

Master thesis

Computing the minimal rebinding effect for nonreversible processes

**Zur Erlangung des akademischen Grades
Master of Science**

Susanne Röhl

Berlin, den 8. Februar 2017

Contents

1	Markov State Models	3
1.1	Markov Process	3
1.2	Transfer Operator	6
1.3	Galerkin Projection	9
1.4	Recrossing Effect	10
2	Metastability and Dominant Structures	12
2.1	Concept of metastability applied to MD	12
2.2	Spectral Approach	15
2.3	Decomposition via membership functions	17
2.4	Dominant cycles (non reversible NESS process)	20
3	Rebinding Effect in a Given Kinetics	21
3.1	Receptor-Ligand System	21
3.2	Rebinding Effect	21
3.3	Molecular Kinetics as a Projection	22
3.4	Minimizing the Rebinding Effect	26
3.5	Approach for nonreversible processes	26
4	Illustrative Examples	27
4.1	An artificial (bivalent) binding Process	27
Bibliography		ii Eine Randnotiz

1 Markov State Models

Stochastic processes, especially Markov processes, are used in a lot of applications in different areas, like biotechnology or Simulations of biomolecular systems (in atomic representation) require often timescales that are far beyond the capacity of computer power currently available (for more detailed example see Anton). To get a simulation result in a reasonable time, it makes sense to consider a reduced model of that stochastic process which maintains the relevant dynamical properties while at the same time being less complex. Such reduced models are called Markov State Models. There has been a lot of investigations/ recent research activity during the last years, see ...

In order to define/create a Markov State Model, we need at first some basic definitions of stochastic processes, especially Markov processes and how their evolution can be described using the transfer operator. The actual dimension reduction of the process happens by applying a Galerkin projection onto the transfer operator. By that action, states of the original process are clustered/ grouped conveniently, such that .. properties/ transition rates?.. are preserved..

1.1 Markov Process

A Markov process is a certain/particular stochastic process with some nice properties which makes it easy to compute with. They are defined on possibly continuous state space and time, in opposite to a Markov chain where both these properties are discrete (finite state space?, transition matrix = stochastic matrix).

Transition function

From now on we will denote by $\mathbb{X} := (\mathbb{X}, \Sigma)$ a measurable space, i.e. a set \mathbb{X} with some σ -algebra Σ defined on it. And $\Omega := (\Omega, \mathcal{A}, \mathbb{P})$ will be a probability space, i.e. a measurable space with a probability measure \mathbb{P} defined on it; for detailed information about these basic notions see

X with Borel-sigma-alg.?

A random variable $X : \Omega \rightarrow \mathbb{X}$ is a measurable function from a probability space Ω into a measurable space \mathbb{X} , i.e. preimages of measurable sets in \mathbb{X} are measurable

in Ω :

$$A \in \Sigma \Rightarrow X^{-1}(A) \in \mathcal{A}.$$

Then the probability measure \mathbb{P} of Ω induces a canonical probability measure on \mathbb{X} , also denoted by \mathbb{P} , by $\mathbb{P}(A) := \mathbb{P}(X \in A) := \mathbb{P}(X^{-1}(A))$ for all $A \in \Sigma$, see (...).

Definition 1.1 (Stochastic Process). A family $(X_t)_{t \in \mathbb{T}}$ of random variables $X_t : \Omega \rightarrow \mathbb{X}$ on some index set \mathbb{T} is called a stochastic process on a state space \mathbb{X} .

einschränkung
 $\mathbb{R}^n + \text{Borel?}$

In order to introduce Markov processes as a special case of stochastic processes, we need a tool to describe the time evolution or propagation of the process. This can be done using the transition function which describes the propagation of the distribution functions of a stochastic process.

Definition 1.2 (Transition function). A function $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$ is a transition function if it fulfills the following properties:

assigned to a
process?

- i) $x \mapsto p(t, x, A)$ is measurable for all $t \in \mathbb{T}$ and $A \in \mathcal{B}(\mathbb{X})$
- ii) $A \mapsto p(t, x, A)$ is a probability measure for all $t \in \mathbb{T}$ and $x \in \mathbb{X}$
- iii) $p(0, x, \mathbb{X} \setminus x) = 0$ for all $x \in \mathbb{X}$
- iv) the Chapman-Kolmogorov equation

?

$$p(t + s, x, A) = \int_{\mathbb{X}} p(t, x, dz) p(s, z, A). \quad (1.1)$$

holds for all $t, s \in \mathbb{T}, x \in \mathbb{X}$ and $A \subset \mathbb{X}$.

Here the first three properties just ensure that we get reasonable (measurable) results and that the process can only be in one state at a certain time and not make a jump (a transition in 0-time).

So the transition function $p(t, x, A)$ can be considered as the probability to get into a certain subset A in a time interval t starting from a point x . That follows from the Chapman-Kolmogorov equation, see (...). That means that we can describe the time evolution of a stochastic process by a transition function. In particular, the transition matrix of a Markov chain (time discrete, finite state space) is a special case of the transition function since it fulfills the above properties.

why? how?

Markov Process

Now we can define Markov processes as a special case of stochastic processes.

Definition 1.3 (Markov Process). A stochastic process $(X_t)_{t \in \mathbb{T}}$ on a state space \mathbb{X} is a Markov process if its transition function fulfills the equation

$$p(t, x, A) = \mathbb{P}(X_{t+s} \in A \mid X_s = x). \quad (1.2)$$

for all $s, t \in T$, $x \in \mathbb{X}$ and $A \subset \mathbb{X}$. If that probability is independent from s , then the Markov process is called (time-)homogeneous.

We will be interested only in homogeneous processes. As we can see from the definition the time evolution of a Markov process is described by its transition function. It is a process that has "no memory" in the sense that only the last known state of the process has an influence of the future of the process; as we can see on the right side of the equation. ?

Indeed, there is a one-to-one relation between transition functions and (time-homogeneous) Markov processes, i.e. every homogeneous Markov process defines a transition function and vice versa, see Meyn and Tweedie[5, chapter 3].

The transition function for a Markov process plays the same role as the transition matrix for a Markov chain; it propagates its distributions? If for the transition function we choose $t = 1$ and transitions into one-elementic subsets, then the transition function corresponds to the 1-step transition matrix of a Markov chain. Having introduced the notion of Markov Processes, we can now define some important properties and give some examples. time homo-geneity?

