Master thesis

Computing the minimal rebinding effect for nonreversible processes

Zur Erlangung des akademischen Grades
Master of Science

Susanne Röhl

Berlin, den 24. März 2017

Contents

1 Markov State Models

Stochastic processes, especially Markov processes, are used in many applications in different areas, like biotechnology or ... . Simulations of biomolecular systems (in atomic representation) often require timescales that are far beyond the capacity of computer power currently available (for detailed example see Anton). To get a 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/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 direction 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 $X : \Omega \to \mathbb{X}$ 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 $X_t : \Omega \to \mathbb{X}$ on some index set $\mathbb{T}$ is called a stochastic process on a state space $\mathbb{X}$.* einschränkung

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. $\mathbb{R}^n$ + Borel?

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

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 ?

$$p(t + s, x, A) = \int_{\mathbb{X}} p(t, x, \mathrm{d}z)p(s, z, A).$$

holds for all $t, s \in \mathbb{T}, x \in \mathbb{X}$ *and* $A \in \mathcal{B}(\mathbb{X})$. $A \subset \mathbb{X}$

In this definition, the first three properties just ensure that we get reasonable (measurable) results and that that the process can only be in one state at the same 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 fullfills 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*

$$p(t, x, A) = \mathbb{P}(X_{t+s} \in A \mid X_s = x). \tag{1.1}$$

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 on the future of the process; as we can see on the right side of (**??**).

?

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[**?** , 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 homogeneity?

Invariant Measure

stat. distr., stat. 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*

$$\int_{\mathbb{X}} p(t, x, A)\mu(\mathrm{d}x) = \mu(A). \tag{1.2}$$

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 $u : \mathbb{X} \to \mathbb{R}$ with $\int_{\mathbb{X}} |u|\mu(\mathrm{d}x) < \infty$ we have*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T u(X_t)\,\mathrm{d}t = \int_{\mathbb{X}} u(x)\mu(\mathrm{d}x). \tag{1.3}$$

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*

$$\int_A p(t,x,B)\mu(\mathrm{d}x) = \int_B p(t,x,A)\mu(\mathrm{d}x) \tag{1.4}$$

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

If the stochastic transition function is absolutely continuous w.r.t. μ, i.e. ... then reversibility corresponds to $p(t,x,y) = p(t,y,x)$ for all $t \in \mathbb{T}$ and μ-a.e. $x, y \in \mathbb{X}$.

rev. for Markov chain bzw. abs. cont. meas.
rev. follows ex. stat./inv. dist.?
det.bal.?

Example (Markov Chain)

To illustrate the previous definitions on an easy example, we set them in relation to the familiar/well-known special case of Markov Chains.

Let X_t be a Markov chain with $\mathbb{T} = \mathbb{N}$ and finite state space $\mathbb{X} = \{1, \dots, n\}$ and with positive invariant measure μ. Since we are considering 1-step transitions, the associated transition function is given by $p(x,y) = p(1,x,y)$ and corresponds to the entries of the transition matrix $T \in \mathbb{R}^{n\times n}$, i.e.

$$T_{x,y} = p(x,y).$$

The propagation of a vector(distribution) $v_0 \in \mathbb{R}^n$ in the unweighted state space can be written as $v_1^T = v_0^T T$, where v_0^T denotes the transposed vector. The invariant measure $\mu \in \mathbb{R}^n$ satisfies $\mu^T = \mu^T T$.

inv. meas. = stat. meas. = stat. dist. = equi. dist.?
weights p.29 (later)

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, we need some convenient integrable spaces on which the transfer operator can act. We are going to define an operator which acts just on general L^r-spaces, i.e. r-integrable spaces. no?

Definition 1.7 (L^r-Spaces)**.** *We define the r-integrable spaces as*

$$L^r_\mu(\mathbb{X}) = \{f : \mathbb{X} \to \mathbb{R} \mid \int_{\mathbb{X}} |f(x)|^r \mu(\mathrm{d}x) < \infty\}$$

for $1 \leq r < \infty$ and

$$L^\infty_\mu(\mathbb{X}) = \{f : \mathbb{X} \to \mathbb{R} \mid \operatorname*{ess\,sup}_{x \in \mathbb{X}} |f(x)|^r \mu(\mathrm{d}x) < \infty\}.$$

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 $f, g \in L^2$ we define complex needed?

$$\langle f, g \rangle_\mu := \int_{\mathbb{X}} f(x) g(x) \mu(\mathrm{d}x).$$

Now let ν_0 be the density function of a given start distribution. Then the density function of a subset $A \subset \mathbb{X}$ at time t is given in terms of the transition function by

$$\nu_t(A) = \int_{\mathbb{X}} \nu_0 p(t, x, A) \mu(\mathrm{d}x).$$

On the other hand, the density ν_t is given by

$$\nu_t(A) = \int_A \nu_t(x) \mu(\mathrm{d}x).$$

Transfer Operator and spectral properties

The two equations above yield in the following intuitive definition of a transfer operator which should "propagate" probability densities according to a given Markov process. But instead of limiting us to density functions, we define the transfer operator as acting on any r-integrable function. propagates

Definition 1.8 (Transfer Operator)**.** *Let $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \to [0, 1]$ 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 $\mathcal{T}^t : L^r_\mu(\mathbb{X}) \to L^r_\mu(\mathbb{X})$ with $t \in T$ and $1 \leq r \leq \infty$ is defined via* subensembles? propagator

$$\int_A \mathcal{T}^t \nu(y) \mu(\mathrm{d}y) = \int_{\mathbb{X}} \nu(x) p(t, x, A) \mu(\mathrm{d}x) \tag{1.5}$$

for all measurable $A \subset \mathbb{X}$ and $\nu \in L^r$. $A \in \Sigma$

The transfer operator is well-defined, see [**?**]. We will announce here already some properties of the transfer operator which will be useful in the following chapter(s). $\mathcal{T}^t$ is a Markov operator; i.e. it conserves the norm, $\|\mathcal{T}^t\nu\|_1 = \|\nu\|_1$, and is positive, $\mathcal{T}^t\nu \geq 0$ for $\nu \geq 0$. $\mathcal{T}^t\nu_0$ 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:

Markov operator?

$$\nu_0 \mapsto \nu_t = \mathcal{T}^t\nu_0.$$

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

$$\mathcal{T}^t\mathbb{1}_{\mathbb{X}} = \mathbb{1}_{\mathbb{X}}.$$

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

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

why only in L2?
stationary measure π

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!

Definition 1.9 (Self-adjoint Operator)**.** *An operator $\mathcal{T}$ on $L^2(\mu)$ is called self-adjoint if for all $f, g \in L^2_\mu(\mathbb{X})$*

$$\langle f, \mathcal{T}g\rangle_\mu = \langle \mathcal{T}f, g\rangle_\mu.$$

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

Proof. □

Theorem 1.11. *Let $\mathcal{T}^t : L^2_\mu(\mathbb{X}) \subset L^1_\mu(\mathbb{X}) \to L^2_\mu(\mathbb{X})$ denote the propagator corresponding to the Markov process X_t. Then $\mathcal{T}^t$ is self-adjoint with repect to the scalar product $\langle\cdot,\cdot\rangle_\mu$ in $L^2_\mu(\mathbb{X})$ if and only if X_t is reversible.*

Proof. Huisinga[**?**] □

Since the spectral radius of any transfer operator is 1, it follows from the previous two theorems that a reversible process has a spectrum $\sigma(\mathcal{T}) \subset [-1, 1]$.

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.

Backwards Operator

Definition 1.12 (Backwards Operator)**.** *The backwards transfer operator* $\mathcal{U}^t : L^r(\mu) \to L^r(\mu)$ *with* $t \in \mathbb{T}$ *and* $1 \leq r \leq \infty$ *is defined by*

$$\mathcal{U}f(x) = \int_{\mathbb{X}} f(y)p(t, x, \mathrm{d}y). \tag{1.6}$$

For all $t \in \mathbb{T}$ we have again

$$\mathcal{U}^t \mathbb{1}_{\mathbb{X}} = \mathbb{1}_{\mathbb{X}}.$$

The operator $\mathcal{U}$ is *adjoint* to $\mathcal{T}^t$, that is they are related via the duality bracket, namely

$$\langle \mathcal{T}^t f, g\rangle_\mu = \langle f, \mathcal{U}^t g\rangle_\mu \tag{1.7}$$

for all $f, g \in L(?)$. Thus, if a Markov process X_t is reversible, then we have

$$\mathcal{T}f(x) = \mathcal{U}f(x)$$

for the corresponding forward and backward operators $\mathcal{T}$ and $\mathcal{U}$.

Infinitesimal Generator

describes the rate a MC moves between states

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

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

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

proof time-indep.?

Definition 1.13 (Infinitesimal Generator)**.** *For the semigroup of propagators or forward transfer operators* $\mathcal{T}^t : L^r_\mu(\mathbb{X}) \to L^r_\mu(\mathbb{X})$ *with* $t \in T$ *and* $1 \leq r \leq \infty$ *we define* $\mathcal{D}(L)$ *as the set of all* $\nu \in L^r_\mu(\mathbb{X})$ *s.t. the strong limit*

$$\mathcal{Q}\nu = \lim_{t\to 0} \frac{\mathcal{T}^t\nu - \nu}{t}.$$

exists. Then the operator $\mathcal{Q} : \mathcal{D}(L) \to L^r_\mu(\mathbb{X})$ *is called the infinitesimal generator corresponding to the semigroup* $\mathcal{T}^t$*.*

The infinitesimal generator is an operator which describes the behaviour of a Markov process in infinitesimal time. That becomes clear by the relation $\mathcal{T}^t = \exp(t\mathcal{Q})$ in L^2. We can say that $\mathcal{Q}$ "generates" the semigroup of transfer operators since the whole semi-group of transfer operators can be derived from it.

ref

So the eigenvalues $1 = \lambda_1, \ldots, \lambda_m$ of the propagator $\mathcal{T}^t$ are related to the eigenvalues $0 = \Lambda_1, \ldots, \Lambda_m$ of the propagator L via

not yet discrete?

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

for all $1 \leq k \leq m$. Their corresponding (associated) eigenfunctions are identical. Thus, the stationary distribution of $\mathcal{T}^t$ is the solution of $\mathcal{Q}\pi = 0$; $\mathcal{Q}\mathbb{1} = 0$.

properties of generator ?

1.3 Galerkin Projection

So far we considered Markov processes on very large (possibly continuous) state spaces. Since we are often/mainly interested in computations/simulations on/of a process, we are now going to create a process on a smaller (namely finite) state space which shall inherit the most important properties of our original process. This can be done by a Galerkin projection/discretization.

discr./finite?

