

Master thesis

# **Computing the minimal rebinding effect for nonreversible processes**

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

Susanne Röhl

Berlin, den 27. Februar 2017

# Contents

<b>1</b>	<b>Markov State Models</b>	<b>3</b>
1.1	Markov Process . . . . .	3
1.2	Transfer Operator . . . . .	6
1.3	Galerkin Projection . . . . .	9
1.4	Recrossing Effect . . . . .	15
<b>2</b>	<b>Dominant Structures</b>	<b>20</b>
2.1	Metastability . . . . .	20
2.2	Spectral Approach . . . . .	23
2.3	Fuzzy Clustering . . . . .	25
2.4	Dominant Cycles . . . . .	28
<b>3</b>	<b>Rebinding Effect in a Given Kinetics</b>	<b>31</b>
3.1	Receptor-Ligand System . . . . .	31
3.2	Rebinding Effect . . . . .	31
3.3	Molecular Kinetics as a Projection . . . . .	32
3.4	Minimizing the Rebinding Effect . . . . .	36
3.5	Approach for nonreversible processes . . . . .	36
<b>4</b>	<b>Illustrative Examples</b>	<b>37</b>
4.1	An artificial (bivalent) binding Process . . . . .	37
	<b>Bibliography</b>	<b>ii</b>

# 1 Markov State Models

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

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

## 1.1 Markov Process

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

### Transition function

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

X with Borel-sigma-alg.?

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

in  $\Omega$ :

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

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

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

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

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

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

assigned to a  
process?

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

?

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

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

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

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

why? how?

## Markov Process

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

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

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

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

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

Indeed, there is a one-to-one relation between transition functions and (time-homogeneous) Markov processes, i.e. every homogeneous Markov process defines a transition function and vice versa, see Meyn and Tweedie[9, chapter 3].

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

## Invariant Measure

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

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

If an invariant measure is additionally a probability measure, then it is also a stationary measure. 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. def.

## Ergodicity

The long-time behaviour of stochastic processes can be described using ergodicity. difference  $\mu$ ,

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

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

for almost all initial values  $X_0 = x_0$ .

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

## Reversibility

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

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

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

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

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

## 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  $\mu$ . 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  $\mu^T = \mu^T T$ .

## 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.

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

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

## $L^r$ -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.

**Definition 1.7** ( $L^r$ -Spaces).

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

for  $1 \leq r < \infty$  and

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

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

complex  
needed?

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

Now let  $\nu_0$  be the density function corresponding to the given start distribution (function). Then the density function of a subset  $A \subset \mathbb{X}$  at time  $t$  is given by

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

On the other hand, the density  $\nu_t$  is given by

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

## Transfer Operator and spectral properties

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

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

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

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

backwards  
operator  
important?

The transfer operator is well-defined, see [5]. We will announce here already some properties of the transfer operator which will be useful in the following chapter(s).  $\mathcal{P}^t$  is a Markov operator; i.e. it conserves the norm,  $\|\mathcal{P}^t \nu\|_1 = \|\nu\|_1$ , and is positive,  $\mathcal{P}^t \nu \geq 0$  for  $\nu \geq 0$ .  $\mathcal{P}^t \nu_0$  describes the transport of the function  $\nu_0$  in time  $t$  by the underlying dynamics given by the process  $X_t$  and weighted with respect to  $\mu$ :

Markov operator?

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

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

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

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

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

why only in L2?  
stationary measure  $\pi$

### Transfer operator of reversible processes

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

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

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

*if and only if  $X_t$  is reversible.*

*Proof.* Huisinga[5] □

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

*Proof.* □

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

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



## Infinitesimal Generator

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

proof

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

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

time-indep.?

