

Master thesis

Computing the minimal rebinding effect for nonreversible processes

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

Susanne Röhl

Berlin, den 1. Februar 2017

Contents

1	Introduction	3
2	Markov State Models	6
2.1	Markov Process	6
2.2	Transfer Operator	9
2.3	Galerkin Projection	11
2.4	Recrossing Effect	13
3	Metastability and Dominant Structures	14
3.1	Concept of metastability	15
3.2	Spectral Approach	16
3.3	Set-based vs Function-based Approach	16
3.4	Optimal decomposition (PCCA+)	17
3.5	Dominant cycles (non reversible NESS process)	18
4	Rebinding Effect in a Given Kinetics	19
4.1	Receptor-Ligand System	19
4.2	Rebinding Effect	19
4.3	Molecular Kinetics as a Projection	20
4.4	Minimizing the Rebinding Effect	23
4.5	Approach for nonreversible processes	23
5	Illustrative Examples	25
5.1	An artificial (bivalent) binding Process	25
	Bibliography	i

1 Introduction

understand the long-term behavior of dynamical systems

We can generalize the notion of conformations for arbitrary dynamical systems. Suppose, that there are two or more “macroscopic states”, i.e. subsets in phase space in which trajectories tend to stay long before switching to another. These sets are called almost invariant [Del99]. They are a curse for methods which try to extract long-term statistical properties from trajectories of finite length, since they can “trap” orbits for a long time, and regions in the phase space may stay unvisited, if the length of the simulation is not sufficient. Since they govern the dynamics of a system over a large timescale, one is also interested in finding these almost invariant sets, and to quantize “how much almost invariant” they are.

Drug Design

A drug becomes good if it has a high binding affinity to the virus/.. . The binding affinity of the drug can be affected by different aspects. Like - shape of the drug molecule (favorable spatial preorganisation of the ligands leads to an entropy loss at ligand binding) - thermodynamical reasons - rebinding effect

Rebinding Effect in Drug Design

drug design: minimize rebinding effect to make drugs better

Minimizing the Rebinding Effect

In [11] Weber solved the problem of minimizing the rebinding effect for reversible stochastic processes. We extend his approach in generalizing it to the case of nonreversible processes.

In that context the main difference between reversible and nonreversible processes is their spectrum. While reversible processes possess only real eigenvalues, the eigenvalues of nonreversible processes can be complex. So a spectral decomposition is not immediately possible for nonreversible processes. Instead we use the Schur decomposition as it has been shown in [1] how to find dominant structures of nonreversible processes.

Markov State Models

Thermodynamical Ensemble/ Boltzmann Distribution, (Potential) Energy Landscape: High Potential Energy = transition region, Low Potential Energy = (stable/ metastable/ invariant) region

Clustering in metastable conformations - Mixture Model Clustering - Hidden Markov Model - PCCA+ = Clustering Algorithm using Membership Functions

Deuffhard/ Schuette: set-based clustering approach Weber (PCCA+): function-based clustering (meshless function), (continuous and overlapping) membership functions, partial density function/ total density function, complexity reduction

Approach

Most of all this thesis is about the mathematical concepts which are used to describe the Rebinding Effect. The results will be embedded in the chemical context of how this rebinding effect affects drug design / how drugs can be made more efficient exploiting this effect / by maximizing this effect. For more detailed chemical backgrounds we will refer to convenient/ suitable papers from the field of chemistry.

General approach: Have a problem in a applied field, namely chemistry, describing molecular dynamics, enhancing drugs.. This problem is abstracted to a purely mathematical/ theoretical level, as seen in this thesis, using stochastic processes/ transfer operators ... These mathematical results can then be used to solve the original problem.

Literature

Basis theory of Metastability and Markov State Models is from Schuette (MSM book), including notation, definitions, transfer operator, dominant spectrum, Galerkin projection...

Structure

First: What is a Markov Process, what is a MSM (Markov State Model), Why is the (dominant) spectrum of importance, What is Metastability, including important definitions (transfer operator, Galerkin Projection,...)

Second: Metastability, the importance of the dominant spectrum, the difference of reversible and irreversible MC (dominant metastable sets vs dominant cycles), Schur Decomposition, Schwerpunkt on irreversible processes, Schwierigkeiten von irreversible processes

Third: Application of the previous chapters to MD (Molecular Dynamics): Rebinding Effect, What is the Rebinding Effect, Molecular Kinetics as a Projection, Minimize the Rebinding Effect

Fourth: An Artificial Example: Computing the Minimal Rebinding Effect, Some Pictures!

Eine Randnotiz

2 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..

