integral-equation theory. *J.Chem.Phys.*, 103(18):8247–8256, 1995.

[47] S. Gupta, A. Irbäck, F. Karsch, and B. Petersson. The acceptance probability in the hybrid Monte Carlo method. *Phys. Lett. B*, 242:437–443, 1990.

[48] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration – Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics.* Springer, corr. 2nd edition, 2004.

[49] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems.* Springer-Verlag, Berlin, Heidelberg, 2nd edition, 1993.

[50] T.A. Halgren. *J. Am. Chem. Soc.*, 114:7827–7843, 1992.

[51] T.A. Halgren. Merck molecular force field. *J. Comp. Chem.*, 17(I-V):490–641, 1996.

[52] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods.* J. Wiley, New York, 1964.

[53] F. A. Hamprecht, Ch. Peter, X. Dura, W. Thiel, and W. F. van Gusteren. A strategy for analysis of (molecular) equilibrium simulations: Configuration space density estimation, clustering and visualization. *J. Chem. Phys.*, 114(5):2079–2089, 2001.

[54] U. Hansmann and Y. Okamoto. Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem. *J. Comput. Chem.*, 14:1333, 1993.

[55] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger. Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem.Phys.Lett.*, 259:321–330, 1996.

[56] M. Haviv and L. van der Heyden. Perturbation bounds for the stationary probabilities of a finite Markov chain. *Adv. Appl. Prob.*, 16:804–818, 1984.

[57] H. Hofer and E. Zehnder. *Symplectic Invariants and Hamiltonian Dynamics.* Birkhäuser Verlag, 1998.

[58] R. Horst and P. M. Pardalos, editors. *Handbook of Global Optimization.* Nonconvex Optimization and Its Applications. Kluwer Academic Publishers, 1995.

[59] R. Horst, P. M. Parlados, and N. V. Thoai. *Introduction to Global Optimization.* Nonconvex Optimization and Its Applications. Kluwer Academic Publishers, 1995.

[60] W. Huisinga. *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics.* PhD thesis, Free University Berlin, 2001.

[61] W. Huisinga, S. Meyn, and Ch. Schütte. Phase transitions & metastability in Markovian and molecular systems. *Ann. Appl. Probab.*, 14:419–458, 2004.

[62] W. Huisinga and B. Schmidt. Metastability and dominant eigenvalues of transfer operators. In C. Chipot, R. Elber, A. Laaksonen, B. Leimkuhler, A. Mark, T. Schlick, C. Schütte, and R. Skeel, editors, *New Algorithms for Macromolecular Simulation*, *Lecture Notes in Computational Science and Engeneering*, 49:167–182, Springer, 2005.

[63] I. Ipsen and C. D. Meyer. Uniform stability of markov chains. *SIAM J. Matrix Anal. Appl.*, 15:1061–1074, 1994.

[64] A. Irbäck. Hybrid Monte Carlo simulation of polymer chains. *J.Chem.Phys.*, 101(2):1661–1667, 1994.

[65] J. A. Izaguirre and S. S. Hampton. Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules. *J. Comput. Phys.*, 200:581–604, 2004.

[66] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1976.

[67] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A34:827–828, 1978.

[68] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[69] T. Kato. *Perturbation Theory for Linear Operators.* Springer, Berlin, 1995. Reprint of the 1980 edition.

[70] A. D. Kennedy and B. Pendleton. Acceptances and autocorrelation in hybrid Monte Carlo. *Nuclear Physics B Proceedings Supplements*, 20:118–121, 1991.

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

[72] S. Kirkland, M. Neumann, and B. Shader. Applications of paz's inequality to perturbation bounds for markov chains. *Lin. Alg. App.*, 268:183–196, 1998.

[73] T. Kohonen. *Self–Organizing Maps.* Springer, Berlin, 2nd edition, 1997.

[74] I. Kolossvary and W. C. Guida. Comprehensive conformational analysis of the four- to twelve-membered ring cycloalkanes:Identification of the complete set of interconversion pathways on the MM2 potential energy hypersurface. *J. Am. Chem. Soc.*, 115:2107–2119, 1993.

[75] H. Kurihara, L. Cheng, B. Zhu, Z. He, H. Shibata, Y. Kiso, T. Tanaka, and X. Yao. Anti-Stress Effect of Oolong Tea in Women Loaded with Vigil. *J. of Health Science*, 49(6):436–443, 2003.

[76] J. C. Langarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, 9(1):112–147, 1998.

[77] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education Limited, 2nd edition, 1999.

[78] A. Lew, J. E. Marsden, M. Ortiz, and M. West. An overview of variational integrators. In L. P. Franca, T. E. Tezduyar, and A. Masud, editors, *Finite Element Methods: 1970's and Beyond*, pages 98–115. CIMNE, ISBN 84-95999-49-8, 2004.