Galerkin Projection

The first step in order to create our desired finite process is to determine a convenient state space $D \subset L^2(\mu)$. Instead of just chosing characteristic functions as the basis of our new state space, we will adopt the concept of a partition of unity. Using that more general idea gives us more possibilities/options in/for later applications.

real?

Definition 1.14 (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 fullfilled:*

why L2?

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

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

a.e.??

Definition 1.15 (Galerkin Projection)**.** *Let* $\{\chi_1, \dots, \chi_n\}$ *be a partition of unity,* $D = \operatorname{span}\{\chi_1, \dots, \chi_n\}$ *the associated finite-dimensional ansatz space and* $\hat{S} \in \mathbb{R}^{n \times n}$ *with* $\hat{S}_{kj} = \langle \chi_k, \chi_j \rangle_\mu$*. The Galerkin projection onto* D *is defined by* $G : L^2(\mu) \to D$

$$G\nu = \sum_{k,j=1}^{n} \hat{S}^{-1}(k,j) \langle \chi_k, \nu \rangle_\mu \chi_j. \tag{1.8}$$

The matrix $\hat{S}$ is invertible since it is the Gramian matrix of linear independent functions. In the easy case that the $\{\chi_1, \dots, \chi_n\}$ are the characteristic functions belonging to a full partition $\{A_1, \dots, A_n\}$, equation (**??**) becomes

"weighted" orthogonal projection?

$$G\nu = \sum_{k=1}^{n} \frac{1}{\mu(A_k)} \langle \chi_k, \nu \rangle_\mu \chi_k,$$

since the χ_i are orthogonal which means that $\chi_k \chi_j = 1$ if $j = k$ and 0 otherwise.

A Galerkin projection can be applied on the transfer operator of a Markov process.

Definition 1.16 (Projected Transfer Operator)**.** *Let* $\mathcal{P} := \mathcal{P}^t$ *be the transfer operator of a Markov process on a state space* $\mathbb{X}$ *with unique invariant measure* μ*,* $\{\chi_1, \dots, \chi_n\}$ *be a partition of unity and* G *the Galerkin projection onto the associated subspace* D*. Then an operator of the form*

$$G\mathcal{P}G : L^2_\mu(\mathbb{X}) \to D$$

is called projected transfer operator and we abbreviate it by $G(\mathcal{P})$*.*

Matrix Representation

left or right?

Since we are interested in transitions inside of our smaller (projected) space, we want to propagate n-dimensional vectors by the projected transfer operator. That's why we consider now the projection of the restricted transfer operator $G\mathcal{P}_{|D} : D \to D$.

def. GPG not $n \times n$, but $\infty \times n$?

We remind that every linear map between finite-dimensional vector spaces can be described by a matrix which is determined by chosen bases. So we can write the projected transfer operator as a $n \times n$-matrix in a way that will be very useful later.

GPG same as GP_D

Theorem 1.17. *Let $\mathcal{P}$ be the transfer operator of a Markov process, $\{\chi_1, \ldots, \chi_n\}$ a partition of unity and $G\mathcal{P}G$ the Galerkin projection of the transfer operator onto the associated subspace. Then $G\mathcal{P}G$ has a matrix representation*

ansatz space vs subspace

skillnad QPQ vs P

$$P = S^{-1}T,$$

where

$$S_{kj} = \frac{\hat{S}(k,j)}{\langle \chi_k, \mathbb{1} \rangle_\mu} = \frac{\langle \chi_k, \chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \tag{1.9}$$

and

$\mathbb{1}_D \mathbb{1}_\mathbb{X}$?

$$T_{kj} = \frac{\langle \chi_k, \mathcal{P}\chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu}. \tag{1.10}$$

Proof. Remember that P_c is a (left) matrix representation of $G\mathcal{P}G$ with respect to a basis $\{\psi_1, \ldots, \psi_n\}$ of D if for any function $f \in D$ with

$f : D \to L^2$ or $f : D \to D$ possible?

$$f = \sum_{i=1}^{n} \alpha_i \psi_i \quad \text{and} \quad (G\mathcal{P}G)f = \sum_{i=1}^{n} \beta_i \psi_i$$

it holds that

$$(\alpha_1, \ldots, \alpha_n) P_c = (\beta_1, \ldots, \beta_n). \tag{1.11}$$

We assume that (**??**) holds and choose a basis $\{\psi_1, \ldots, \psi_n\}$ of D with

$$\psi_k = \frac{\chi_k}{\langle \chi_k, \mathbb{1}_\mathbb{X} \rangle_\mu}.$$

By applying the projected transfer operator to f, we try to find the coefficient vector $(\beta_1, \ldots, \beta_n)$. Therefor we exploit the definition of a Galerkin projection and

the definition of our basis:

$$\begin{aligned}(G\mathcal{P}G)f &= \sum_{k=1}^{n} \alpha_k (G\mathcal{P})f \\ &= \sum_{k,l,j=1}^{n} \alpha_k \hat{S}^{-1}(j,l)\langle \chi_j, \mathcal{P}\psi_k\rangle_\mu \chi_l \\ &= \sum_{k,l,j=1}^{n} \alpha_k \hat{S}^{-1}(j,l)\langle \chi_j, \mathcal{P}\psi_k\rangle_\mu \langle \chi_l, \mathbb{1}_{\mathbb{X}}\rangle_\mu \psi_l \\ &= \sum_{l=1}^{n} \beta_l \psi_l.\end{aligned}$$

That means that

$$\begin{aligned}\beta_l &= \sum_{k,j=1}^{n} \alpha_k \hat{S}^{-1}(j,l)\langle \chi_l, \mathbb{1}_{\mathbb{X}}\rangle_\mu \langle \chi_j, \mathcal{P}\psi_k\rangle_\mu \\ &= \sum_{k=1}^{n} \alpha_k \underbrace{\sum_{j=1}^{n} \hat{S}^{-1}(j,l)\langle \chi_l, \mathbb{1}_{\mathbb{X}}\rangle_\mu \frac{\langle \chi_j, \mathcal{P}\chi_k\rangle_\mu}{\langle \chi_k, \mathbb{1}_{\mathbb{X}}\rangle_\mu}}_{P_{kl}}. \end{aligned} \tag{1.12}$$

The underbraced term is equal to P_{kl} because of (**??**). Now we can compare it to the (k,l)-th entry of TS^{-1} which is computed via matrix multiplication as

not correct! change S,T?

$$\begin{aligned}(TS^{-1})_{kl} &= \sum_{j=1}^{n} T_{kj}(S^{-1})_{jl} \\ &= \sum_{j=1}^{n} \frac{\langle \chi_k, \mathcal{P}\chi_j\rangle}{\langle \chi_k, \mathbb{1}\rangle}(\hat{S}^{-1})_{jl}\langle \chi_j, \mathbb{1}\rangle.\end{aligned}$$

We see that this corresponds to the (k,l)-th entry of P_c, compare with (**??**). □

Theorem 1.18. *The matrices S and T are stochastic.*

non-neg.

Proof. In order to be stochastic, each row must sum up to 1. We exploit the partition of unity property $\sum_j \chi_j = 1$ for all j and the aforementioned property $\mathcal{P}\mathbb{1} = \mathbb{1}$ of a transfer operator:

$$\sum_{j=1}^{n} S_{kj} = \frac{\langle \chi_k, \sum_j \chi_j\rangle_\mu}{\langle \chi_k, \mathbb{1}\rangle_\mu} = \frac{\langle \chi_k, \mathbb{1}\rangle_\mu}{\langle \chi_k, \mathbb{1}\rangle_\mu} = 1,$$

$$\sum_{j=1}^{n} T_{kj} = \frac{\langle \chi_k, \mathcal{P} \sum_j \chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} = \frac{\langle \chi_k, \mathcal{P}\mathbb{1} \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} = 1.$$

□

Since S and T are both stochastic matrices, they have $\mathbb{1}_D$ as right eigenvector to the eigenvalue 1. It implies that the same holds for P, i.e. the product $S^{-1}T$ is at least pseudostochastic, i.e. its rows sum up to 1. But nonnegativity is not assured since inverting S can provoke/evoke/produce/cause negative entries. The non-negativity depends on the choice of the partition of unity. There are examples s.t. $S^{-1}T$ is a stochastic matrix.

Example

A good way to understand the concept of a Galerkin Projection is to consider the case of a full partition discretization.

Theorem 1.19 (Full Partition Discretization)**.** *Let $\{\chi_1, \ldots, \chi_n\}$ be a partition of unity that is induced by a full partition, i.e. the χ_i are the characteristic functions of pairwise disjoint sets A_i s.t. $\cup A_i = \mathbb{X}$. Then the matrix representation P of $G\mathcal{P}G$ is a stochastic matrix consisting of the transition probabilities between the partition sets, i.e.*

$$P(k,l) = \mathbb{P}(X_t \in A_l \mid X_0 \in X_k).$$

Proof. The entries of the Gram matrix $\hat{S}$ are $\mu(A_k)$ on the diagonal and 0 everywhere else. We can deduce that S is the unit matrix, while $P = T$ is a stochastic matrix. For the entries of P we get: makes role of S, T clear

$$P(k,l) = \frac{\langle \chi_l, \mathcal{P}\chi_k \rangle_\mu}{\langle \chi_k, \mathbb{1}_{\mathbb{X}} \rangle_\mu} = \frac{\mathbb{P}_\mu(\{X_t \in A_l\} \cap \{X_0 \in A_k\})}{\mathbb{P}_\mu(X_0 \in A_k)} = \mathbb{P}(X_t \in A_l \mid X_0 \in X_k).$$

□

In this case, P is a stochastic matrix and hence the transition matrix of a Markov chain. Its state space consists just of the partition sets A_i. The stationary distribution of the so defined Markov chain P_c is just the projection of the invariant measure μ onto D. ref to 1-1-rel? what for?

For a full partition discretization, the matrix S is a diagonal matrix. If we choose a partition of unity that is *close* to a full partition, i.e. we choose *almost characteristic functions*, then the matrix S is not diagonal, but close to that. We will later see the consequences of that fact regarding to the examination/investigations of the *rebinding effect*.

Properties of Galerkin Projection

As the matrix representation of a projected transfer operator is in general *not* a stochastic matrix, which is equivalent to being the transition matrix of a Markov chain (see ..), we see immediately that the process can lose its Markovianity by projecting it (onto a subspace).

This possible loss of Markovianity is certainly a really undesirable effect. But before examining that later in section 2.4, let us now first analyze further, hopefully *nice*, properties of the matrix representation P.