Invariant Measure

Definition 1.4 (Invariant measure). Let $(X_t)_{t \in \mathbb{T}}$ be a Markov process. The probability measure μ is invariant w.r.t. $(X_t)_{t \in \mathbb{T}}$ if for all $t \in \mathbb{T}$ and $A \subset \mathbb{X}$ we have stat. distr., stat. measure

$$\int_{\mathbb{X}} p(t, x, A) \mu(dx) = \mu(A). \quad (1.3)$$

In other words, a measure is invariant wrt a Markov process if the probability to be in any subset of the state space is the same as the probability to get into that subset by the evolution of the Markov process for any fixed transition time.

Ergodicity

The long-time behaviour of stochastic processes can be described using ergodicity. difference μ , \mathbb{P} ?

Definition 1.5 (ergodic process). Let $(X_t)_{t \in \mathbb{T}}$ be a Markov process with invariant probability measure μ . Then $(X_t)_{t \in \mathbb{T}}$ is ergodic w.r.t. μ if for all functions $u : \mathbb{X} \rightarrow \mathbb{R}$ with $\int_{\mathbb{X}} |u| \mu(dx) < \infty$ we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T u(X_t) dt = \int_{\mathbb{X}} u(x) \mu(dx). \quad (1.4)$$

for almost all initial values $X_0 = x_0$.

So a Markov process is ergodic if its time average (left side) is the same as its average over the probability space (right side), known in the field of thermodynamics as its ensemble average. In an ergodic process, the state of the process after a long time is nearly independent of its initial state.

Reversibility

A very nice property of Markov processes is reversibility. A process is reversible if it fulfills the detailed balance equation...; it means that they keep the same probability law even if their movement is considered backwards in time.

Definition 1.6 (reversible process). *Let $(X_t)_{t \in \mathbb{T}}$ be a Markov process with invariant probability measure μ . Then $(X_t)_{t \in \mathbb{T}}$ is reversible w.r.t. μ if*

$$\int_A p(t, x, B) \mu(dx) = \int_B p(t, x, A) \mu(dx) \quad (1.5)$$

for all $t \in \mathbb{T}$ and $A, B \subset \mathbb{X}$. If μ is unique, then X_t is simply called reversible.

1.2 Transfer Operator

With the previously defined transition function, we have a tool to describe the propagation of **distributions** of stochastic processes. Now we are going to introduce an operator that propagates **probability densities** of Markov processes.

L^r -Spaces

As probability densities are defined about their integral properties (integral = 1) (?), we need some (suitable/ convenient?) integrable spaces on which the transfer operator can act.

Definition 1.7 (L^r -Spaces).

$$L_\mu^r(\mathbb{X}) = \{u : \mathbb{X} \rightarrow \mathbb{R} \mid \int_{\mathbb{X}} |u(x)|^r \mu(dx) < \infty\} \quad (1.6)$$

for $1 \leq r < \infty$ and

$$L_\mu^\infty(\mathbb{X}) = \{u : \mathbb{X} \rightarrow \mathbb{R} \mid \operatorname{ess\,sup}_{x \in \mathbb{X}} |u(x)|^r \mu(dx) < \infty\}. \quad (1.7)$$

The space L^2 is the only of the L^r -spaces which can be equipped with a canonical scalar product and thus becomes a Hilbert space (see..). For $u, v \in L^2$ we define

$$\langle u, v \rangle_\mu := \int_{\mathbb{X}} u(x)v(x)\mu(dx). \quad (1.8)$$

Now let ν_0 be the density function corresponding to the given start distribution (function). Then the density function of a subset $A \subset \mathbb{X}$ at time t is given by

$$\nu_t(A) = \int_{\mathbb{X}} \nu_0 p(t, x, A) \mu(dx). \quad (1.9)$$

On the other hand, the density ν_t is given by

$$\nu_t(A) = \int_A \nu_t(x) \mu(dx). \quad (1.10)$$

Transfer Operator and spectral properties

The two equations above yield in the following intuitive definition of a transfer operator which should "propagate" probability densities. / or just any functions/-subensembles in integrable spaces according to a given Markov process?

Definition 1.8 (Transfer Operator). *Let $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$ be the transition function of a Markov Process $(X_t)_{t \in \mathbb{T}}$ and μ be an invariant measure of $(X_t)_{t \in \mathbb{T}}$. The semigroup of propagators or (forward) transfer operators $\mathcal{P}^t : L_\mu^r(\mathbb{X}) \rightarrow L_\mu^r(\mathbb{X})$ with $t \in T$ and $1 \leq r \leq \infty$ is defined via*

$$\int_A \mathcal{P}^t \nu(y) \mu(dy) = \int_{\mathbb{X}} \nu(x) p(t, x, A) \mu(dx) \quad (1.11)$$

for measurable $A \subset \mathbb{X}$.

The transfer operator is well-defined, see [2]. We will announce here already some properties of the transfer operator which will be useful in the following chapter(s). \mathcal{P}^t is a Markov operator; i.e. it conserves the norm, $\|\mathcal{P}^t \nu\|_1 = \|\nu\|_1$, and is positive, $\mathcal{P}^t \nu \geq 0$ for $\nu \geq 0$. $\mathcal{P}^t \nu_0$ describes the transport of the function ν_0 in time t by the underlying dynamics given by the process X_t and weighted with respect to μ :

$$\nu_0 \mapsto \nu_t = \mathcal{P}^t \nu_0. \quad (1.12)$$

Since μ is invariant(?), we have that the characteristic function of the state space is invariant under the action of \mathcal{P}^t , i.e.

$$\mathcal{P}^t \mathbb{1}_{\mathbb{X}} = \mathbb{1}_{\mathbb{X}} \quad (1.13)$$

It means that \mathcal{P}^t has the eigenvalue 1 which corresponds to its eigenfunction $\mathbb{1}_{\mathbb{X}}$.

Furthermore \mathcal{P} is a bounded operator with operator norm $\|\mathcal{P}\|_2 = 1$ and $\mathcal{P} \mathbb{1}_{\mathbb{X}} = \mathbb{1}_{\mathbb{X}}$. This implies that the spectrum $\sigma(\mathcal{P})$ of \mathcal{P} is contained in the unit circle of the complex plane; i.e. we have $|\lambda| \leq 1$ for all $\lambda \in \sigma(\mathcal{P}) \subset \mathbb{C}$.

propagates subensembles? transfer operator: propagation of densities

backwards operator important? anschauung? Markov operator?

why only in L2?