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

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

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

The infinitesimal generator is an operator which describes the behaviour of a Markov process in infinitesimal time. That becomes clear by the relation  $\mathcal{P}^t = \exp(t\mathcal{Q})$  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  $1 = \lambda_1, \dots, \lambda_m$  of the propagator  $\mathcal{P}^t$  are related to the eigenvalues  $0 = \Lambda_1, \dots, \Lambda_m$  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  $\mathcal{P}^t$  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 choosing 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 1.12** (Partition of Unity). A family of measurable functions  $\{\chi_1, \dots, \chi_n\} \subset L^2(\mu)$  is called a partition of unity if the following two conditions are fulfilled: why L2?

i) The  $\chi_i$  are non-negative and linear independent

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

**Definition 1.13** (Galerkin Projection). Let  $\{\chi_1, \dots, \chi_n\}$  be a partition of unity,  $D = \text{span}\{\chi_1, \dots, \chi_n\}$  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  $G : L^2(\mu) \rightarrow D$

$$G\nu = \sum_{k,j=1}^n \hat{S}^{-1}(k, j) \langle \chi_k, \nu \rangle_\mu \chi_j. \quad (1.14)$$

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 (3.6) 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  $\chi_i$  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 1.14** (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  $\mu$ ,  $\{\chi_1, \dots, \chi_n\}$  be a partition of unity and  $G$  the Galerkin projection onto the associated subspace  $D$ . Then an operator of the form

$$G\mathcal{P}G : L_\mu^2(\mathbb{X}) \rightarrow D$$

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

## Matrix Representation

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  $G\mathcal{P}|_D : D \rightarrow D$ .

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.

left or right?  
def. GPG  
not  $n \times n$ , but  
 $\infty \times n$ ?  
GPG same  
as  $GP_D$

**Theorem 1.15.** Let  $\mathcal{P}$  be the transfer operator of a Markov process,  $\{\chi_1, \dots, \chi_n\}$  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

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

where

$$S_{kj} = \frac{\hat{S}(k, j)}{\langle \chi_k, \mathbb{1} \rangle_\mu} = \frac{\langle \chi_k, \chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \quad (1.15)$$

and

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

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

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

it holds that

$$(\alpha_1, \dots, \alpha_n)P_c = (\beta_1, \dots, \beta_n). \quad (1.17)$$

We assume that (1.17) holds and choose a basis  $\{\psi_1, \dots, \psi_n\}$  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  $(\beta_1, \dots, \beta_n)$ . Therefor we exploit the definition of a Galerkin projection and the definition of our basis:

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

ansatz space  
vs subspace  
skillnad  
QPQ vs P

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

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

That means that

$$\begin{aligned}\beta_l &= \sum_{k,j=1}^n \alpha_k \hat{S}^{-1}(j, l) \langle \chi_l, \mathbb{1}_{\mathbb{X}} \rangle_{\mu} \langle \chi_j, \mathcal{P}\psi_k \rangle_{\mu} \\ &= \sum_{k=1}^n \alpha_k \underbrace{\sum_{j=1}^n \hat{S}^{-1}(j, l) \langle \chi_l, \mathbb{1}_{\mathbb{X}} \rangle_{\mu} \frac{\langle \chi_j, \mathcal{P}\chi_k \rangle_{\mu}}{\langle \chi_k, \mathbb{1}_{\mathbb{X}} \rangle_{\mu}}}_{P_{kl}}.\end{aligned}\quad (1.18)$$

The underbraced term is equal to  $P_{kl}$  because of (1.17). 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 (1.18).  $\square$

**Theorem 1.16.** *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:

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

$\square$

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

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

*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

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

□

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  $\mu$  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 1.15 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 1.18.
- 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

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

$$\begin{aligned} P \text{ self-adj.?} \\ \langle Av, w \rangle &= \\ \langle v, Aw \rangle \end{aligned}$$

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  $T$ .

- 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.

and so, eigenvalues of  $S$  and  $T$  are real and in  $[-1, 1]$

**Theorem 1.18.** *The matrix representation  $P$  from Thm 1.15 has the left eigenvector*

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

*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

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

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

## 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 1.19.** *Let  $\mathcal{Q} : L^2(\mu) \rightarrow L^2(\mu)$  be a generator of a semigroup of transfer operators with unique invariant measure  $\mu$  and satisfying  $\mathcal{Q}\mathbb{1}_X = 0$ . Let  $\chi$  be a partition of unity with a projection  $G$  onto the associated subspace spanned by  $\chi$ . Then the projected generator  $G\mathcal{Q}G$  has the matrix representation  $Q = RS^{-1}$  with the stochastic mass matrix  $S$  from (1.15) and*

not commu-  
tative

$$R(k, j) = \frac{\langle \chi_k, \mathcal{Q}\chi_j \rangle_\mu}{\langle \chi_k, \mathbb{1} \rangle_\mu} \quad (1.19)$$