[79] S. Li and W. K. Liu. *Meshfree Particle Methods*. Springer, Berlin, Heidelberg, 2004.

[80] G. R. Liu. *Mesh Free Methods – Moving beyond the Finite Element Method*. CRC Press, 2002.

[81] R. Lowen. Convex fuzzy sets. *Fuzzy Sets and Systems*, 3:291–310, 1980.

[82] D. S. Mackey and N. Mackey. On the determinant of symplectic matrices. *Numerical Analysis Report No. 422*, February 2003. Manchester Centre for Computational Mathematics.

[83] V. I. Manousiouthakis and M. W. Deem. Strict detailed balance is unnecessary in Monte Carlo simulation. *J. Chem. Phys.*, 110:2753–2756, 1999.

[84] E. Marinari and G. Parisi. Simulated tempering: A new monte carlo scheme. *Europhys. Lett.*, 19:451, 1992.

[85] G. J. McLachlan and K. E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel-Dekker, New York, 1988.

[86] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley, New York, 2000.

[87] E. Meerbach, A. Fischer, and Ch. Schütte. Eigenvalue bounds on restrictions of reversible nearly uncoupled markov chains. *submitted to: J. Lin. Alg. App.*, 2003. available via http://page.mi.fu-berlin.de/~biocomp/publications.html.

[88] B. Mehlig, D. W. Heermann, and B. M. Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Phys.Rev. B*, 45(2):679–685, 1992.

[89] M. Meila and J. Shi. A random walks view of spectral segmentation, 2001.

[90] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J.Chem.Phys.*, 21:1087–1092, 1953.

[91] C. D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Rev.*, 31:240–272, 1989.

[92] C. D. Meyer. Sensitivity of the stationary distribution of a Markov chain. *SIAM Journal on Matrix Analysis and Applciations*, 15:715–728, 1994.

[93] H. Meyer. Die Implementierung und Analyse von HuMFree – einer gitterfreien Methode zur Konformationsanalyse von Wirkstoffmolekülen. Master's thesis, Free University Berlin, 2005. In German.

[94] H. Meyer, M. Weber, and A. Riemer. ZIBgridfree. Software package for HMC simulation and conformation analysis based upon C++ classes of amiraMol [109] using the Merck Molecular Force Field [50, 51] implemented by T. Baumeister and parametrized by F. Cordes. Robust Perron Cluster Analysis implemented by M. Weber and J. Schmidt-Ehrenberg. VERX (extrapolation method based on Verlet) implemented by U. Nowak, Status: January 2005. Software owned by Zuse Institute Berlin.

[95] L. Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford*, 11:50–59, 1960.

[96] L. Montagnier, R. Olivier, and C. Pasquier, editors. *Oxidative Stress in Cancer, AIDS, and Neurodegenerative Diseases*, volume 1 of *Oxidative Stress and Disease*. Marcel Dekker, 1st edition, 1998.

[97] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer J.*, 7:308–313, 1965.

[98] E. Novak. Is there a curse of dimension for integration? In C.-H. Lai, P. E. Bjorstad, M. Cross, and O. B. Widlund, editors, *Eleventh Int. Conf. on Domain Decomposition Methods*, pages 88–95. Domain Decomposition Press, Bergen, 1999.

[99] M. Orešič and D. Shalloway. Hierarchical characterization of energy landscapes using Gaussian packet states. *J. Chem. Phys.*, 101(11):9844–9857, 1994.

[100] J. O'Rourke. *Computational geometry in C.* Cambridge Univeristy Press, 1994.

[101] P. H. Peskun. Optimum Monte Carlo sampling using Markov chains. *Biometrika*, 60:607–612, 1973.

[102] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–15, January 1986.

[103] Ch. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer New York, Berlin, Heidelberg, 1999.

[104] R.Y. Rubinstein. *Simulation and the Monte Carlo Method.* J. Wiley, New York, 1981.

[105] J. Rust. Using randomization to break the curse of dimensionality. *Econometrica*, 65(3):487–516, 1997.

[106] K. Schäfer. *Statistische Theorie der Materie*, volume 1. Vandenhoeck & Ruprecht in Göttingen, 1960. In German.

[107] T. Schlick and H. H. Gan, editors. *Computational Methods for Macromolecules: Challenges and Applications – Proc. of the 3rd Intern. Workshop on Algorithms for Macromolecular Modelling*, Berlin, Heidelberg, New York, 2000. Springer.

[108] J. Schmidt-Ehrenberg. A density mixture model for molecular conformation distributions. Technical Report ZR-05-02, Zuse Institute Berlin, 2005. In preparation.