We already know that the matrices S and T from Theorem **??** are stochastic matrices. This leads to some good properties of P:

- The eigenvalue $\lambda = 1$ of P has the associated right-eigenvector $e = (1, \dots, 1)^T$ and left-eigenvector $\hat{\mu}^T$ from theorem **??**.
- If $\mathcal{P}$ is self-adjoint in L^2, then $G\mathcal{P}G$ as well. Then the matrices S and T are self-adjoint with respect to the discrete scalar product

  P self-adj.? $\langle Av, w\rangle = \langle v, Aw\rangle$

  $$\langle u, v\rangle_{\hat{\mu}} = \sum_{i=1}^{n} u_i v_i \hat{\mu}_i.$$

  Since self-adjointness of the operator is equivalent to reversibility of the corresponding process, detailed balance equation (e.g. $\hat{\mu}_k T_{kl} = \hat{\mu}_l T_{lk}$ for all $k, l = 1, \dots, n$) is fulfilled for S and T.

  and so, eigenvalues of S and T are real and in $[-1, 1]$
- If the transfer operator has a simple and dominant eigenvalue 1 and the continuous part of the spectrum is bounded away from the discrete part, then the process is irreducible and aperiodic which is inherited by the matrix T. In particular T has the simple and dominant eigenvalue $\lambda = 1$ which is the only eigenvalue with $|\lambda| = 1$ and the discrete invariant density $\hat{\mu}$ is the unique invariant density of T.
- As seen in the last example a full-partition projection yields the transition matrix $P = T$ of a Markov chain describing transitions between the partition sets.

Theorem 1.20. *The matrix representation P from Thm **??** has the left eigenvector*

$$\hat{\mu} = \langle \mathbb{1}, \chi_k\rangle_{\mu} = \int_{\mathbb{X}} \chi_k(x)\mu(\mathrm{d}x).$$

Proof. We observe that $\hat{\mu}^T S = \hat{\mu}^T$ and $\hat{\mu}^T T = \hat{\mu}^T$ since

$$(\hat{\mu}^T S)_j = \sum_{k=1}^{n} \langle \mathbb{1}, \chi_k \rangle_\mu \frac{\langle \chi_k, \chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} = \langle \mathbb{1}, \chi_j \rangle_\mu = \hat{\mu}_j$$

and

$$(\hat{\mu}^T T)_j = \sum_{k=1}^{n} \langle \chi_k, \mathcal{P}\chi_j \rangle_\mu = \langle \mathbb{1}, \mathcal{P}\chi_j \rangle_\mu = ... = \hat{\mu}_j.$$

We can deduce that $\hat{\mu}^T P = \hat{\mu}^T T S^{-1} = \hat{\mu}^T S^{-1} = \hat{\mu}^T S S^{-1} = \hat{\mu}^T$. □

Summary

Summarizing, the discretization of the propagator can be interpreted as a *coarse graining* procedure. ... blabla ..... The discretization maintains/keeps/inherits the most important properties of the transfer operator/propagator.

Projected infinitesimal generator

The Galerkin projection of an infinitesimal generator yields a similar matrix representation as it has been the case for the transfer operator. It can also be written as the product of two stochastic matrices, one of them inverted.

Theorem 1.21. *Let $\mathcal{Q} : L^2(\mu) \to L^2(\mu)$ be a generator of a semigroup of transfer operators with unique invariant measure μ and satisfying $\mathcal{Q}\mathbb{1}_\mathbb{X} = 0$. Let χ be a partition of unity with a projection G onto the associated subspace spanned by χ. Then the projected generator $G\mathcal{Q}G$ has the matrix representation $Q = RS^{-1}$ with the stochastic mass matrix S from (**??**) and*

not commutative

$$R(k,j) = \frac{\langle \chi_k, \mathcal{Q}\chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \tag{1.13}$$

*The eigenvalue problem of Q is equivalent to the generalized eigenvalue problem $Ru = \Lambda S u$. For both Q and R the largest eigenvalue is $\lambda = 0$. The associated right eigenvector is $e = (1, \ldots, 1)^T$, the associated left eigenvector is $\hat{\mu}^T$ from theorem **??**.*

Proof. The matrix representation of $G\mathcal{Q}G$ can be shown similar to the proof of theorem **??**. For the other properties see... □

There are obiously many possibilities/options to make a Galerkin discretization/projection of the propagator/generator of a given process because we can define it on an arbitrary partition of unity $\chi_1, \ldots, \chi_n$. In chapter **??** we are going to see which choice of χ gives us a *good* discretization of our process in the sense that it maintains certain desired properties; in our case the long-time behaviour of the process using so called *metastability*.

1.4 Recrossing Effect

In the last section, we have seen that the projected transfer operator of a Markov process respectively its matrix representation inherit some of the most important properties of the original process. Unfortunately, not everything works out in such a good way. For instance, one further desirable property would be a commuting behaviour in/concerning propagation and projection of the process. That is, it should make no difference in which order these two operations are executed. Unfortunately/but it will turn out that this is not the case.

We will make that clear by the example of a full-partition discretization which has already been partly examined in the previous sections.

Initial Situation

Assume we are given a Markov process $(X_t)_{t\in\mathbb{T}}$ on a continuous or very large state space $\mathbb{X}$. In order to get a discrete process out of it, we are going to discretize the time onto $\mathbb{N}$ and the state space onto a finite set $\{1,\dots,n\}$. Discretizing the time can be done naturally without problems since for every lag-time $\tau > 0$, the process $(X_{k\tau})_{k\in\mathbb{N}}$ is again Markovian. why?

However, the state-space discretization has to be examined/observed a bit more differentiated/elaborated/sophisticated. We apply the Galerkin projection as described in section **??**, but we will see that some problems/difficulties in our model can occur; depending on the order of projection and propagation. On the one hand, we can assign a Galerkin Projection $G\mathcal{P}G$ to a given transfer operator $\mathcal{P}$ which results in a matrix representation P_c (see theorem **??**). Probability distributions can be propagated by multiplication with P_c^k which corresponds to the operator $(G\mathcal{P}G)^k$. This model will be called $(\widehat{X}_k)$. On the other hand, we can firstly propagate the process via the transfer operator $\mathcal{P}^k := \mathcal{P}(\tau)^k$. Projecting it afterwards yields the new operator $G\mathcal{P}^kG$. This process will be called $(\widetilde{X}_k)$. fehlerfortpflanzu

A desirable behaviour of our models would be that $(\widehat{X}_k)$ and $(\widetilde{X}_k)$ have the same trajectory when started on the same initial distributions $\widehat{X}_0$ and $\widetilde{X}_0$. It will turn out that this is normally not the case. Another desirable property would be Markovianity of both models, since this is the case for the original process (X_t). But we have alredy seen in theorem **??** that the matrix representation P_c of $G\mathcal{P}G$ is in general not a transition matrix, i.e. $(\widehat{X}_k)$ is not Markovian. naja

Example: Full-partition discretization

Look at these two different models for the example of a full-partition discretization. Assume we have fixed a lag-time $\tau > 0$ and are given the transfer operator $\mathcal{P} := \mathcal{P}(\tau)$

of a Markov process $(X_t)_{t\in\mathbb{T}}$.

Consider the operator $G\mathcal{P}^kG$, i.e. first propagate the (original) process and project it afterwards. Then for all k-multiples of τ, we assign the current state of the original process X_t to the projected process $\widetilde{X}_k$:

$$\widetilde{X}_k = i \Leftrightarrow X_{k\tau} \in A_i.$$

$(\widetilde{X}_k)$ describes the snapshot dynamics of (X_t) with lag time τ between the sets $A_1, \ldots, A_n$. The so defined process is not necessarily Markovian as we will see in the next application.

Let $(\widehat{X}_k)_{k\in\mathbb{N}}$ be the Markov chain that is described by the transition matrix P_c, i.e. the matrix representation of the discretized transfer operator. Since it is a model of the non-Markovian process $(\widetilde{X}_k)_{k\in\mathbb{N}}$, there will be some differences between $(\widehat{X}_k)$ and $(\widetilde{X}_k)$. *Markov State Model*

Notice that just in this particular example $G\mathcal{P}G$ is a Markov chain. Normally TS^{-1} corresponding to the projected process is not a a transition matrix. So in the normal case, neither of prop-proj and proj-prop. are Markovian! But can still differ! Which we are going to examine at the end of this section.

Example: Double Well Potential

Consider a double-well potential $V(x) = (x^2 - 1)^2$ with a full-partition into two sets A and B around the local minima of the energy landscape, as shown in figure **??**. Now compare the two model-processes $\widetilde{X}_k$ and $\widehat{X}_k$. We already know that $\widehat{X}_k$

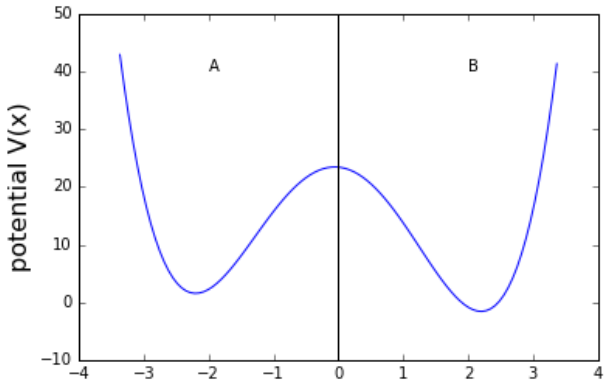

Figure 1.1: full-partition of a double-well potential

is a Markov chain. How about $\widetilde{X}_k$? Lets investigate in possible memory effects. For a small lag-time $\tau = 0.1$ compare the probabilities $\widetilde{X}$?

$$\mathbb{P}_\mu[\widetilde{X}_{(k+1)\tau} \in A \mid \widetilde{X}_{k\tau} \in B] \text{ and } \mathbb{P}_\mu[\widetilde{X}_{(k+1)\tau} \in A \mid \widetilde{X}_{k\tau} \in B, \widetilde{X}_{(k-1)\tau} \in A].$$

We get v densities?

$$\mathbb{P}_\mu[\widetilde{X}_{(k+1)\tau} \in A \mid \widetilde{X}_{k\tau} \in B] = \int_A v_B^\tau(x)\,\mathrm{d}x = ..., \tag{1.14}$$

$$\mathbb{P}_\mu[\widetilde{X}_{(k+1)\tau} \in A \mid \widetilde{X}_{k\tau} \in B, \widetilde{X}_{(k-1)\tau} \in A] = \int_A v_{BA}^\tau(x)\,\mathrm{d}x = ... \tag{1.15}$$

So we see that for such a short lag-time τ, the process $\widetilde{X}_k$ is not memoryless and hence not a Markov process. That effect is intuitively clear. Equation (**??**) describes the probability to get from B to A, so it is averaged(?) over all possible starting points in B. Comparing to that in (**??**), being in A immediately one time-step before being in B increases the probability that the process is still in the transition region. And to get to A from the transition region in B is just more likely than from any other region inside B.