Transfer operator of reversible processes

The following two theorems give us an important insight about the spectrum of the transfer operator. Since self-adjointness of the transfer operator is equivalent to reversibility of the process, we know that only reversible processes guarantees a real spectrum!

Theorem 1.9. *Let $\mathcal{P}^t : L_\mu^2(\mathbb{X}) \subset L_\mu^1(\mathbb{X}) \rightarrow L_\mu^2(\mathbb{X})$ denote the propagator corresponding to the Markov process X_t . Then \mathcal{P}^t is self-adjoint with respect to the scalar product $\langle \cdot, \cdot \rangle_\mu$ in $L_\mu^2(\mathbb{X})$, i.e.*

$$\langle u, \mathcal{P}^t v \rangle_\mu = \langle \mathcal{P}^t u, v \rangle_\mu \text{ for all } u, v \in L_\mu^2(\mathbb{X}),$$

if and only if X_t is reversible.

Proof. Huisinga[2] □

Theorem 1.10. *A self-adjoint operator has only real eigenvalues; $\sigma(\mathcal{P}) \subset \mathbb{R}$.*

Proof. □

Since the spectral radius of any transfer operator is 1, it follows for a reversible process that $\sigma(\mathcal{P}) \subset [-1, 1]$.

Later in this thesis, we are going to be very interested in examining the spectrum of the transfer operator of a given Markov process. Unfortunately we also have to consider nonreversible processes, so with a nonreal (complex) spectrum which will be a bit harder to compute with.

Infinitesimal Generator

For $\mathbb{T} = \mathbb{R}$ the Chapman-Kolmogorov property of the transition functions makes the family $\{\mathcal{P}^t\}_{t \in \mathbb{R}}$ a continuous semigroup due to

def sem. grp.

$$\mathcal{P}^{t+s} = \mathcal{P}^t \mathcal{P}^s.$$

This leads to the following definition of the the infinitesimal generator.

time-indep.?

Definition 1.11 (Infinitesimal Generator). *For the semigroup of propagators or forward transfer operators $\mathcal{P}^t : L_\mu^r(\mathbb{X}) \rightarrow L_\mu^r(\mathbb{X})$ with $t \in T$ and $1 \leq r \leq \infty$ we define $\mathcal{D}(L)$ as the set of all $\nu \in L_\mu^r(\mathbb{X})$ s.t. the strong limit*

$$L\nu = \lim_{t \rightarrow \infty} \frac{\mathcal{P}^t \nu - \nu}{t}.$$

exists. Then the operator $L : \mathcal{D}(L) \rightarrow L_\mu^r(\mathbb{X})$ is called the infinitesimal generator corresponding to the semigroup \mathcal{P}^t .

So the infinitesimal generator is an operator which describes the behaviour of the stoch. process (Markov chain) in infinitesimal time. That becomes clear by the relation $\mathcal{P}^t = \exp(tL)$ in L^2 , so we can say L "generates" the semigroup of transfer operators.

So the eigenvalues $\lambda_1, \dots, \lambda_m$ of the propagator \mathcal{P}^t are related to the eigenvalues $\Lambda_1, \dots, \Lambda_m$ of the propagator L in the following way:

$$\lambda_k = \exp(t\Lambda_k) \quad (1.14)$$

not yet discrete?

for all $1 \leq k \leq m$. Their corresponding (associated) eigenfunctions are identical. So the spectrum of the infinitesimal generator is $\Lambda_1 = 0, \dots$

Further properties of infinitesimal generator/ samband med transfer operator/ spectrum. stationary distribution of \mathcal{P}^t : solution of $Q\pi = 0$

1.3 Galerkin Projection

Until now we have a process on a continuous state space. Since we are interested in computations with our process, we are now creating a smaller (namely finite) state space which shall inherit (important) properties of our original process. This can be done by a Galerkin projection/ discretization.

p.70

need(?) to get a discrete/finite state space

At first, we need to choose an appropriate/ desired/ favoured ansatz space $\chi = \{\chi_1, \dots, \chi_n\} \subset L^2(\mu)$.

Partition of Unity

Now we are going to change from the original state space \mathbb{X} of our process to a finite (dimensional?) subspace $D \subset \mathbb{X}$. By a projection of the transfer operator onto a finite subspace we get a stochastic process defined on this subset which preserves/ has still many properties of the original process. But in general it doesn't preserve the Markov property. But the projected process is still a good approximation of the original process.

MA Jo p. 15 - 16

We project our process with a Galerkin projection on a subspace with the aid of a suitable/convenient basis of the subspace.

Definition 1.12 (Partition of Unity). *A family of measurable functions $\{\chi_1, \dots, \chi_n\} \subset L^2(\mu)$ is called a partition of unity if the following two conditions are fulfilled:*

why L2?

i) *The χ_i are non-negative and (pairwise?) linear independent*

ii) $\sum_{i=1}^n \chi_i(q) = 1$ for all $q \in \mathbb{X}$

bei jo anders; a.e.??

Definition 1.13 (Galerkin Projection). *Let $\{\chi_1, \dots, \chi_n\}$ be a partition of unity and $D = \text{span}\{\chi_1, \dots, \chi_n\}$ the associated finite-dimensional ansatz space. Then the Galerkin projection onto D is defined by $Q : L^2(\mu) \rightarrow D$ by*

$$Q\nu = \sum_{k,j=1}^n (\langle \chi_k, \chi_j \rangle_\mu)^{-1} \langle \chi_k, \nu \rangle_\mu \chi_j. \quad (1.15)$$

Theorem 1.14. *Let $\{\chi_1, \dots, \chi_n\}$ be a partition of unity and $D = \text{span}\{\chi_1, \dots, \chi_n\}$ the associated finite-dimensional ansatz space. Then the Galerkin projection onto D is defined by $Q : L^2(\mu) \rightarrow D$ by*

$$P_{kj} = \frac{\langle \chi_k, \mathcal{P}\chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \quad (1.16)$$

Example: if we choose a partition as the basis of the subspace on which we want to project. Then we get a Markov chain whose states are just the partition sets. Such a Markov chain keeps/ maintains some important properties of the process on \mathbb{X} ; the stationary distribution of the so defined Markov chain P_Q is just the projection of the invariant measure μ onto D .

know this
only after
definition
of galerkin
proj.