*The eigenvalue problem of  $Q$  is equivalent to the generalized eigenvalue problem  $Ru = \Lambda Su$ . For both  $Q$  and  $R$  the largest eigenvalue is  $\lambda = 0$ . The associated right eigenvector is  $e = (1, \dots, 1)^T$ , the associated left eigenvector is  $\hat{\mu}^T$  from theorem 1.18.*

*Proof.* The matrix representation of  $G\mathcal{Q}G$  can be shown similar to the proof of theorem 1.15. For the other properties see...  $\square$

There are obviously 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  $\chi_1, \dots, \chi_n$ . In chapter 2 we are going to see which choice of  $\chi$  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 1.3, 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 1.15). 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  $(\hat{X}_k)$ . 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  $(\tilde{X}_k)$ .

fehlerfortpflanzung

A desirable behaviour of our models would be that  $(\hat{X}_k)$  and  $(\tilde{X}_k)$  have the same trajectory when started on the same initial distributions  $\hat{X}_0$  and  $\tilde{X}_0$ . 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 already seen in theorem 1.15 that the matrix representation  $P_c$  of  $G\mathcal{P}G$  is in general not a transition matrix, i.e.  $(\hat{X}_k)$  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  $\tau$ , we assign the current state of the original process  $X_t$  to the projected process  $\tilde{X}_k$ :

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

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

Let  $(\hat{X}_k)_{k \in \mathbb{N}}$  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  $(\tilde{X}_k)_{k \in \mathbb{N}}$ , there will be some differences between  $(\hat{X}_k)$  and  $(\tilde{X}_k)$ . *Markov State Model*

Notice that just in this particular example  $\mathcal{GPG}$  is a Markov chain. Normally  $TS^{-1}$  corresponding to the projected process is not 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 1.1. Now compare the two model-processes  $\tilde{X}_k$  and  $\hat{X}_k$ . We already know that  $\hat{X}_k$

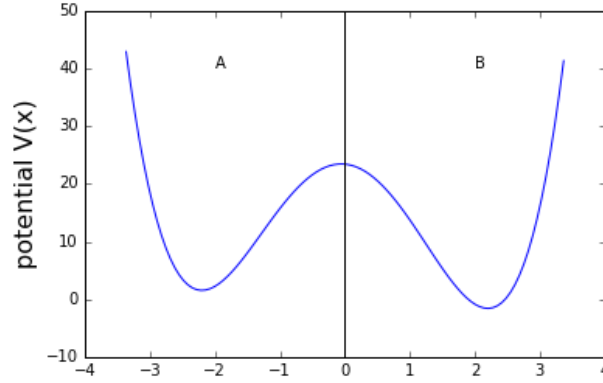


Figure 1.1: full-partition of a double-well potential

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

$\tilde{X}$ ?

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

We get

$v$  densities?

$$\mathbb{P}_\mu[\tilde{X}_{(k+1)\tau} \in A \mid \tilde{X}_{k\tau} \in B] = \int_A v_B^\tau(x) dx = \dots, \quad (1.20)$$

$$\mathbb{P}_\mu[\tilde{X}_{(k+1)\tau} \in A \mid \tilde{X}_{k\tau} \in B, \tilde{X}_{(k-1)\tau} \in A] = \int_A v_{BA}^\tau(x) dx = \dots \quad (1.21)$$

So we see that for such a short lag-time  $\tau$ , the process  $\tilde{X}_k$  is not memoryless and hence not a Markov process. That effect is intuitively clear. Equation (1.20)

describes the probability to get from  $B$  to  $A$ , so it is averaged(?) over all possible starting points in  $B$ . Comparing to that in (1.21), 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  $\hat{X}$  could be called a *short-time memory*.