On the other hand, if we choose a large lag-time $\tau = 100$, then the past transition from A to B took place a long time ago. So we cannot certainly know if the process is still in the critical transition region; during that long lag-time it could also have been gone anywhere else. That means that the memory effect included in $\widetilde{X}$ could be called a *short-time memory*.

Comparing that to $\widehat{X}$. $\widehat{X}$ is a Markov chain on the two possible states (=partition sets). Its transition matrix consists of the transition probabilities between these two sets within time τ. But as these probabilities are built from an originally continuous state space, they are just averaged over the whole space. That means, the probability to get from A to B under $\widehat{X}$ is always the same which is not an appropriate description of the original process. In fact/for instance, being in the transition region (i.e. close to $x = 0$) inside of A yields a much higher probability to get into B in comparison to start inside of the energy minimum of A. But these differences are not included in our Markov State Model.

Loss of Markov Property = Recrossing Effect

later = rebinding + proj.prop != prop.proj.? ?

This loss of Markovianity of a process when projecting it onto a finite subspace is called *Recrossing Effect*. It is due to the fact that the projected transfer operator form in general NOT a semi-group. Hence, in general we have

$$(P_c)^k \neq (P^k)_c.$$

An important question when it comes to projections of Markov 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?

$$
\begin{array}{ccc}
 & \mathcal{P}(\tau) & \xrightarrow{\tau\to\tau k} & (\mathcal{P}(\tau))^k \\
 & \Big\downarrow{\scriptstyle proj.} & & \Big\downarrow{\scriptstyle proj.} \\
G(\mathcal{P}(\tau))\widehat{=} & P_c(\tau) & \xrightarrow{\tau\to\tau k} & (P_c(\tau))^k
\end{array}
$$

Figure 1.2: Projecting/propagating a transfer operator (non-commutative)

Discretization Error (Density Propagating Error)

what about eigenv. err.?

We will present here shortly how the discretization error can be estimated in general. For our purposes that will not play an important role, since we can zurückgreifen on a transfer operator by Weber[**?**] which allows a projection without(?) error.

The maximal possible error between the distributions of $\widetilde{X}_k$ and $\widehat{X}_k$ after k time-steps is (independently of initial distribution) given by

which norm?

$$E(k) = \|G(\mathcal{P}^k) - (G(\mathcal{P}))^k\|.$$

Theorem 1.22. *Assume the discrete/dominant spectrum of a transfer operator $\mathcal{P}$ is given/denoted/ordered by $1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_n$. Then the projection error can be bounded from above in terms of the second-largest eigenvalue by*

$$\|(G(\mathcal{P}))^k - \Pi_0\| \leq \lambda_1^k,$$

where Π_0 is the orthogonal projection of ... .

For a proof, see Schütte and Sarich[**?** , p.72]. In the following chapter we will see further/deeper relations between the spectrum of the transfer operator and ... properties. We will see how to choose partition for a MSM s.t. the approximation error becomes small.

what is markov state model? def!

P_c ?

2 Dominant Structures in Markov Processes

In section **??**, we already mentioned that under certain conditions (high complexity/ long-time simulations) a reduction of dimension of a given Markov process is required. But we don't yet know how to choose a partition of unity such that the corresponding Galerkin projection yields a reasonable Markov State Model. In order to solve that problem, we are now going to introduce the concept of *metastability* which is a certain behaviour of the trajectory of stochastic processes and describes the characteristics of rare transitions between specific subsets after a long duration of stay inside. We will see why it makes sense to choose these *metastable sets* as clustering sets in the aforementioned Galerkin Projection and how they can be detected using the spectrum of the transfer operator of the process. We will also see that the optimal metastable decomposition is not sharp/crisp but soft/fuzzy.

clustering fcts.

Unfortunately, metastability of nonreversible processes is much harder to detect, since we are not guaranteed a real spectrum of the transfer operator. So we need the concept of dominant cycles instead of dominant sets. We will give a short outlook how that case could be handled.

2.1 Metastability

There exist several different definitions of metastability. Shortly said, metastability is the property of a Markov process which consists of subsets/regions of the state space s.t. transitions between these subsets are rare events while the duration of stay inside of each of these subsets is rather long. Some possible characterizations of that behaviour are based on large hitting times or small exit rates, see Schütte and Sarich[**?** , chapter 3] which gives a good overview of the most common definitions.

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.

Analoguously, we can define a *metastable* or *almost invariant subset* as a subset $A \subset \mathbb{X}$ 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. Thus, a full partition $A_1, \ldots, A_m$ of the state space $\mathbb{X}$ is called *metastable* if

metastable decomposition

$$\sum_{k=1}^{m} \mathbb{P}_\mu(X_{t_f} \in A_k \mid X_0 \in A_k) \approx m. \tag{2.1}$$

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.

Obviously, being "close to 1" or "close to m" are rather vague statements. But that lack of concreteness will be eliminated later, since we will only be interested in the "best" metastable decomposition. That means that we want to obtain a decomposition where the probability to stay inside of a metastable set is as close as possible to 1, resulting in the sum (**??**) being as close as possible to m. Instead, we will have to determine the number of subsets we are looking for.

?

Also the choice of the timescale t_f is not specified in general and will depend on the particular system in consideration.

Have projected transfer operator P by Galerkin projection with $\{\chi_1, \ldots, \chi_n\}$. Then the trace of P is referred to as *metastability* of the conformations $\{\chi_1, \ldots, \chi_n\}$.

later?

Metastability in Molecular Dynamic Systems

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

conformation = spatial arrangement?

This behaviour is shown in figure **??**, taken from ... . This example represents an bla? system. It consists of ? molecules and can take ? values. As we can see, the process has two conformations (red and blue) where the process stays for a long time and rare transitions between these two conformations.

special case of metastability?

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 **??**, long-time simulations

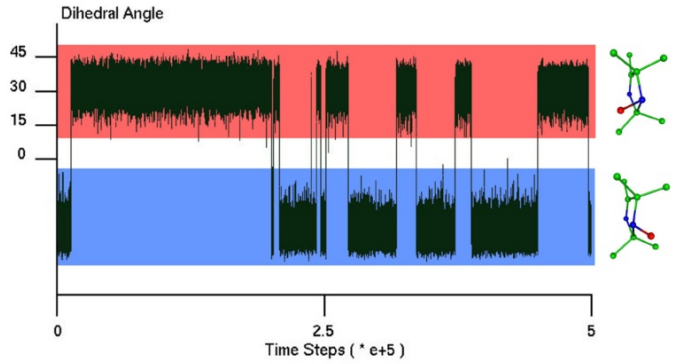

Figure 2.1: Example of a molecule with two conformations

of such complex systems are not feasible even with the best computers nowaday, see (...).

Thus, in order to be able to compute some long-time simulations of a given MD system, a reduction of complexity is needed/required. This can be achieved by a clustering of the state space as described in section **??**. 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. wich states of the original process should be grouped together for the reduced model.

Galerkin = clustering?

As we are mainly interested in the long-time behaviour of a given process, it seems reasonable to cluster states of a metastable set together and create a new process where each (macro) state corresponds to one of the metastable sets. The transition probabilities of the clustered process/reduced model should correspond to the transition rates of the original process between its metastable sets. As metastability describes a behaviour on long timescales, our newly created process should maintain the long-time behaviour of the original process, but *forget* about its short-time transitions, i.e. transitions inside of a conformation/metastable set.

micro/macro states

Since there is not one unique metastable decomposition of the state space, we need to find a decomposition which is in some sense "the best"; then we can use it to create a reduced model. In sections **??** and **??** we will see how to find such a decomposition.

Advantages / Disadvantages

weg?

Most importantly, our newly created process will have the desired property of a reduced dimension/complexity since the model acts on a smaller state space while maintaining the crucial property of the original process (transitions between metastable sets = long-time behaviour). 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 these (at least for our case!) superfluous informations. But there is also a disadvantage, as already mentioned in section **??**, by projecting a process it can lose its Markov property.

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[**?**]. 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 $L^2_\mu(\mathbb{X})$.

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

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:

C1 The essential spectral radius of $\mathcal{P}$ is less than one; i.e. $r_{\text{ess}} < 1$.

C2 The eigenvalue $\lambda = 1$ pf $\mathcal{P}$ 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[**?**]. 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 absolut 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.

has periodic structures? excludes modeling and interpretation problems

C2 = ergodic?

Theorem 2.1 (Schütte[**?** , theorem 4.16])**.** *The transfer operator $\mathcal{P} : L^2 \to L^2$ of a reversible process with properties* **C1** *and* **C2** *has the following spectrum:*

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

with $-1 < a \leq b < \lambda_n \leq \cdots \leq \lambda_1 = 1$ and isolated, not necessarily simple eigenvalues of finite multiplicity that are counted according to multiplicity.

why discr. rechts?

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/ Optimal Decomposition

Theorem 2.2 (Schütte[**?** , theorem 4.16])**.** *The metastability of an arbitrary decomposition $\mathcal{D} = \{A_1, \ldots, A_m\}$ of the state space $\mathbb{X}$ can be bounded from above by*

$$p(A_1, A_1), + \cdots + p(A_m, A_m) \leq 1 + \lambda_2 + \cdots + \lambda_m.$$

Theorem **??** lets us think about the problem if there exists an optimal decomposition with highest possible metastability. In fact, this problem might unfortunately be ill-conditioned.

The number of metastable sets is determined by the number of dominant eigenvalues.

Relation of dominant eigenfunctions to metastable decomposition

Theorem 2.3 (Schuette[**?**])**.** *Each single eigenfunction induces a metastable decomposition*

Proof. □

The zeros of an eigenfunction (respectively change of sign) induce a metastable decomposition of the state space. Different eigenfunctions result in different decompositions.

eigenfcts vs committor functions

This theorem gives us a first demonstrative relation of the metastability of a process to its eigenvectors. This result is not yet optimal/very good. In the next sections, we will see that linear combinations of eigenvectors result in much better metastability.

In figure **??**, we get a good overview of that relation. We have a potential/energy landscape which has 4 energy minima, i.e. 4 regions where the process could be *trapped* such that it is hard to get outside again. The transition matrix shows this metastable behavious since we can see 4 regions which large probabilities to stay inside and very small probabilities to go to a different region. Furthermore we see that the process has 4 dominant eigenvalues. Three of them have a change of sign, which induces the metastable decomposition.