2.1 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 a continuous state space and continuous 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 :

$$A \in \Sigma \Rightarrow X^{-1}(A) \in \mathcal{A}. \quad (2.1)$$

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 2.1 (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 2.2 (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) p fulfills the Chapman-Kolmogorov equation, namely for all $t, s \in \mathbb{T}, x \in \mathbb{X}$ and $A \subset \mathbb{X}$:

?

$$p(t + s, x, A) = \int_{\mathbb{X}} p(t, x, dz) p(s, z, A). \quad (2.2)$$

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 2.3 (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*

$$p(t, x, A) = \mathbb{P}(X_{t+s} \in A \mid X_s = x). \quad (2.3)$$

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

So the time evolution of a Markov process is described by its transition function. From this definition, it is clear that such a process possesses no "memory". The change of the process after an arbitrary time step t depends only on the last known state of the process at time s and not of any previous states, as we can see on the right side of equation

So a Markov process is a process without memory; its evolution depends only on the current space and is independent of the history. We are only interested in (time-)homogeneous Markov processes, i.e. processes which...

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[4, chapter 3].

So the transition function for a Markov process plays the same role as the transition matrix for a Markov chain; propagating distributions?

Having introduced the notion of Markov Processes, we can now define some important properties and give some examples.

time homo-
geneity

Invariant Measure

Definition 2.4 (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*

$$\int_{\mathbb{X}} p(t, x, A) \mu(dx) = \mu(A). \quad (2.4)$$

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 a fixed transition time.

Stationary
distribution
for Markov
Chain -
invariant
measure
for Markov
Process?
skillnad
stationary
measure/
stat. distr.?

Ergodicity

The long-time behaviour of stochastic processes can be described using ergodicity.

Definition 2.5 (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*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T u(X_t) dt = \int_{\mathbb{X}} u(x) \mu(dx). \quad (2.5)$$

for almost all initial values $X_0 = x_0$.

general: mea-
sure μ oder
BProb?

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 2.6 (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*

$$\int_A p(t, x, B) \mu(dx) = \int_B p(t, x, A) \mu(dx) \quad (2.6)$$

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

2.2 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.

skillnad generator

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 2.7 (L^r -Spaces).

$$L_\mu^r(\mathbb{X}) = \{u : \mathbb{X} \rightarrow \mathbb{R} \mid \int_{\mathbb{X}} |u(x)|^r \mu(dx) < \infty\} \quad (2.7)$$

for $1 \leq r < \infty$ and

$$L_\mu^\infty(\mathbb{X}) = \{u : \mathbb{X} \rightarrow \mathbb{R} \mid \operatorname{ess\,sup}_{x \in \mathbb{X}} |u(x)|^r \mu(dx) < \infty\}. \quad (2.8)$$

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

$$\langle u, v \rangle_\mu := \int_{\mathbb{X}} u(x)v(x)\mu(dx). \quad (2.9)$$

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

$$\nu_t(A) = \int_{\mathbb{X}} \nu_0 p(t, x, A) \mu(dx). \quad (2.10)$$

On the other hand, the density ν_t is given by

$$\nu_t(A) = \int_A \nu_t(x) \mu(dx). \quad (2.11)$$

Transfer Operator

The two equations above yield in the following intuitive definition of a transfer operator which should "propagate" probability densities.

Definition 2.8 (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 $P^t : L_\mu^r(\mathbb{X}) \rightarrow L_\mu^r(\mathbb{X})$ with $t \in T$ and $1 \leq r \leq \infty$ is defined via*

$$\int_A P^t \nu(y) \mu(dy) = \int_{\mathbb{X}} \nu(x) p(t, x, A) \mu(dx) \quad (2.12)$$

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

We will announce here already some properties of the transfer operator which will be very useful in the following chapter(s). 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 2.9. *Let $P^t : L_\mu^2(\mathbb{X}) \subset L_\mu^1(\mathbb{X}) \rightarrow L_\mu^2(\mathbb{X})$ denote the propagator corresponding to the Markov process X_t . Then P^t is self-adjoint with respect to the scalar product $\langle \cdot, \cdot \rangle_\mu$ in $L_\mu^2(\mathbb{X})$ if and only if X_t is reversible.*

Proof.

□

Theorem 2.10. *A self-adjoint operator has only real eigenvalues; $\sigma(P) \subset \mathbb{R}$*

Furthermore $\sigma(P) \subset [-1, 1]$

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.

transfer operator: propagation of densities

backwards operator important? anschauung?

self-adj = real spectrum see...

essential spectral radius

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

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

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

time-indep.?