There are obviously many possibilities to make a Galerkin discretization/ projection because we can define it on any arbitrary partition of unity χ_1, \dots, χ_n . In chapter 2 we are going to show how the number of clusters (= metastable sets) can be determined using the spectrum of the transfer operator (eigenvalues). Furthermore we will show how to construct functions χ_1, \dots, χ_n (=membership functions) that give us a decomposition which is as good as possible (linear combination of eigenfunctions).

1.4 Recrossing Effect

Loss of Markov Property

By projection onto a (finite?) subspace we lose an important property of the transfer operator bzw generator; namely its Markov property. Hence, in general we have

nicht immer?

$$(P_Q)^k \neq (P^k)_Q.$$

Dieses Beispiel, in welchem das überschreiten der Barriere für den nächsten Zeitraum eine höhere Wahrscheinlichkeit einer erneuten Überschreitung der Barriere bedeutet, wird auch als Recrossing Problem bezeichnet und ist in...

-¿ spoling Markov property???

An important question when it comes to projections of stochastic processes onto lower-dimensional state spaces is shown in the following diagram. Does it make a difference if we first project the process and then propagate it and vice versa?

$$\begin{array}{ccc}
 P(\tau) & \xrightarrow{\tau \rightarrow \tau k} & (P(\tau))^k \\
 \downarrow \text{projection} & & \downarrow \text{projection} \\
 P_C(\tau) & \xrightarrow{\tau \rightarrow \tau k} & (P_C(\tau))^k
 \end{array}$$

Weber shows in habilitation that under a certain Galerkin Projection using membership functions (not set-based family) leads to a commuting diagram of projection and propagation. + Markov Property is preserved?

non-
reversible?

Recrossing Effect

If we consider a (continuous) Markov process (X_t) and project it onto a finite-dimensional state space via a Galerkin projection, then the process can lose his Markov property; that phenomena is called the recrossing effect.

It depends on the chosen partition of unity (the new finite-dim. state space) if the recrossing effect occurs. If we chose a full partition for that ansatz space, then our projected process will maintain its Markovianity.

2 Metastability and Dominant Structures

What we have done so far in chapter 2.4, was to state the fact that there exist classes of stochastic processes that are either too complex and/or need such long-time simulations s.t. a dimension reduction (e.g. Galerkin projection) is required.

So far, everything for reversible processes

In this chapter, we are going to describe metastability which is a certain/characteristic behaviour of stochastic processes (Markov chains?) that often occurs in molecular dynamic systems (also other applications like ... , see ...). We will see why it makes sense to choose these metastable sets (or functions) as clustering sets in the aforementioned Galerkin Projection. We will also see how these metastable sets (conformations) can be detected using the spectrum (i.e. eigenvalues and eigenfunctions) and which metastable decomposition is optimal (no unique solution) (PCCA+).

Unfortunately, that can only be done for reversible processes. In the non-reversible case, everything is a bit more difficult and we need to detect dominant cycles instead of dominant sets. We will give a short outlook how that case could be handled.

2.1 Concept of metastability applied to MD

There exist several different definitions of metastability in the literature. Shortly said, metastability is the property of a Markov process which consists of (a decomposition into?) subsets s.t. transitions between these subsets are rare events while the duration of stay inside of each these subsets is long. Some possible characterizations of that behaviour are based on large hitting times (see...) or small exit rates (...).

Mathematical concept of metastability

To describe the concept of metastability, it is a good way to start with so called *stable* or *invariant subsets*. A subset of the state space of a Markov process is called stable or invariant if it cannot be left, i.e. if $\mathbb{P}(X_t \in A \mid X_0 \in A) = 1$ for all t . Analogously, we can define a *metastable* or *almost invariant subset* as a subset in which the process will stay for a very long time before exiting it into any other

subset; i.e. $\mathbb{P}(X_{t_f} \in A \mid X_0 \in A) \approx 1$ for a convenient timescale t_f .

Obviously, being "close to 1" is a rather vague statement. But that lack of concreteness will be eliminated later, since we will only be interested in the "best" metastable decomposition, so for each metastable set we try to get that probability as close to 1 as possible. Instead, we will have to determine the number of subsets we are looking for.

Let A_1, \dots, A_m be a full partition of the state space \mathbb{X} . Then this partition and each of its partition sets are called *metastable* if

$$\sum_{k=1}^m \mathbb{P}_\mu(X_{t_f} \in A_k \mid X_0 \in A_k) \approx m.$$

Then each of the sets A_k is almost invariant with respect to timescale t_f ; the probability to stay in one of the partition sets being started there is almost 1, while the probability to change between any two different partition sets is almost 0.

Importance of metastability to MD systems

Metastability is a concept which is very important for stochastic processes describing molecular dynamic systems. Such systems/processes often have the characteristic behaviour that their trajectory stays inside of a certain conformation/region/subset for a very long time before switching/changing/exiting to another region. Furthermore, transitions between these conformations are rare. So these regions correspond to our above definition of metastable sets if we choose a convenient timescale.

Problem of long-time simulations of MD systems

As transitions between metastable sets are a rather rare event, we need to make long-time simulations of our process in order to get reasonable results about these changes of conformations. But, as mentioned in Chapter 2, long-time simulations of such complex systems are not feasible even with the best computers nowadays, see (...).

So, in order to be able to compute some long-time simulations of a given MD system, a reduction of complexity like described in section 2.4 is needed/ required. This can be achieved by a clustering of a state space. Different states of the state space will be clustered appropriately s.t. we get a process on a smaller state space.

Clustering into metastable sets

So far, we didn't mention how we should choose the clusters for the Galerkin projection, i.e. which states of the original states should be grouped together for the reduced/projected model.

Galerkin =
clustering???

As we are mainly interested (for the moment) in the long-time behaviour of our process, we are not interested in short-time transitions, i.e. transitions inside of a conformation/ metastable set. So knowing now the concept of metastability, it seems directly/intuitively appropriate/reasonable to group/cluster states of a metastable together and create a new process where each state corresponds to one of the metastable sets.

Since we want to maintain/keep the long-time-behaviour/properties of the original space/system/process, the transition probabilities of the clustered/reduced process/model should correspond to the transition rates of the original process between its metastable sets. In doing to, we get a new process which possesses the long-time-behaviour of the original process, but "forgets" its short-time-behaviour, i.e. transitions inside of a conformation.