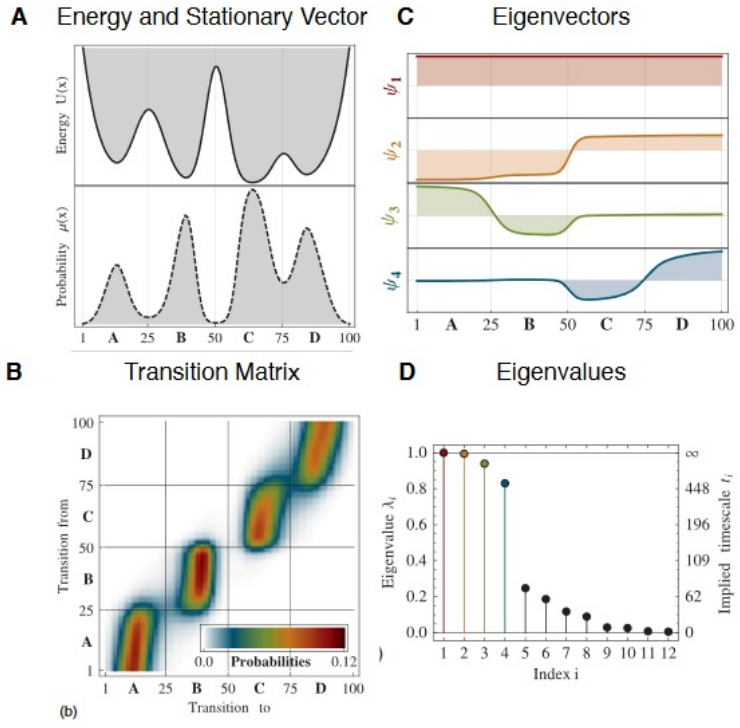

Figure 2.2: Relation of eigenvalues and eigenvectors to metastability of a process

Disadvantages

So the spectral approach is suitable/convenient to characterize metastability of Markov processes. But there are two disadvantages. 1: the result is only appliable on reversible processes, because real eigenvalues are only guaranteed if the transfer operator is self-adjoint. 2: eigenvector problem of the transfer operator has only see global solutions.

Most of all, the previous approach in computing a metastable decomposition doesn't include/consider the transition regions of the process, so we need to refine/improve it.

2.3 Fuzzy Clustering

The above considerations result in a metastable full decomposition of the state space, i.e. each state is assigned to exactly one of the partition sets. 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 rigoruous description of actual behaviour of the process.

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

Fuzzy sets are sets whose elements have degrees of membership.

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 [**?**]

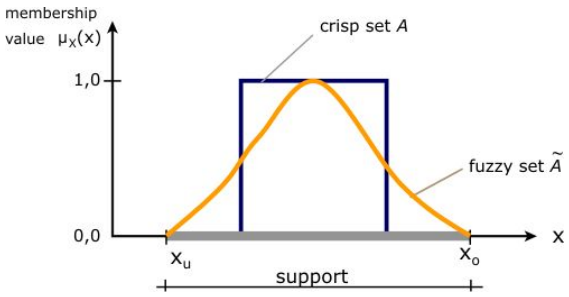

Figure 2.3: Crisp vs Fuzzy Sets/ Clustering

by defining macro states as *overlapping partial densities*. They can be identified by membership functions $\chi_1, \ldots, \chi_n : \mathbb{X} \to [0,1]$. Each state of the original state space shall be assigned to the different macro states with a certain *degree of membership*.

Definition 2.4 (Membership Function)**.** *For any set $\mathbb{X}$, a membership function on $\mathbb{X}$ is a function $m : \mathbb{X} \to [0,1]$.*

Membership functions can be used to decompose the space into metastable sets/clusters. 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 degrees, which can be interpreted as kind of 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 section **??**. That gives us a Markov State Model where each (macro) state is a metastable set of the original process.

so far: membership fct indep. of metast. set

Example 2.5. *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 dominant eigenfunctions might have an overlap.

Statistical Weights

what for?

For each macrostate we can define a statistical weight

$$w_i = \int_{\mathbb{X}} \chi_i(q)\chi_i(q)\,\mathrm{d}\mu(q) \text{ or } \int \chi_i \mathbb{1}\,\mathrm{d}\mu(q)$$

i.e. ....

$D = \mathrm{diag}(w_1, \ldots, w_n)$ is the diagonal matrix of the statistical weights of the membership functions. Then $T = D^{-1}\langle \chi, \mathcal{P}(\tau)\chi \rangle_\mu$, compare theorem **??**.

Perron Cluster Analysis

The term *Perron Cluster Analysis* denotes the objective of clustering a Markov process into metastable sets using the *Perron eigenvalues* respective *Perron eigenfunctions*, which means eigenvalues close to 1 and the corresponding eigenfunctions. Perron Cluster Analysis respectively its algorithmic implementation PCCA (*Perron Cluster Cluster Analysis*) has been developed by Deuflhard et al[**?**] which used the sign structure of the dominant eigenvalues of the transition matrix. That approach has been improved by Deuflhard and Weber[**?**] who transformed the system of eigenvectors into a system of membership functions which results in a soft/fuzzy clustering of the state space of the original process; their algorithm is called PCCA+ (*Robust Perron Cluster Analysis*). Originally, PCCA+ was formulated only for discrete Markov chains, but Weber[**?**] extended it even on continuous processes.

set-based approach?

We are considering the set of dominant eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ with the correponding set of eigenfunctions $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$. They fulfill the eigenvalue problem $\mathcal{P}(\tau)\mathcal{X} = \mathcal{X}\Lambda$ of the transfer operator $\mathcal{P}(\tau)$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. The set of membership functions $\chi = \{\chi_1, \ldots, \chi_n\}$ can be built as a linear combination $\mathcal{X}\mathcal{A}$ of the dominant eigenfunctions, i.e.

eigenvector matrix?

membership function matrix χ

$$\chi_j(q) = \sum_{i=1}^{n} \mathcal{A}_{ij}\mathcal{X}_i(q), \;\; j = 1, \ldots, n. \tag{2.2}$$

Here, $\mathcal{A} = \{\mathcal{A}_{ij}\}_{i,j=1,\ldots,n} \in \mathbb{R}^{n\times n}$ is a real matrix which should be chosen in such a way that the resulting membership functions fulfill the required properties/constraints; i.e. positivity and partition of unity. There are infinitely many transformations $\mathcal{A}$ 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 $\mathcal{A}$ that satisfies some optimality condition. The algorithm PCCA+ computes the matrix $\mathcal{A}$ as the solution of a convex maximization problem, see Weber[**?**].

or only conv.

Weber shows in [**?**] that the choice $\chi = \mathcal{X}\mathcal{A}$ (any linear combination of the eigenfunctions $\mathcal{X}$) preserves the Markov property of the process.

lin.comb.?

even: discretiz. error vanishes

Iteration Error

2.4 Dominant Cycles

The previously described decomposition of Markov processes into metastable sets and in particular its algorithmic computation via PCCA+ is unfortunately only possible/valid for reversible processes. In this section, we give an overview about a similar topic for a special case of nonreversible processes (NESS).

In general, there exist two possible ways to uniquely describe a Markov chain; via a transition matrix or via cycles. For non-reversible processes, this cycle representation is the preferred method.

NESS processes

Definition 2.6 (NESS process)**.** *A Markov process is called nonequilibrium steady state (NESS) process if it is nonreversible, but still has a steady state, given by an invariant measure μ w.r.t. which the process is ergodic.*

what is that?

$p_\tau(A, B) = p(\tau, A, B) = \mathbb{P}(X_\tau = B \mid X_0 = A)$

As a NESS process is nonreversible, there are regions where the detailed balance equation is not fulfilled, i.e. there is an effective probability flow $p(\tau, A, B) - p(\tau, B, A) \neq 0$ between some subsets $A, B \subset S$ of the state space.

Flow of a process

why only finite?

In the following we consider an irreducible and aperiodic (i.e. ergodic) Markov chain on the finite state space $S = \{1, \ldots, n\}$ given by the transition matrix P. Since this Markov chain is irreducible and aperiodic, it possesses a unique invariant measure μ that is positive everywhere. Then μ is the normalized eigenvector of P for the unique eigenvalue $\lambda = 1$.

see ..

Definition 2.7 (Flow Matrix)**.** *The probability flow associated to a Markov process is given by the flow matrix*

$$F = DP,$$

where P is the transition matrix of the process and D the diagonal matrix $D_{ii} = \mu_i$ with the entries of the invariant measure μ.

So the (steady state) probability flow from state i to j is given by $F_{ij} = \mu_i P_{ij}$. If the process is reversible, the flow matrix F is symmetric due to the detailed balance equation. For a NESS process, F is not symmetric since there are states $i, j \in S$ with $F_{ij} \neq F_{ji}$.

Cycle Decomposition

This flow must be decomposable into elementary cycles.

why?

Definition 2.8 (Cycle of a process)**.** *A k-cycle γ on S is an ordered sequence (up to cyclic permutations) of k connected states $\gamma = (i_1, \dots, i_k)$ with length $|\gamma| = k$, i.e. the probability to get to the next state is always positive: $\mathbb{P}_{i_j,i_{j+1}} > 0$ and $\mathbb{P}_{i_k,i_1} > 0$. Cycles without repetition/self-intersections are called simple cycles. The set of all simple cycles is denoted by $\mathcal{C}$.* 1-step prob.?

We want to make a cycle decomposition of the flow F, see Kalpazidou[**?**].

Definition 2.9 (Cycle/flow Decomposition)**.** *A collection $\mathcal{C}_+ \subset \mathcal{C}$ of cycles γ with real positive weights $w(\gamma)$ is a flow decomposition if for every edge $(i,j) \in S^2$ we have*

$$F_{ij} = \sum_{\gamma \supset (i,j)} w(\gamma),$$

where $(i,j) \subset \gamma$ if the edge (i,j) is in γ.

In order to make sense in a probabilistic context, we define the weight w of a cycle γ in the following way. Given a (realization of?) Markov chain $(X_t)_{t\in\mathbb{T}}$, we count the number of times N_T^γ the process passes through a cycle γ up to time T.

Definition 2.10 (Weight of a Cycle)**.**

$$w(\gamma) = \lim_{T\to\infty} \frac{N_T^\gamma}{T}.$$

Since we are considering/assuming an ergodic process, this limit exists a.s. jian qian

Dominant cycles/sets

We will see that dominant cycles have similar properties as dominant sets for reversible processes, i.e. large eigenvalues with $|\lambda| \approx 1$. But now the eigenvalues are lying in the complex plane and might be non-real (pairs of complex eigenv.).

Definition 2.11 (Dominant Cycle)**.** *= metastable cycle?*

So a cycle is dominant if there is a high probability inside the Markov chain to follow this cycle.