Infinitesimal Generator

Definition 2.11 (Infinitesimal Generator). *For the semigroup of propagators or forward transfer operators $P^t : L_\mu^r(\mathbb{X}) \rightarrow L_\mu^r(\mathbb{X})$ 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*

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

exists. Then the operator $L : \mathcal{D}(L) \rightarrow L_\mu^r(\mathbb{X})$ is called the infinitesimal generator corresponding to the semigroup 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 $P^t = \exp(tL)$ 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 P^t are related to the eigenvalues $\Lambda_1, \dots, \Lambda_m$ of the propagator L in the following way:

not yet discrete?

$$\lambda_k = \exp(t\Lambda_k) \tag{2.13}$$

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 P^t : solution of $Q\pi = 0$

2.3 Galerkin Projection

p.70

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.

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

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

Partition of Unity

MA Jo p. 15

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.

- 16

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

Definition 2.12 (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 2.13 (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*

$$Q\nu = \sum_{k,j=1}^n (\langle \chi_k, \chi_j \rangle_\mu)^{-1} \langle \chi_k, \nu \rangle_\mu \chi_j. \quad (2.14)$$

Theorem 2.14. *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*

$$P_{kj} = \frac{\langle \chi_k, \mathcal{P}\chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \quad (2.15)$$

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 3 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).

2.4 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...

-i 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.

3 Metastability and Dominant Structures

In the last chapter, we explained why and how we can create a reduced model of a stochastic process. In this chapter, we are going to describe how we can determine the sets after which we cluster the process (=conformations) and why these conformations (=metastable sets) play such an important role.

So far, everything for reversible processes

In this chapter, we are going to the important concept of metastability and its relation to the spectrum of the transfer operator. There is a large difference in describing metastable structures in reversible and nonreversible processes; so we are going to distinguish between these two cases.

Why is metastability important for MD?

Metastability is a concept which is very important for molecular dynamics. There are many stochastic processes describing molecular systems in which the system stays for a really long time in a certain subset. Then after a long time, it goes into a different subset where again it stays a long time before leaving this subset again. Such a partition of a state space into subsets with long Aufenthaltsdauer bzw small transition rates/probabilities is called is called a metastable decomposition of the space. The jeweiligen subsets are called metastable of conformations (see ..).

example

Knowing these metastable sets of a process, it is a common task to cluster the process wrt to its metastable sets. That means that a certain area (subset of state space) is clustered/ considered as one "macro" state. By such a discretization, we get a process which is much easier to handle/ to compute with. The reduction of dimension (deletion of unnecessary information) gives us a much smaller state space than before. Going from micro states to macro states. Of course, by a reduction of dimension we lose some of the information we had before. But there exist good ways to find such a metastable decomposition which maintains most of the important properties of the process. But One disadvantage: can lose the Markov property..

In this chapter, we will give some different definitions of the concept of metastability, before showing its relation to the spectrum of the transfer operator of the process. We will show different approaches (set vs function- based) to compute such metastable decompositions + PCCA+ which computes an optimal solutions. Afterwards we will compare the case of reversible and nonreversible (NESS) process. We will see that reversible processes are much easier to handle with and give an outlook how the nonreversible case could be tackled. (in that case, metastability is given by cycles instead of sets)

3.1 Concept of metastability

There exist several different definitions of metastability. All of them give a good idea about the concept, so we will present some of them to make clear why metastability is such an important concept, especially in molecular dynamics.

Some views on metastability and relation to MD

Metastability: A subset is not invariant but almost invariant

Large hitting times

Small exit rates

Metastable full partition

Importance of metastability for MD

fore many applications, not every value is interesting but just conformations/ metastable clusters/sets are interesting.

For example. In many applications, we have a process on an extremely large state space. But it happens that there exists subsets of the state space with long Verweildauer inside these subsets and rare transitions between these subsets. Such subsets are called metastable sets.

So: computations on a large state space are very hard + not every state may be interesting, but rather a collection of states.

So: we want to cluster our stochastic process into such metastable sets. But such a clustering is not a eindeutig thing. For example: in which metastable set should we assign a transition region (e.g. a region which is close to several metastable sets).

Beispiel

loss of information by clustering?
but try to maintain important properties?

3.2 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

Theorem 3.1 (Schuette?). *The transfer operator $\mathcal{P} : L^2 \rightarrow L^2$ of a reversible process with properties 1 and 2 has the following spectrum:*

maybe in chapter 2?

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

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

A transfer operator has spectral radius 1, which means that all eigenvalues have $\lambda_i \leq 1$. We will consider only ergodic process which means that the eigenvalue 1 is unique. If the transfer operator is self-adjoint, then all eigenvalues are real; otherwise they lie in the complex plane and are possibly nonreal.

Relation of dominant eigenvalues to metastable sets

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.

Relation of dominant eigenfunctions to metastable decomposition

Theorem 3.2. *Each single eigenfunction induces a metastable decomposition*

Proof.