Advantages / Disadvantages

First of all, our newly created process will have the desired property of a reduced dimension/complexity since the model acts on a (much) smaller state space while maintaining its crucial property (transitions between metastable sets = long-time behaviours). So the computation effort for (long-time) simulations is definitely decreased. Furthermore, we get a better overview of our process, since it is always easier to consider a process on a few states in comparison to a process on a very large or even continuous state space. Since fast/ short-time transitions (transitions inside conformation/ metastable set) are not our research goal, we just omit this (at least for our case!) superfluous informations. So our advantages in clustering are

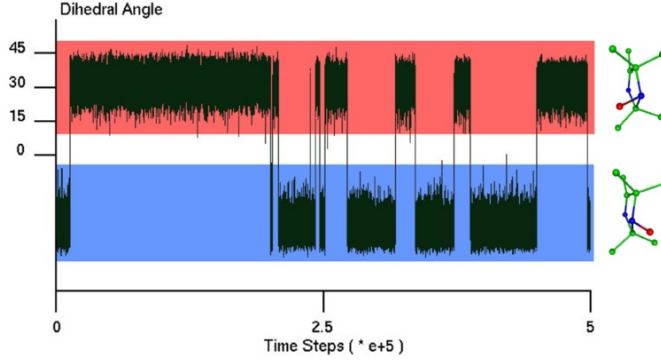
- i) reduction of dimension (main reason b/c otherwise computation are just not possible)
- ii) better overview of the crucial behaviour of the process (rare transitions)

But there is also a disadvantage, as we mentioned in 2.4, by projecting a process it can lose its Markov property.

Example

To explain metastability in Molecular Dynamics, consider the following easy example, figure Our system consists of ..? molecule and can take ...? values. As we can see from figure ..., showing a long-time-simulation of that process,

under cer-
tain/ which
conditions
not?
Disadv:
no good
metastabil-
ity -¿ no
good model?



2.2 Spectral Approach

In this section we will see that the spectrum of the transfer operator is highly connected to the metastability of the corresponding Markov process. Namely, the number of metastable sets can be determined by the number of eigenvalues close to 1. And the corresponding eigenfunctions allow a metastable decomposition.

partition?

Existence of dominant eigenvalues

For further informations, see Kato[4]. We are interested in large eigenvalues which are close to 1 and separated from the rest of the spectrum. The *discrete spectrum* $\sigma_{\text{discr}}(\mathcal{P})$ is the set consisting of all eigenvalues $\lambda \in \sigma(\mathcal{P})$ that are isolated and of finite multiplicity. The *essential spectral radius* $r_{\text{ess}}(\mathcal{P})$ is defined as follows

$$r_{\text{ess}}(\mathcal{P}) = \inf\{r \geq 0 \mid \lambda \in \sigma(\mathcal{P}) \text{ with } |\lambda| > r \text{ implies } \lambda \in \sigma_{\text{discr}}(\mathcal{P})\}.$$

The existence of dominant eigenvalues requires that the essential/ continuous part of the spectrum is bounded away from the dominant elements of the discrete spectrum. Let us now consider the transfer operator $\mathcal{P} = \mathcal{P}^t$ for some fixed t in the Hilbert space $L^2_\mu(\mathbb{X})$.

To ensure that the process we are considering actually possesses metastable sets, we need to pose some conditions on the spectrum of the transfer operator:

L^1 Huisinga
diss; why L^2 ,
 L^1 enough?

- C1** The essential spectral radius of \mathcal{P} is less than one; i.e. $r_{\text{ess}} < 1$.
- C2** The eigenvalue $\lambda = 1$ of \mathcal{P} is simple and dominant; i.e. $\eta \in \sigma(\mathcal{P})$ with $|\eta| = 1$ implies $\eta = 1$.

We will not go into further details for which processes the two above conditions are fulfilled; some criteria about it can be found in Huisinga[2]. Since these conditions are required for the later investigations, we will just assume that they are true.

We need condition **C1** to ensure that the continuous part of the spectrum is bounded away from the discrete eigenvalues. Otherwise they would not be dominant anymore and the process would be rather durchmischt than having any metastable sets. Condition **C2** however is important because a transfer operator with more than one eigenvalue of absolute value 1 can be decomposed into stable/invariant sets, i.e. subsets which cannot be left. In that case we could just consider the different stable sets as independent processes. But that is not interesting for us. Instead, we want to know more about almost invariant sets and their critical/transition regions.

Theorem 2.1 (Schuette[7]). *The transfer operator $\mathcal{P} : L^2 \rightarrow L^2$ of a reversible process with properties **C1** and **C2** has the following spectrum:*

$$\sigma(\mathcal{P}) \subset [a, b] \cup \{\lambda_n\} \cup \dots \cup \{\lambda_2\} \cup \{1\}$$

with $-1 < a \leq b < \lambda_n \leq \dots \leq \lambda_1 = 1$.

This theorem assures us the existence of a discrete set of dominant eigenvalues. In the following we will see that this property results in metastability.

Relation of dominant eigenvalues to metastable sets

The number of metastable sets can be determined by counting the dominant eigenvalues.

Relation of dominant eigenfunctions to metastable decomposition

Theorem 2.2. *Each single eigenfunction induces a metastable decomposition*

Proof.

□

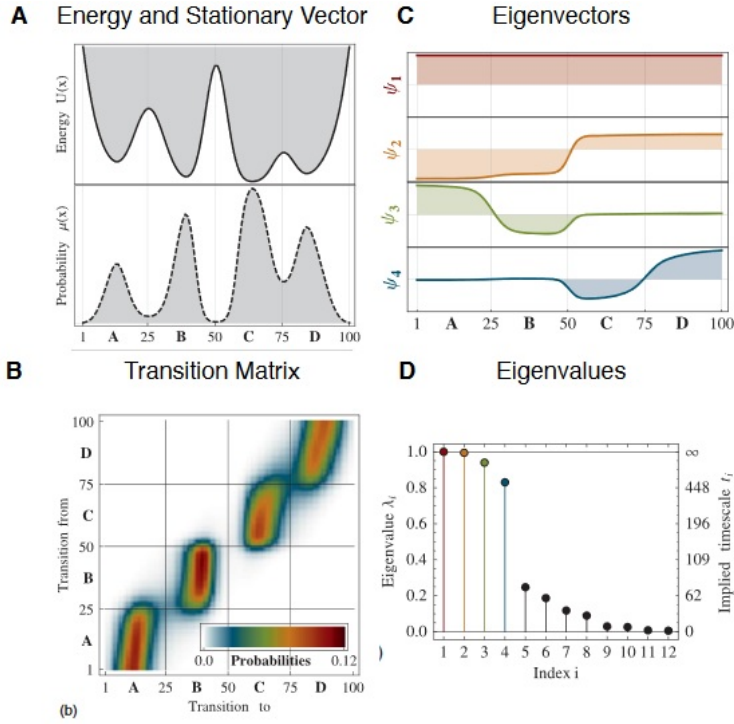
The zeros of an eigenfunction (bzw. Vorzeichenwechsel) induce a metastable decomposition of the state space.