[109] Johannes Schmidt-Ehrenberg, Daniel Baum, and Hans-Christian Hege. Visualizing dynamic molecular conformations. In *IEEE Visualization 2002*, pages 235–242. IEEE Computer Society Press, 2002.

[110] B. J. Schulz, K. Binder, and M. Mueller. Flat histogram method of Wang-Landau and n-fold way. *Int. J. Mod. Phys. C13*, page 477, 2002.

[111] Ch. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules.* Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1999.

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

[113] Ch. Schütte, R. Forster, E. Meerbach, and A. Fischer. Uncoupling-coupling techniques for metastable dynamical systems. In *Proc. of the 15th Int. Conf. on Domain Decomposition Methods (accepted)*, 2004. available via http://page.mi.fu-berlin.de/~biocomp/publications.html.

[114] Ch. Schütte, W. Huisinga, and P. Deuflhard. Transfer operator approach to conformational dynamics in biomolecular systems. In B. Fiedler, editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 191–223. Springer, 2001.

[115] M. A. Schweitzer. Ein Partikel-Galerkin-Verfahren mit Ansatzfunktionen der Partition of Unity Method. Master's thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Angewandte Mathematik, for: Prof. Dr. M. Griebel, December 1997. In German.

[116] D. Shalloway. Macrostates of classical stochastic systems. *J. Chem. Phys.*, 105(22):9986–10007, 1996.

[117] D. Shephard. A two-dimensional function for irregularly spaced data. *In Proc. 23rd AMC National Conference*, pages 517–523, 1968.

[118] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[119] B. W. Silverman. *Density Estimation*. London: Chapman and Hall, 1986.

[120] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29:111–138, 1961.

[121] R. D. Skeel and D. J. Hardy. Practical construction of modified Hamiltonians. *SIAM Journal on Scientific Computing*, 23(4):1172–1188, 2002.

[122] D. Stalling, M. Westerhoff, and H.-Ch. Hege. Amira - a highly interactive system for visual data analysis. In Christopher R. Johnson and Charles D. Hansen, editors, *Visualization Handbook*. Academic Press, November 2004.

[123] G.M. Torrie and J.P. Valleau. Monte carlo study of a phase-separating liquid mixture by umbrella sampling. *J.Chem.Phys.*, 66(4):1402–1408, February 1977.

[124] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte-carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.*, 23:187–199, 1977.

[125] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[126] D. Verma and M. Meila. A Comparison of Spectral Clustering Algorithms. Technical Report 03-05-01, University of Washington, 2003.

[127] P. Virnau and M. Müller. Calculation of free energy through successive umbrella sampling. *J. Chem. Phys.*, 120(23):10925–10930, 2004.

[128] P. Virnau and M. Müller. A successive umbrella sampling algorithm to sample and overcome free energy barriers. *Computer Simulation Studies in Condensed Matter Physics*, 17, 2004. D. P. Landau and S. P. Lewis and H.-P. Schüttler (eds.) (in press).

[129] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050, 2001.

[130] M. Weber. Clustering by using a simplex structure. Technical Report ZR-04-03, Zuse Institute Berlin, 2004.

[131] M. Weber and T. Galliat. Characterization of transition states in conformational dynamics using Fuzzy sets. Technical Report ZR-02-12, Zuse Institute Berlin, March 2002.

[132] M. Weber and S. Kube. Robust Perron Cluster Analysis for Various Applications in Computational Life Science. In M. R. Berthold et al., editors, *CompLife 2005*, pages 57–66. Springer-Verlag Berlin Heidelberg, 2005.

[133] M. Weber and H. Meyer. ZIBgridfree – Adaptive Conformation Analysis with qualified Support of Transition States and Thermodynamic Weights. Technical Report ZR-05-17, Zuse Institute Berlin, 2005.

[134] M. Weber, W. Rungsarityotin, and A. Schliep. Perron cluster analysis and its connection to graph partitioning for noisy data. Technical Report ZR-04-39, Zuse Institute Berlin, 2004. Submitted to: *J. of Classification*.

[135] M. Weber, W. Rungsarityotin, and A. Schliep. An indicator for the number of clusters using a linear map to simplex structure. *Accepted for: From Data and Information Analysis to Knowledge Engineering, 29th Annual Conference of the German Classification Society*, 2005.

[136] J. M. Wendlandt and J. E. Marsden. Mechanical integrators derived from a discrete variational principle. *Physica D*, 106:223–246, 1997.

[137] Z. Wu and R. Leahy. An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation. *IEEE Tr. Pattern Analysis and Machine Intelligence*, 15(11):1101–1113, 1993.

[138] H. Zhang. A new hybrid Monte Carlo algorithm for protein potential function test and structure refinement. *Proteins*, 34:464–471, 1999.

Danksagung