□

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

eigenfunctions
vs committor
functions

3.3 Set-based vs Function-based Approach

Set-based Approach

An intuitive approach to decompose the state space would be to determine a certain number of sets which form a partition of the state space, where each partition

corresponds to one conformation/ metastable set. (each state assigned to exactly one of the metastable sets).

The problem with that approach is that also the transition regions of the process 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 does not make so much sense.

Function-based Approach

So this set-based approach in decompose 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 one metastable set. That makes sense, because a state in a transition region can go with certain probabilities to different conformations.

Membership functions

using membership functions (linear combinations of eigenfunctions).

These membership functions have to fulfill certain properties, like positivity and partition of unity.

As membership functions are a partition of unity, we can 1) create a metastable decomposition (maybe with overlap, so no partition) and 2) apply Galerkin Projection on it (partition of unity necessary).

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

membership functions: continuous and overlapping

As we can see, these membership functions form a partition of one; i.e. having this family of functions, we can apply the Galerkin projection as defined in chapter 2. That gives us a Markov State Model where each (macro) state is a metastable set of the original process.

3.4 Optimal decomposition (PCCA+)

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

Second largest eigenvalue results in the best decomposition and so on ...

But Better: linear combination of eigenfunctions (might have an overlap) (membership functions, PCCA+). That will be described more in the next subsection 2.4 where, in aim to get a really good MSM, we try to find the best possible decomposition (?) into metastable sets (? not a partition anymore)

PCCA+

Finding optimal solution for finding conformations.

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

3.5 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.

4 Rebinding Effect in a Given Kinetics

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)..

4.1 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

4.2 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 necessarily Markovian. The Markovianity can be spoiled by the Rebinding Effect. If a Receptor-Ligand system dissolves, due to the favorable spatial situation (?) it is more likely to rebind again than to stay dissolved.

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

impact of multivalency on the rebinding effect (Weber, Chem.)
Zusammenhang zum Absatz davor?

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

4.3 Molecular Kinetics as a Projection

In general, to describe molecular systems we can distinguish between two point of views: considering micro states or macro states of the system.

Mathematical
description
of molecular
kinetics

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

$$\bar{\rho}(q, p) \propto \exp(-\beta H(q, p)), \quad (4.1)$$

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.

$$\sum_{i=1}^n \chi_i(q) = 1. \quad (4.2)$$

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

So: membership function "=" macro state?

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):

$$w_i = \int_{\Omega} \chi_i(q) \pi_q(q) dq \quad (4.3)$$

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 [9] defines a transfer operator which is self-adjoint! due to the detailed balance condition.

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

$$\mathcal{P}(\tau)f(q) = \int_{\mathbb{R}^{3N}} \left(\int_{\Omega} f(\tilde{q}) \Psi^{-r}(\tilde{q} | (q, p)) d\tilde{q} \right) \pi_p(p) dp. \quad (4.4)$$

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
transfer
operator
/ Markov
operator.
almost same

?

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 .

$$\chi_j(q) = \sum_{i=1}^N A_{ij} \mathcal{X}_i(q), \quad j = 1, \dots, n, \quad (4.5)$$

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.

$$P(\tau) = (\langle \chi, \chi \rangle_\pi)^{-1} (\langle \chi, \mathcal{P}(\tau) \chi \rangle_\pi) \quad (4.6)$$

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

$$\mathcal{Q} = \lim_{\tau \rightarrow 0} \frac{\mathcal{P}(\tau) - \mathcal{I}}{\tau}. \quad (4.7)$$

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

$$\mathcal{P}(\tau) = \exp(\tau \mathcal{Q}). \quad (4.8)$$

Weber shows in [9] 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 (4.6). We get a $n \times n$ -rate matrix

$$Q_c = \dots = (\langle \chi, \chi \rangle_\pi)^{-1} (\langle \chi, \mathcal{Q} \chi \rangle_\pi). \quad (4.9)$$

4.4 Minimizing the Rebinding Effect

For reversible processes, this problem is solved by Weber and Fackeldey [11] 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).

4.5 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???

5 Illustrative Examples

5.1 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] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series. Springer, 1993.
- [5] A. NIELSEN, *Computation schemes for transfer operators*, Ph.D. thesis, Freie Universität Berlin, (2015).
- [6] 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.
- [7] 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.
- [8] M. WEBER, *Meshless methods in conformation dynamics*, Ph.D. thesis, Freie Universität Berlin, (2006).
- [9] M. WEBER, *A subspace approach to molecular markov state models via a new infinitesimal generator*, Habilitation thesis, Freie Universität Berlin, (2011).
- [10] 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.
- [11] M. WEBER AND K. FACKELDEY, *Computing the minimal rebinding effect included in a given kinetics*, Multiscale Modeling & Simulation, 12 (2014), pp. 318–334.