Different eigenfunction results in a different decomposition. (partition into sets??)

has periodic structures?
excludes modeling and interpretation problems
C2 = ergodic?

why discr. rechts?

eigenfunctions vs committor functions



Disadvantages

So the spectral approach is suitable/ convenient to characterize metastability of Markov processes. But: there are two disadvantages. 1: the result is only applicable on reversible processes, because a self-adjoint transfer operator is needed in order to get real eigenvalues. 2: eigenvector(?) problem of the transfer operator has only global solutions.

2.3 Soft/ Fuzzy Clustering

The above considerations (metastable decomposition induced by zeros of eigenfunctions) result in a metastable full decomposition of the state space, i.e. each state is assigned to exactly one of the partition set. We will see now, that there exist better solutions, considering/ including the fact that transition regions can belong to several metastable sets. So for this new approach (even though already examined/ investigated/ analyzed in recent research, see ...), there may be some overlap in the assignment of states to metastable sets.

Set-based vs. Function-based Approach

An easy intuitive approach to decompose the state space would be to determine a certain number of metastable sets which form a full partition of the state space, such that each state is assigned to exactly one of the metastable sets. The problem with that approach is that also the transition regions of the process would have to be assigned to one of these partition sets. But why would you assign a state in a transition region to one adjacent metastable set and not to the/one other? So such an assignment is not a rigorous description of actual behaviour of the process.

So this *set-based approach* in decomposing the process has been replaced by the *function-based approach*. That means that each state of the process is assigned with a certain probability to each metastable set. That makes sense, because a state in a transition region can go with certain probabilities to different conformations.

Membership functions (=almost characteristic function)

Assuming we have already determined that our process consists of n metastable sets (by knowing its n dominant eigenvalues). We will follow the approach of [9] by defining macro states as *overlapping partial densities*. They can be identified by membership functions $\chi_1, \dots, \chi_n : \mathbb{X} \rightarrow [0, 1]$. Each state of the original state space shall be assigned to the different macro states with a certain *degree of membership*.

Definition 2.3. *membership function χ_i = partition of unity (see 2.4) and non-negative*

Membership functions can be used to decompose the space into metastable sets. In this case, the assignment of a state to a metastable set must not be unique, but a state can belong to different metastable sets with certain probabilities(?). That model takes into consideration the existence of transitions regions which cannot be uniquely assigned to one macro state/ metastable set.

Since membership functions form a partition of unity, we can apply the Galerkin projection as defined in chapter 1. That gives us a Markov State Model where each (macro) state is a metastable set of the original process.

Example 2.4. *A characteristic function is a membership function. It induces a full partition as metastable decomposition.*

Individual eigenfunctions \mathcal{X} do not overlap since they are orthogonal. But the membership functions χ_i as linear combinations of the eigenfunctions might have an overlap.

so far: membership fct indep. of metastable set

Perron Cluster Analysis

Perron Cluster Analysis vs Robust Perron Cluster Analysis (=PCCA+?)

So now we need to choose/determine convenient/suitable membership functions s.t. the metastable decomposition becomes as good as possible. Normed eigenfunctions are one possibility (no overlap). But better: linear combinations of eigenfunctions.

Perron Cluster Analysis has been developed by Deuffhard et al (aggregates in nearly uncoupled MC) which used the sign structure of the dominant eigenvalues of the stochastic matrix. That approach has been approved by / Deuffhard and Weber (oder Weber und Galliat?) who transformed the system of eigenvectors into a system of membership functions which results in a soft/ fuzzy clustering?

The required membership functions can be built as a linear combination of the (dominant) eigenfunctions (convex problem?). multiplied with a matrix A which is the result of Perron Cluster Cluster Analysis (PCCA+).

We are considering the set of dominant eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ with the corresponding set of eigenfunctions $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$. They fulfill the eigenvalue problem $\mathcal{P}(\tau)\mathcal{X} = \mathcal{X}\Lambda$ of the transfer operator $\mathcal{P}(\tau)$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$.

Let \mathcal{X} be the eigenvector matrix, i.e. the i -th column of \mathcal{X} is an eigenvector corresponding to the eigenvalue λ_i . Weber showed in [10] that the choice $\chi = \mathcal{X}A$ (linear combination of the eigenfunctions \mathcal{X}) preserves the Markov property, i.e.

$$\chi_j(q) = \sum_{i=1}^n A_{ij} \mathcal{X}_i(q), \quad j = 1, \dots, n,$$

where $A = \{A_{ij}\}_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$ is the solution of a convex maximization problem (PCCA+), see Weber[9]. In order to fulfill the partition of unity of each λ_i , the matrix A has to be chosen s.t. A is row-stochastic.

Optimal decomposition (PCCA+)

Since there are so many possibilities of finding membership functions (all convex combinations of eigenfunctions), we want to find the best (or a good) solution which gives us a good decomposition into metastable sets.

PCCA+

Finding membership functions which result in optimal solution for metastable decomposition (but PCCA+ works only on finite state spaces(?)).

There are infinitely many transformations A of the eigenvectors resulting in a soft membership matrix χ satisfying the positivity and partition of unity constraints.

only for finite matrices? or also cont. operator w/ eigenfct.? set-based approach?

why not eigenfct.? only dom. spec. jede konvexe lin.komb.? bzw even: discretization error vanishes

Consequently, we have to determine the transformation A that satisfies some optimality condition.

Do we need PCCA+ to find the best(?) membership functions. And with the help of these membership functions we can apply a Galerkin Projection?

2.4 Dominant cycles (non reversible NESS process)

Approaches for nonreversible processes

What is the dominant structure of a nonreversible process? cycle? What is egentli-
gen the problem with nonreversible processes? non-real eigenvalues

How can the computation of dominant cycles help to detect metastable sets/
metastables cycles?