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 leading Schur vectors (spectral clustering with PCCA+).

Schur Decomposition

We have the same situation/aim as in the previous sections: we have a Markov process on a large state space S and we want to decompose it into a smaller state space consisting of clusters that belong to dominant structures of the process.

Definition 2.12 (Schur Decomposition)**.** *Let $P \in \mathbb{R}^{n\times n}$ be a transition matrix. Then it can be written as*

$$P = XRX^{-1},$$

where X is a unitary matrix and U is an upper triangular matrix, which is called a Schur form/matrix/decomposition of P.

Since R is similar to P, both matrices have the same spectrum. Since R is triangular, their eigenvalues are the diagonal entries of R.

A Schur Decomposition is not unique. As P is a real matrix, its non-real eigenvalues come in complex conjugate pairs. This fact can be used to build a *real Schur form*, where X and R are both real matrices. But then R is no longer triagonal, but only *quasi-triangular*, allowing 2×2-blocks on its diagonal. The eigenvalues of the 2×2-blocks are exactly the complex conjugate eigenpairs of P. ... picture see ..

Definition 2.13 (Schur Vector)**.** *Let $\widetilde{R} \in \mathbb{R}^{m\times m}$ be a submatrix of R (top left part of R). Then*

$$P = \widetilde{X}\widetilde{R}\widetilde{X}^{-1},$$

where $\widetilde{X} \in \mathbb{R}^{n\times m}$ consists of the first m columns of X. These vectors will be denoted as the dominant Schur Vectors of P. Schur Values? why?

Using PCCA+

Djurdjevac Conrad et al[**?**] propose an algorithm in order to determine/get the desired membership vectors χ.

i) Compute a real Schur decomposition $(\widetilde{X}, R)$ of ...

ii) Sort the Schur values and the 2×2-blocks such that they are in a descending order

iii) Determine the submatrix $\widehat{X}$ and solve PCCA+ equation (...) in order to get the membership functions χ

Computation of metastable cycles/sets

I. Beckenbach, L. Eifler, K. Fackeldey, A. Gleixner, A. Grever, M. Weber, J. Witzig: Mixed-Integer Programming for Cycle Detection in Non-reversible Markov Processes(2017)

K. Fackeldey, M. Weber: GenPCCA – Markov State Models for Non-Equilibrium Steady States. WIAS Report, 29:70-80, 2017.

M. Weber and K. Fackeldey. G-PCCA: Spectral clustering for non- reversible markov chains. ZIB-Report 15-35, Zuse Institute Berlin, 2015

3 Rebinding Effect in a Given Kinetics

In this chapter we are going to examine a special type of molecular dynamic systems, namely receptor-ligand systems. To describe these systems rigorously we can use all the mathematical concepts/objects defined in the previous chapters.

To give a short overview about what is going to happen here. The kinetics of 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 (section **??**). This operator will be projected onto a finite-dimensional state space (section **??**) which may spoil the Markov Property of the process (section **??**). This chapter is mainly based on Weber and Fackeldey[**?**].

nop

optimization

TODO: In which model are we working? Hamiltonian (Γ)? Hamiltonian w/ randomized momentum (Ω w/ any momentum)? Langevin? Diffusion Dynamics? TODO: Ensemble, Conformation Space

3.1 Receptor-Ligand System

We will present the *receptor-ligand-system* as particular molecular dynamic system and describe it mathematically using a differential equation. We will discuss the so called *Rebinding Effect* and set in in relation to the Recrossing effect known from section **??**. As an outlook/motivation, we will explain how these two concepts can be set together in the important application of *drug design* and how the rebinding effect can help to improve the effect/efficiency of drugs.

What is a MD system? Molecular Dynamics vs Kinetics

Weber p.10

In the previous chapters, we have often mentioned molecular dynamic systems (MD systems) and some of their properties without actually explaining what such a system is. A *molecular system* consists of atoms that are connected by *covalent bonds*. bla bla. The potential energy function, or *energy landscape*, of a molecular system results in a dynamical behaviour on different timescales. The fastest timescales (vibrations of covalent bonds) are around 10^{-15} seconds. bla. up to nanoseconds or,

see ..

for protein folding, up to microseconds or seconds or even longer..

What is a Receptor-Ligand system?

A receptor can be a protein or blabla. As we are mainly interested in the mathematical description of a system, we will not specify further, which kind of biochemic/molecular receptor we are considering. Instead for us, a receptor will just be an *object* to which a ligand can bind.

A *receptor* is (in general) a proteine molecule. A molecule that binds to a receptor is called a *ligand*. Each receptor will only bind with ligands of a particular structure.

Ligand binding is an (chemical) equilibrium process, i.e. the reaction rates of the forward and backward reactions are equal. That means that the concentrations of the reactants and the products are constant (*dynamic equilibrium*). see ..

A ligand (L) can bind to a receptor (R) and form a receptor-ligand complex (LR) which can dissolve again into its two original components. These reactions can be represented in the following form

$$L + R \rightleftharpoons LR,$$

which corresponds to the law of mass action. The dissociation constant k_d see ..

$$k_d = \frac{[L] \cdot [R]}{[LR]},$$

where $[L], [R]$ and $[LR]$ are the concentrations of $(L), (R)$ and (LR), respectively. This constant is commonly used to describe the affinity between a ligand (L) and a protein (P), i.e. how strongly/tightly the ligand can bind to his particular protein/receptor. If the dissociation constant is small, then there is a high binding affinity between the ligand and the receptor. The association constant k_a is just the inverse of the dissociation constant

$$k_a = \frac{[LR]}{[L] \cdot [R]}.$$

There are different factors which can lead to a high or low *binding affinity*. which?

Bivalent Ligand

Bivalent ligands consists of two (drug-like?) molecules connected by an (inert?) linker.

Mathematical Description of Receptor-Ligand-System

Starting from the reaction equation (**??**), we can deduce that the ligand can be found in two different (macro) states: "unbound" (L) or "bound" (LR). Then the probabilities of the ligand to be in one of these states can be described by the probability vector $x^T = \frac{1}{s}(([L],[LR]))$, where $s = [L] + [LR] = \text{const.}$. This leads to an ordinary differential equation

$$\dot{x}^T = x^T Q_c.$$

The matrix Q_c consists of the rates of reaction,

$$Q_c = \begin{pmatrix} -k_a[R] & k_a[R] \\ k_d & -k_d \end{pmatrix},$$

where k_a and k_d are the association and dissociation constants. Thus, it is the transition rate matrix corresponding to a Markov chain, i.e. it describes a memoryless process.

We will later see that this mathematical description of a receptor-ligand-system is not accurate, since in fact, such a process *will* have some kind of memory.

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

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 from a chemical and a mathematical point of view. In chemistry, there are several reasons/factors for the rebinding effect discussed.

Relation to Recrossing Effect

We remember that we described the recrossing effect in section **??**. There, we also had a process on macro states described by a transition matrix and thus being a Markov chain. But in reality, the (clustered) process was not memoryless. The same phenomena occurs with the rebinding effect. We have a transition matrix, even though our process has a memory. We want to quantify this effect.

Basics of Drug Design

The term *drug design* describes the development of new medications based on the knowledge of a biological target. The drug is often a small molecule which can bind to a protein molecule (target/disease/receptor) and thus activates or inhibits its function (disease modifying). So drug design is basically about designing a molecule which is complementary in shape and charge to the biomolecular target and therefore will bind to it, see Strømgaard et al[**?**]. More precisely, drug design describes the design of ligands, i.e. molecules that will bind tightly to the given target, see Tollenaere[**?**].

In order to be an efficient drug, we aim/wish for a a high *binding affinity*, which is a measure of the strength of the chemical bond. That means that the designed ligands should easily bind to the receptors, remain in a binding or rebind quickly after being dissolved. If the binding affinity of a drug is too low, a higher concentration of the drug is needed instead, which is undesired because of possible side effects. There are many factors that influence/affect the binding affinity of a drug/ligand, such as ... . The rebinding effect has been recently investigated to increase the binding affinity of a ligand, mathematically described by Weber et al[**? ?**] as well as chemically by e.g. Vauqelin[**?**]. This effect will be examined in this thesis. As a *high* rebinding effect is aimed/wished, we will derive a lower bound for this effect, i.e. we will *minimize* it.

falsch?

Drug Design

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. Thus, 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 and Fackeldey[**?**] for reversible processes.

How: from bound to unbound

3.2 Molecular Kinetics as a Projection

Mol. Kinetics Weber p.10

In this section, we will basically embed the mathematical concepts/results of chapter **??** into a chemical/physical context in order to get a rigorous description of molecular (dynamic/kinetic?) systems. In considering such systems, we can distinguish between two point of views: we will see how we can get from the *atomistic (=microsopic)* to a *macroscopic* scale/point of view by a projection.

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* $\mathbb{R}^{3N}$. In the following, we consider systems in *thermodynamical equilibrium*. One possible model is given by the *Boltzmann distribution* $\pi : \Omega \times \mathbb{R}^{3N} \to \mathbb{R}$, a probability distribution assigning to each micro state a probability depending on its energy and temperature, see McQuarrie[**?**]. It can be expressed as

spatial/position space

prob. dens. fct.?

$$\pi(q,p) = \frac{1}{Z} \exp\left(-\beta H(q,p)\right), \tag{3.1}$$

where $\beta = 1/(k_B T)$ is the inverse of the temperature T multiplied with the Boltzmann constant k_B and $Z = \int_\Gamma \exp -\beta H(q,p)$ is the normalization factor. The Hamilton function denoted by H is given by $H(q,p) = K(p) + V(q)$, the sum of the kinetic energy $K(p)$ and the potential energy $V(q)$. Thus, the Boltzmann distribution π can be decomposed into $\pi = \pi_p \pi_q$,

d?

$$\pi(q,p) = \underbrace{\frac{1}{Z_p} \exp\left(-\beta K(p)\right)}_{\pi_p} \cdot \underbrace{\frac{1}{Z_q} \exp\left(-\beta V(q)\right)}_{\pi_q},$$

where $\pi_p : \mathbb{R}^{3N} \to \mathbb{R}$ is the probability density function of the kinetic part in the momentum space $\mathbb{R}^{3N}$ and $\pi_q : \Omega \to \mathbb{R}$ is the probability density function of the potential part in the configurational space Ω.

As we are interested in examining conformations/metastable sets, which are objects in configurational space, we will restrict ourselves to Ω:

> "A conformation $C \subset \Omega$ will be identified with the particular metastable sub-ensemble $\mu_{C \times \mathbb{R}^{3N}}$ corresponding to the particular subset $C \times \mathbb{R}^{3N} \subset \Gamma$. Hence, for every position $q \in C$, the conformation contains all states with $q \in \Omega$ and arbitrary $p \in \mathbb{R}^{3N}$."