What is a metastable cycle?

Definition dominant cycle

Schur Decomposition

Dominant structures will be defined utilizing the dominant Schur vectors of the transition matrix instead of its eigenvectors.

A membership matrix can be defined as a linear combination of these Schur vectors.

3 Rebinding Effect in a Given Kinetics

In which model are we working??? Hamiltonian (Gamma)? Hamiltonian w/ randomized momentum (Omega w/ any momentum)? Langevin? Diffusion Dynamics? ..

In this chapter we are going to examine a special type of a molecular (kinetics?) system, namely receptor-ligand systems. To describe these systems we can use all the previously defined mathematical objects.

To give a short overview about what is going to happen here. A molecular system can be described via a differential equation. The solution of this differential equation is a (Markov?) process which can be described via a transfer operator (chapter 2). This operator will be projected onto a finite-dimensional state space (Galerkin projection, chapter 2) which (maybe?) spoils the Markov Property of the process.

Here we will try to tackle this subject for nonreversible (NESS) processes also.

This chapter is mathematically based on Weber (Quantifying Rebinding Effect)..

3.1 Receptor-Ligand System

Ligand (L) can bind to a receptor (R) and form a receptor-ligand complex (LR)

together with reb. eff.

What is a molecular system? State space, Phase Space, Ensemble, Configurational Space, Conformation Space, reduced density

3.2 Rebinding Effect

The rebinding effect has been characterized as a memory effect which leads to an additional thermodynamic weight of the bound state.

In fact, a stochastic process describing a receptor-ligand molecular system IS NOT necesarrily Markovian. The Markovianity can be spoiled by the Rebinding

impact of multivalency on the rebinding effect (Weber, Chem.)

Effect. If a Receptor-Ligand system dissolves, due to the favorable spatial situation (?) it is more likely to rebind again than to stay dissolved.

Zusammenhang
zum Absatz
davor?

There are several papers (...) describing the rebinding effect on a chemical and a mathematical view. In chemistry, there are several reasons/ factors for the rebinding effect discussed.

An important application of receptor-ligand processes is drug design. In short: A drug consists of ligands which should bind to the receptors of the virus. If the drug creates many bindings, the virus is "bound" and cannot attack the human (cell?) anymore. So many bindings are a favorable thing. So a high rebinding effect enhances the (overall?) binding affinity of the process/ system which is good for the efficiency of a drug. We want a high rebinding effect. So in this chapter, we examine the minimal rebinding effect for a given Kinetics. This task has been solved by Weber... for reversible processes.

How: bound -> unbound

3.3 Molecular Kinetics as a Projection

Mathematical
description
of molecular
kinetics

In general, to describe molecular systems we can distinguish between two point of views: considering micro states or macro states of the system. Here, we will basically embed the theoretical concepts of chapter 1 and 2 in the physical/ chemical framework (phase space etc.).

Micro States

A micro state of a molecular system with N atoms can be represented in a $6N$ -dimensional phase space $\Gamma = \Omega \times \mathbb{R}^{3N}$, consisting of the configurational space $\Omega = \mathbb{R}^{3N}$ and the momentum space \mathbb{R}^{3N} .

Boltzmann distribution

$$\bar{\rho}(q, p) \propto \exp(-\beta H(q, p)), \quad (3.1)$$

where $\beta = 1/(k_B T)$ is the inverse of the temperature T multiplied with the Boltzmann constant k_B . The Hamilton function denoted by H is given by $H(q, p) = K(p) + V(q)$, so by the sum of the kinetic energy $K(p)$ and the potential energy $V(q)$.

Macro States

Group/cluster a collection of the micro states having the same or similar values in one observable (e.g. bound, unbound state of receptor-ligand system). By grouping micro states, the (corresponding?) macro states yield **TODO: statistical weights and entropic information (?)**. Macro states need not be distinct sets. Approach here: overlapping partial densities, which can be identified as membership functions χ_1, \dots, χ_n .

General Approach here

In this section, we are going to describe the mathematical model which we will use to describe a molecular system and its propagation/ evolution/ time changes? As it is a stochastic process (on molecular level small arbitrary changes that cannot be described deterministically), we will get a transfer operator which describes the propagation of EACH state (=micro state) of the process. I.e. we have a state space which is enormous large.

Since we want to make computations on that process, we need to make it smaller. This can be done using the Galerkin Projection as describes in chapter 2. So we get a process on a finite state space which should maintain the most important properties of the continuous process. (Markovianity?)

This discretization can be achieved by using membership functions (linear combinations of eigenfunctions of the transfer operator corresponding to the dominant eigenvalues). These membership functions can be computed using PCCA+. BUT: PCCA+ works only for reversible processes.

So we will at first handle only the reversible case, since it is much easier to deal with.

At the end, we get our (finite) transition matrix and/or the infinitesimal generator of the process. This important tool will help us to maximize/minimize our optimization problem at the end.

Membership functions

There exists different approaches for ... set-based vs function based. Here we use membership functions which can have certain overlap.

Membership functions $\chi_1, \dots, \chi_n : \Omega \rightarrow [0, 1]$ which form a partition of unity, i.e.

$$\sum_{i=1}^n \chi_i(q) = 1. \quad (3.2)$$

define macro states as overlapping partial densities, which can be identified as membership functions.

here overlap is good?
membership functions, overlapping partial densities, partition of unity

Statistical Weights

We can assign a statistical weight to each macro state (= membership function χ_i):

$$w_i = \int_{\Omega} \chi_i(q) \pi_q(q) dq \quad (3.3)$$

and entropic information?
what is that good for?

Statistical weight w_i = probability to be in conformation/ (metastable) macro state i

Transfer Operator

For reversible(!) processes Weber [10] defines a transfer operator which is self-adjoint! due to the detailed balance condition.

trivial?

In his thesis, Weber defined the transfer operator as follows:

$$\mathcal{P}(\tau)f(q) = \int_{\mathbb{R}^{3N}} \left(\int_{\Omega} f(\tilde{q}) \Psi^{-r}(\tilde{q} | (q, p)) d\tilde{q} \right) \pi_p(p) dp. \quad (3.4)$$

So for the nonreversible approach we cannot use this operator? Or: we can use this operator but it is not self-adjoint anymore? (so complex eigenvalues)..

We assume that the discrete spectrum of the transfer operator has n dominant eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ which are all close to 1 and bounded away from the essential spectrum (see Schuette?). The corresponding eigenfunctions are denoted by $\mathcal{X} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$

?