In this sense, conformations/metastable sets contain no information on momenta and are determined in configurational space only. We are considering a reduced model in position space with a *reduced density* $\pi_q = \int_{\mathbb{R}^{3N}} \pi(q,p)\, \mathrm{d}p$.

?

TODO: difference conformation vs. metastable set

Macro States via Membership Functions

As the phase space and even the configurational space are very large, we aim to reveal the underlying discrete Markov State Model by group/cluster a collection of the micro states having the same or similar values in one observable. Such a

collection of micro states will be called a *macro state*. For instance, that could be the states/observables "bound" or "unbound" for a receptor-ligand system.

entropic inf.?

We apply the function-based clustering method presented in section **??**. We define macro states as overlapping partial densities, which can be identified as membership functions $\chi_1, \ldots, \chi_n$. The membership functions $\chi_1, \ldots, \chi_n : \Omega \to [0,1]$ form a partition of unity, i.e.

overlap = good?

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

By grouping micro states, the (corresponding) macro states yield *statistical weights*

what is that good for?

$$w_i = \langle \chi_i, \mathbb{1} \rangle_\pi = \int_\Omega \chi_i(q) \pi_q(q) \, \mathrm{d}q.$$

e vs. $\mathbb{1}$

The statistical weight w_i corresponds to the "probability to be in conformation χ_i".

Transfer Operator

Each micro state $(q,p) \in \Gamma$ determines a *probability density function* $\Psi^{-\tau}(\cdot \mid (q,p))$ describing the possible evolutions of the system in configurational space Ω in time τ, being started at the initial state (q,p). Weber[**?**] defines a transfer operator $\mathcal{P}(\tau) : L^{1,2}_{\pi_q}(\Omega) \to L^{1,2}_{\pi_q}(\Omega)$ for the propagation of (membership) functions via

why not densities?

$$\mathcal{P}(\tau) f(q) = \int_{\mathbb{R}^{3N}} \left(\int_\Omega f(\tilde{q}) \Psi^{-\tau}(\tilde{q} \mid (q,p)) \, \mathrm{d}\tilde{q} \right) \pi_p(p) \, \mathrm{d}p. \tag{3.3}$$

In this definition, the density function $\Psi^{-\tau}(\cdot \mid (q,p))$ can be interpreted as a transition function as defined in section **??**. We have to notice that this transfer operator corresponds to the *backward operator* from section **??**.

?

It is a *generalized* transfer operator in the sense that it includes deterministic as well as stochastic dynamical models. In order to describe deterministic dynamics, the density function $\Psi^{-\tau}$ has to be chosen as a Dirac delta function, since an initial state $(q(0), p(0))$ determines exactly the future states in configurational space.

It is important to remark that the transfer operator $\mathcal{P}(\tau)$ also defines a projected *Markov operator* $\overline{\mathcal{P}}(\tau)$ acting in configurational space Ω, see Weber[**?**], by

propagator sec **??** def

$$\overline{\mathcal{P}}(\tau) = \pi_q \circ \mathcal{P}(\tau) \circ (\pi_q)^{-1}, \tag{3.4}$$

which propagates density functions. The previous equation shows that the space of membership functions is connected to the space of density functions by multiplication with π_q. We will keep that relation in mind, but just use $\mathcal{P}$ in the following.

As $\mathcal{P}(\tau)$ in (**??**) propagates **membership functions**, stationarity is characterized by the equation $e = \mathcal{P}(\tau)e$ for the constant function $e = 1$ in Ω. For the Markov operator $\overline{\mathcal{P}}(\tau)$ in (**??**) propagating **densities**, stationarity can be characterized by $\pi = \overline{\mathcal{P}}(\tau)\pi$, where π is the Boltzmann density. These two operators are *adjoint* operators. This can also be seen by the fact that a discretization of $\mathcal{P}(\tau)$ results in a matrix P_c, while a discretization of $\overline{\mathcal{P}}(\tau)$ will result in the transposed matrix P_c^T.

Boltzmann dens. = dist.?

Maybe: Properties of transfer operator for reversible Processes

Detailed Balance

$$\pi_q(\tilde{q}) \cdot \int_{\mathbb{R}^{3N}} \Psi^{-r}(q \mid (\tilde{q}, p))\pi_p(p)\,\mathrm{d}p = \pi_q(q) \cdot \int_{\mathbb{R}^{3N}} \Psi^{-r}(\tilde{q} \mid (q, p))\pi_p(p)\,\mathrm{d}p \qquad (3.5)$$

Markov State Model for reversible Processes

For now, we consider a reversible process. Then due to the detailed balance condition (**??**), the corresponding transfer operator $\mathcal{P}$ is **self-adjoint** and thus has a real spectrum, see theorem **??** (follows from linearity and self-adjointness) and $\sigma(\mathcal{P}) \subset [-1, 1]$ (since $\|\mathcal{P}f\|_{\pi_q} \leq \|f\|_{\pi_q}$). In order to apply the spectral approach from section **??**, we assume that the **discrete spectrum** of the transfer operator $\mathcal{P}$ has n **dominant eigenvalues** $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ which are all close to 1 and bounded away from the essential spectrum. The corresponding dominant eigenfunctions are denoted by $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$ and therefore the eigenvalue problem is $\mathcal{P}(\tau)\mathcal{X} = \mathcal{X}\Lambda$, with the eigenvalue matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

self-adj. wrt π_q

As we have seen in chapter **??**, the number of metastable sets of a process can be determined by the number of dominant eigenvalues; i.e. we are going to create a 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 $P(\tau)$. In order to get from our continuous operator $\mathcal{P}(\tau)$ to a discrete matrix $P(\tau)$, we need at first to determine the size and shape of the membership functions χ_i. As described in section **??**, this can be done by computing a linear combination of the dominant eigenfunctions via

?

$$\chi_j(q) = \sum_{i=1}^{N} A_{ij}\mathcal{X}_i(q),\ j = 1, \ldots, n, \qquad (3.6)$$

where $A = \{A_{ij}\}_{i,j=1,\ldots,n}$ 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 χ_i might have an overlap; they are not orthogonal!

PCCA+ only for finite/ discrete state spaces?
Ref?

Galerkin Projection

Having computed the membership functions χ_i, we can project $\mathcal{P}(\tau)$ to a low-dimensional Markov State Model $P_c(\tau)$ by the Galerkin discretization

$$P_c(\tau) = G(\mathcal{P}(\tau)) = (\langle \chi, \chi \rangle_\pi)^{-1}(\langle \chi, \mathcal{P}(\tau)\chi \rangle_\pi). \tag{3.7}$$

We are interested in the iteration error under this projection. As mentioned in section **??**, this error is zero if the Galerkin discretization of $(\mathcal{P}(\tau))^k$ is equal to the iteration $(P_c(\tau))^k$. In that case, the diagram **??** commutes. The following theorem shows that there is no discretization error under the projection (**??**), i.e. we have $(\mathcal{P}(\tau))^k = (P(\tau))^k$, which implies that Markovianity is preserved. ?

Theorem 3.1 (Weber [**?** , Theorem 2])**.** *Let $\mathcal{P}(\tau)$ be the π_q-self-adjoint transfer operator defined in* (**??**) *with a set $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$ of normalized eigenfunctions s.t. $\mathcal{P}(\tau)\mathcal{X} = \mathcal{X}\Lambda$ and a set of functions $\chi = \mathcal{X}A$ that is a linear combination of the eigenfunctions $\mathcal{X}$ with a regular $n \times n$-transformation matrix A from* (**??**)*. Then the iteration error for the Galerkin discretization $P_c(\tau) = G(\mathcal{P}(\tau))$ in* (**??**) *vanishes.*

Proof. □

It follows that the above projection represents the correct dynamical long-time behaviour of the original process and that the matrix $P_c(\tau)$ is the correct Markov State Model. We can use the matrix representation $P_c(\tau) = S^{-1}T$ from theorem **??**. Then S and T are stochastic matrices with ?

$$\begin{aligned} T &= D^{-1}\langle \chi, \mathcal{P}(\tau)\chi \rangle_\pi = D^{-1}A^T\Lambda A \quad \text{and} \\ S &= D^{-1}\langle \chi, \chi \rangle_\pi = D^{-1}A^T A, \end{aligned}$$

where $D = \mathrm{diag}(w_1, \ldots, w_n)$ is the diagonal matrix of statistical weights in (**??**).

Measuring the Rebinding Effect

Interpretation

Even though $P_c(\tau) := P(\tau)$ is the correct Markov State Model, it cannot be interpreted as a transition matrix, since the inverse matrix of S is not necessarily stochastic. The matrix $T = D^{-1}\langle \chi, \mathcal{P}\chi \rangle$ however can be interpreted as a transition matrix. Then the difference between $P_c(\tau)$ and T is given by

$$SP_c(\tau) = T.$$

Thus, the "disturbance" of ... can be measured by the matrix S. The more the matrix S differs from the identity matrix, the more the correct projection $P(\tau)$

differs from the transition matrix T. Thus, the rebinding effect can be measured by the matrix S. The trace of S is at most n. Optimizing trace(S) is equivalent to optimzing the *crispness* of the conformations χ (Röblitz).

Infinitesimal Generator to transition rate matrix

Often (...) it is more convenient to consider/examine/investigate transition rate matrices instead of transition matrices/ infinitesimal generators instead of transfer operators. We can define the same/similar/analogous Galerkin Projection on the corresponding infinitesimal generator.

Conceptually, $\mathcal{Q}$ is connected to the computation of transition rates.

The transfer operator $\mathcal{P}(\tau)$ defines a time-independent operator $\mathcal{Q}$ via

$$\mathcal{Q} = \lim_{\tau \to 0} \frac{\mathcal{P}(\tau) - \mathcal{I}}{\tau},$$

which is the infinitesimal generator of $\mathcal{P}$: Chapman

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

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

Since the eigenfunctions of $\mathcal{Q}$ and $\mathcal{P}$ are the same and their eigenvalues are related via $\exp(\xi_i) = \lambda_i$, we can apply the same Galerkin Projection for the infinitesimal generator as for the transfer operator in (**??**). We get a $n \times n$-rate matrix

$$Q_c = A^{-1} \Xi A = (\langle \chi, \chi \rangle_\pi)^{-1} (\langle \chi, \mathcal{Q}\chi \rangle_\pi), \tag{3.8}$$

where Ξ is the diagonal matrix consisting of the n leading eigenvalues $0 = \xi_1 > \xi_2 \geq \cdots \geq \xi_n$ of $\mathcal{Q}$ and A is the transformation matrix of (**??**), which analoguously transforms the eigenfunctions of $\mathcal{Q}$ into membership functions of the macro states.