Relation between Transfer Operator and Markov Operator

A transfer operator $\mathcal{P}(\tau)$ also defines a (corresponding) projected Markov operator $\bar{\mathcal{P}}(\tau)$ acting in configurational space Ω , see [10];

$$\bar{\mathcal{P}}(\tau) = \pi_q \circ \mathcal{P}(\tau) \circ (\pi_q)^{-1}.$$

Discrete Markov State Model

Now we want to create a discrete Markov State Model of our process. We want the states of this discrete model to correspond to the metastable conformations/sets of the process.

Since each of the n dominant eigenvalues of the transfer operator corresponds to a metastable set (no? because membership fct instead of set?)...

Using the dominant spectrum of the transfer operator, we want to create a discrete Markov State Model on n states. The state space of this Model should

consist of the macro states of our Molecular System and its transition behaviour should be described via a $n \times n$ -transition matrix $P(\tau)$ (i.e. row-stochastic matrix).

So somehow we need to get from our continuous operator $\mathcal{P}(\tau)$ to a discrete matrix $P(\tau)$, while we want to preserve the most important properties of the process. (Markovianity?)

For doing this, we have at first to determine the size and shape of the membership function χ_i .

$$\chi_j(q) = \sum_{i=1}^N A_{ij} \chi_i(q), \quad j = 1, \dots, n, \quad (3.5)$$

where $A = \{A_{ij}\}_{i,j=1,\dots,n}$ is the solution of PCCA+ (convex maximization problem). This choice of membership functions preserves Markovianity of the process when projecting. As a linear combination of eigenfunctions, the membership functions might have an overlap!

PCCA+ can be applied only for clustering processes on finite/discrete state spaces
Ref?

Galerkin Projection

The Galerkin Projection helps us to reduce our continuous stochastic process to a discrete process by a projection.

$$P(\tau) = (\langle \chi, \chi \rangle_\pi)^{-1} (\langle \chi, \mathcal{P}(\tau) \chi \rangle_\pi) \quad (3.6)$$

Normally the Markov Property is spoiled by a projection on a finite-dimensional space, but with this Galerkin Projection Markovianity is preserved....

Infinitesimal Generator

Using the transfer operator $\mathcal{P}(\tau)$ we can define a time-independent operator \mathcal{Q} via

$$\mathcal{Q} = \lim_{\tau \rightarrow 0} \frac{\mathcal{P}(\tau) - \mathcal{I}}{\tau}. \quad (3.7)$$

,... then \mathcal{Q} is the infinitesimal generator of \mathcal{P} :

$$\mathcal{P}(\tau) = \exp(\tau \mathcal{Q}). \quad (3.8)$$

Weber shows in [10] that such an infinitesimal generator exists for a discretization in terms of membership functions.

Since the eigenfunctions of the infinitesimal generator are the same as for the transfer operator and their eigenvalues are in the relation $\exp(\chi_i) = \lambda_i$, we can now apply the same Galerkin Projection for the infinitesimal generator as for the transfer operator in (3.6). We get a $n \times n$ -rate matrix

$$Q_c = \dots = (\langle \chi, \chi \rangle_\pi)^{-1} (\langle \chi, \mathcal{Q} \chi \rangle_\pi). \quad (3.9)$$

3.4 Minimizing the Rebinding Effect

For reversible processes, this problem is solved by Weber and Fackeldey [12] with the spectral approach.

With the tools from chapter 2 (Schur Decomposition) we give an approach how this problem can be solved for nonreversible processes (NESS processes) using Schur Decomposition to get rid of the possibly nonreal eigenvalues, see Djurdevac et al[1](2016).

3.5 Approach for nonreversible processes

detecting dominant cycles of the process?

Galerkin Projection

with the help of a Schur decomposition we can make the eigenvalues real???

4 Illustrative Examples

4.1 An artificial (bivalent) binding Process

One can distinguish between a monovalent binding process and a multivalent binding process (see ...), where multivalent processes are often considered as having a better binding affinity (see..).

For the monovalent case, the mathematical modeling of its kinetics is well understood.

Whenever the receptor molecules are spatially preorganized, the corresponding binding process is denoted as multivalent.

(especialle bivalent or polyvalent case often observed in nature) These systems are of significant interest for pharmaceutical and technical applications. If the ligands are linked to each other in an appropriate way to match the preorganized receptor molecules and, thus, are also presented multivalently, then extremely high binding affinities are often observed.

So we consider here a bivalent process, as the the easiest multivalent case.

Bibliography

- [1] N. DJURDJEVAC CONRAD, M. WEBER, AND C. SCHÜTTE, *Finding dominant structures of nonreversible markov processes*, Multiscale Modeling & Simulation, 14 (2016), pp. 1319–1340.
- [2] W. HUISINGA, *Metastability of markovian systems*, Ph.D. thesis, Freie Universität Berlin, (2001).
- [3] S. L. KALPAZIDOU, *Cycle representations of Markov processes*, vol. 28, Springer Science & Business Media, 2007.
- [4] T. KATO, *Perturbation Theory for Linear Operators*, Classics in Mathematics. Springer, 1995.
- [5] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series. Springer, 1993.
- [6] A. NIELSEN, *Computation schemes for transfer operators*, Ph.D. thesis, Freie Universität Berlin, (2015).
- [7] C. SCHÜTTE AND M. SARICH, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, vol. 24 of Courant Lecture Notes, American Mathematical Soc., 2013.
- [8] G. VAUQUELIN, *Rebinding: or why drugs may act longer in vivo than expected from their in vitro target residence time*, Expert Opinion on Drug Discovery, 5 (2010), pp. 927–941.
- [9] M. WEBER, *Meshless methods in conformation dynamics*, Ph.D. thesis, Freie Universität Berlin, (2006).
- [10] M. WEBER, *A subspace approach to molecular markov state models via a new infinitesimal generator*, Habilitation thesis, Freie Universität Berlin, (2011).
- [11] M. WEBER, A. BUJOTZEK, AND R. HAAG, *Quantifying the rebinding effect in multivalent chemical ligand-receptor systems*, The Journal of Chemical Physics, 137 (2012), 054111.

- [12] M. WEBER AND K. FACKELDEY, *Computing the minimal rebinding effect included in a given kinetics*, Multiscale Modeling & Simulation, 12 (2014), pp. 318–334.