The matrix Q_c can be interpreted as a transition rate matrix. ?

3.3 Minimizing the Rebinding Effect

As the computations of eigenfunctions of a continuous operator $\mathcal{Q}$ is an extensive task, we assume that the transition rates Q_c can be measured experimentally.

As mentioned before, the rebinding effect is a desired/favorable property for some processes, e.g. in drug design, where high rebinding increases the efficiency of a drug. In order to get high rebinding effects, we will now derive a lower bound for it; i.e. minimize the rebinding effect.

Stability of the system/process in terms of determinants

If the eigenvalues ξ_i of Q_c are close to 0, then the macro states are very stable in the sense that the probability to stay inside of that state is high/close to 1. The trace of Q_c corresponds to the sum of the dominant eigenvalues of $\mathcal{Q}$. Thus, we can measure the stability of the molecular system by considering $F := -\text{trace}(Q_c)$. If F is high, then the process is fast and less stable. If F is close to 0, then the process is slow and very stable. We want to set the indicator for stability F in relation to the matrices S and T from theorem **??** (matrix representation of Galerkin projection).

Theorem 3.2. *If Q_c is a projected infinitesimal generator of a process and $P_c(\tau)$ the corresponding projected transfer operator with a matrix representation $P_c(\tau) = S^{-1}T$ from theorem **??**, then the stability $F := -\text{trace}(Q_c)$ can be measured by*

$$F = \tau^{-1}(\log(\det(S)) - \log(\det(T))).$$

Proof. We use the relation $\exp(\text{trace}()) = \det(\exp())$, the fact that Q_c "generates" $P_c(\tau)$ and theorem **??** to see that

see .. why $exp(\tau Q_c) = P_c(\tau)$?

$$\begin{aligned} F &= -\text{trace}(Q_c) \\ &= -\tau^{-1}\log(\exp(\text{trace}(\tau Q_c))) \\ &= -\tau^{-1}\log(\det(\exp(\tau Q_c))) \\ &= -\tau^{-1}\log(\det(P_c(\tau))) \\ &= \tau^{-1}(\log(\det(S)) - \log(\det(T))). \end{aligned}$$

□

Thus, both determinants of the stochastic matrices S and T influence the stability of the system, but in converse directions.

If $\det(T)$ is close to 1 (unit matrix), then the process is stable/slow ($F = \log\det(S) + 0$ is low). If $\det(T)$ is close to 0 (much overlap), then the process is unstable/fast ($F = \log\det(S) + \infty$ is high).

If $\det(S)$ is close to 1 (almost unit matrix), then $F = 0 + \log\det T$ is high, i.e. the process is unstable/fast. If $\det(S)$ is close to 0, then $F = -\infty + \log\det T$ is low, i.e. the process is stable/slow. That means that a higher overlap in S leads to a slower process!

As we figured out in section **??**, the rebinding effect can be measured by the matrix S. We can deduce now, that a system with strong rebinding, i.e. S deviating from the unity matrix/ having small determinant, is more stable.

Optimization Problem

What is the meaning of the previous results/relation of S to stability of the system?

If we have a determinant $\det(S) = 1$, we have no rebinding. The smaller this determinant, the higher the rebinding effect. As we are interested in a high/increased rebinding effect, it would be nice to get a small determinant of S.

But how is S determined? We were given a transfer operator $\mathcal{P}$ which was projected onto a finite-dimensional state space using membership functions χ_i. These membership functions have been computed as a linear combination of the eigenfunctions with a regular matrix A. In summary, the choice of this matrix A determines S respectively the size of its determinant.

So far, A was assumed to be computed via PCCA+, i.e. such that it results in a optimal metastable decomposition/clustering. We want to know if there are possible choices of A resulting in a "better", i.e. higher, rebinding effect.

Task: find out which choice of A results in the "best", i.e. highest, rebinding effect, measured by S_{opt}. This problem is equivalent to finding the smallest possible determinant of S.

Optimization Problem (Maximizing determinant of S)

The eigenvalue problem of Q_c is given by

$$Q_c X = X\Xi,$$

where the first column of X corresponds to the first eigenvector $X_1 := (1, \dots, 1)^T$. By (**??**), we see that A^{-1} is an eigenvector matrix of Q_c as well. The columns of the matrix A^{-1} consists of multiples of the eigenvectors X_i. So we have

$$A^{-1} = \begin{pmatrix} 1 & & & \\ \vdots & \alpha_2 X_2 & \cdots & \alpha_3 X_3 \\ 1 & & & \end{pmatrix}$$

with $\alpha_2, \dots, \alpha_n \in \mathbb{R}$. We know from theorem **??** that a $\det(S)$ close to 1 results in a low rebinding effect. Thus, in order to find a lower bound for the rebinding effect, we try to maximize $\det(S)$/ minimize $|\det(S) - 1|$. Our minimization/optimization problem is then given by

$$\boxed{\min_{\alpha_1,\dots,\alpha_n \in \mathbb{R}} |\det(S) - 1|}, \tag{3.9}$$

where we have to include several side constraints. As the inverse matrix A^{-1} consists of linear combinations of eigenvectors X_i, we have to consider

$$\boxed{\alpha_1 = 1 \quad \text{and} \quad A^{-1}_{ij} = \alpha_i X_{ij} \quad \forall i, j}.$$

Furthermore, S is a stochastic matrix, see theorem **??**, and its structure is given in terms of the linear transformation matrix A, so we have two further constraints

$$\boxed{S = D^{-1}A^TA \quad \text{and} \quad S_{ij} \geq 0 \ \ \forall i,j}.$$

Interpretation

The result of the optimization problem (**??**) is an *optimal overlap matrix* S_{opt} with $\det(S_{\text{real}}) \leq \det(S_{\text{opt}}) \leq 1$ for any *real overlap matrix* S_{real}. In our following computations, S_{real} will be the solution we get from PCCA+ ($S_{\text{real}} = D^{-1}A^TA$). The real ocurring rebinding effect, measured by S_{real}, is high if the determinant of S_{real} is low. Thus, a small determinant of S_{opt} increases the rebinding effect. ?

Unfortunately, for a reversible Q_c, the solution of optimization problem (**??**) gives us no information, as the following theorem shows.

Theorem 3.3 (Weber and Fackeldey[**?** , Theorem 1])**.** *Let $Q_c \in \mathbb{R}^{n\times n}$ be a reversible matrix that stems from a clustering with positive definite overlap matrix S. Then there exists a matrix $A \in \mathbb{R}^{n\times n}$ in optimization problem* (**??**) *such that* $\det(S_{\text{opt}}) = \det(D^{-1}A^TA) = 1$.

Optimization Problem (Maximizing trace of S)

Instead of maximizing $\det(S)$, we can also maximize $\text{trace}(S)$.

... gives an optimal matrix $S_{\text{opt}} = D^{-1}YBY^T$

Determinant of stochastic matrix

Conclusion

A nontrivial rebinding effect can be estimated only if the kinetics Q_c of a system is nonreversible. why?

3.4 Approach for nonreversible processes

With the tools from section **??** (Schur Decomposition and G-PCCA+) 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[**?**](2016).

Transfer Operator

Have the transfer operator $\mathcal{P}$ from (**??**) given, but from theorem **??** we know that the transfer operator of a nonreversible process is not self-adjoint.

Schur Decomposition

Applying a Schur Decomposition, we can create real eigenvalues from the possibly nonreal eigenvalues of the transfer operator.

detecting dominant cycles of the process? G-PCCA+?

Galerkin Projection

Now that we have real eigenvalues, we can apply the Galerkin Projection as usual.

4 Illustrative Examples

We want to apply the results from chapter **??** on some easy examples.

4.1 A Transition Network Graph

4.2 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] L.-T. Da, F. K. Sheong, D.-A. Silva, and X. Huang, *Application of markov state models to simulate long timescale dynamics of biological macromolecules*, in Protein Conformational Dynamics, Springer, 2014, pp. 29–66.

[2] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled markov chains*, Linear Algebra and its Applications, 315 (2000), pp. 39–59.

[3] P. Deuflhard and M. Weber, *Robust perron cluster analysis in conformation dynamics*, Linear algebra and its applications, 398 (2005), pp. 161–184.

[4] N. Djurdjevac Conrad, M. Weber, and C. Schütte, *Finding dominant structures of nonreversible markov processes*, Multiscale Modeling & Simulation, 14 (2016), pp. 1319–1340.

[5] G. H. Golub and C. F. van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore and London, 1996.

[6] W. Huisinga, *Metastability of markovian systems*, Ph.D. thesis, Freie Universität Berlin, (2001).

[7] S. L. Kalpazidou, *Cycle representations of Markov processes*, vol. 28, Springer Science & Business Media, 2007.

[8] T. Kato, *Perturbation Theory for Linear Operators*, Classics in Mathematics. Springer, 1995.

[9] D. A. McQuarrie, *Statistical Mechanics*, University Science Books, California, 2000.

[10] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series. Springer, 1993.

[11] A. Nielsen, *Computation schemes for transfer operators*, Ph.D. thesis, Freie Universität Berlin, (2015).

[12] S. Röblitz and M. Weber, *Fuzzy spectral clustering by pcca+: application to markov state models and data classification*, Advances in Data Analysis and Classification, 7 (2013), pp. 147–179.

[13] M. Sarich, *Projected transfer operators*, PhD thesis, Freie Universität Berlin, 2011.

[14] M. Sarich, R. Banisch, C. Hartmann, and C. Schütte, *Markov state models for rare events in molecular dynamics*, Entropy, 16 (2013), pp. 258–286.

[15] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *A direct approach to conformational dynamics based on hybrid monte carlo*, Journal of Computational Physics, 151 (1999), pp. 146–168.

[16] 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.

[17] K. Strømgaard, P. Krogsgaard-Larsen, and U. Madsen, *Textbook of Drug Design and Discovery*, CRC Press, 2002.

[18] J. Tollenaere, *The role of structure-based ligand design and molecular modelling in drug discovery*, Pharmacy World and Science, 18 (1996), pp. 56–62.

[19] 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.

[20] M. Weber, *Meshless methods in conformation dynamics*, Ph.D. thesis, Freie Universität Berlin, (2006).

[21] M. Weber, *A subspace approach to molecular markov state models via a new infinitesimal generator*, Habilitation thesis, Freie Universität Berlin, (2011).

[22] 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.

[23] M. Weber and K. Fackeldey, *Computing the minimal rebinding effect included in a given kinetics*, Multiscale Modeling & Simulation, 12 (2014), pp. 318–334.