Comparing that to  $\hat{X}$ .  $\hat{X}$  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  $\tau$ . 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  $\hat{X}$  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

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?

$$G\mathcal{P}G \hat{=} \begin{array}{ccc} \mathcal{P}(\tau) & \xrightarrow{\tau \rightarrow \tau k} & (\mathcal{P}(\tau))^k \\ \downarrow \text{proj.} & & \downarrow \text{proj.} \\ P_c(\tau) & \xrightarrow{\tau \rightarrow \tau k} & (P_c(\tau))^k \end{array}$$

later = re-  
binding  
+ proj.prop  
!=  
prop.proj.?  
?

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

## Discretization Error (Density Propagating Error)

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  $\tilde{X}_k$  and  $\hat{X}_k$  after  $k$  time-steps is (independently of initial distribution) given by

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

**Theorem 1.20.** *Assume the discrete/dominant spectrum of a transfer operator  $\mathcal{P}$  is given/denoted/ordered by  $1 = \lambda_0 > \lambda_1 \geq \dots \geq \lambda_n$ . Then the projection error can be bounded from above in terms of the second-largest eigenvalue by*

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

where  $\Pi_0$  is the orthogonal projection of ... .

For a proof, see Schütte and Sarich[14, 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 about  
eigenvalue error?

which norm?

what is  
markov state  
model? def!  
 $P_c$   
?

## 2 Dominant Structures

In section 1.3, 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  
facts.

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 (a decomposition into?) subsets 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 (see...) or small exit rates (...).

#### Mathematical concept of metastability

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

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

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

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

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

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

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

### Importance of metastability to MD systems

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

### Problem of long-time simulations of MD systems

As transitions between metastable sets are a rather rare event, we need to make long-time simulations of our process in order to get reasonable results about these changes of conformations. But, as mentioned in Chapter 1, long-time simulations of such complex systems are not feasible even with the best computers nowadays, see (...).

So, in order to be able to compute some long-time simulations of a given MD system, a reduction of complexity like described in section 1.3 is needed/ required. This can be achieved by a clustering of a state space. Different states of the state space will be clustered appropriately s.t. we get a process on a smaller state space.

## Clustering into metastable sets

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

Galerkin =  
clustering???

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

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

## Advantages / Disadvantages

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

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

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

## Example

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

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

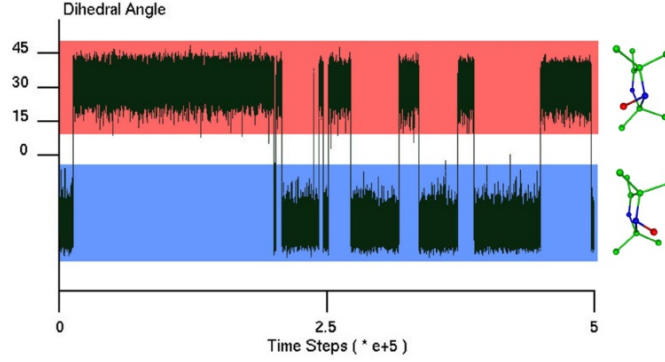


Figure 2.1: Example of a molecule with two conformations

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

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

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

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

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

- C1** The essential spectral radius of  $\mathcal{P}$  is less than one; i.e.  $r_{\text{ess}} < 1$ .
- C2** The eigenvalue  $\lambda = 1$  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[5]. Since these conditions are required for the later investigations, we will just assume that they are true.

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

**Theorem 2.1** (Schuette[14]). *The transfer operator  $\mathcal{P} : L^2 \rightarrow L^2$  of a reversible process with properties **C1** and **C2** has the following spectrum:*

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

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

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

### Relation of dominant eigenvalues to metastable sets

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

### Relation of dominant eigenfunctions to metastable decomposition

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

*Proof.*

□

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

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

### Disadvantages

So the spectral approach is suitable/ convenient to characterize metastability of Markov processes. But: there are two disadvantages. 1: the result is only applicable on reversible processes, because a self-adjoint transfer operator is needed in order

has periodic structures?  
excludes modeling and interpretation problems

C2 = ergodic?

why discr. rechts?

eigenfunctions vs committor functions



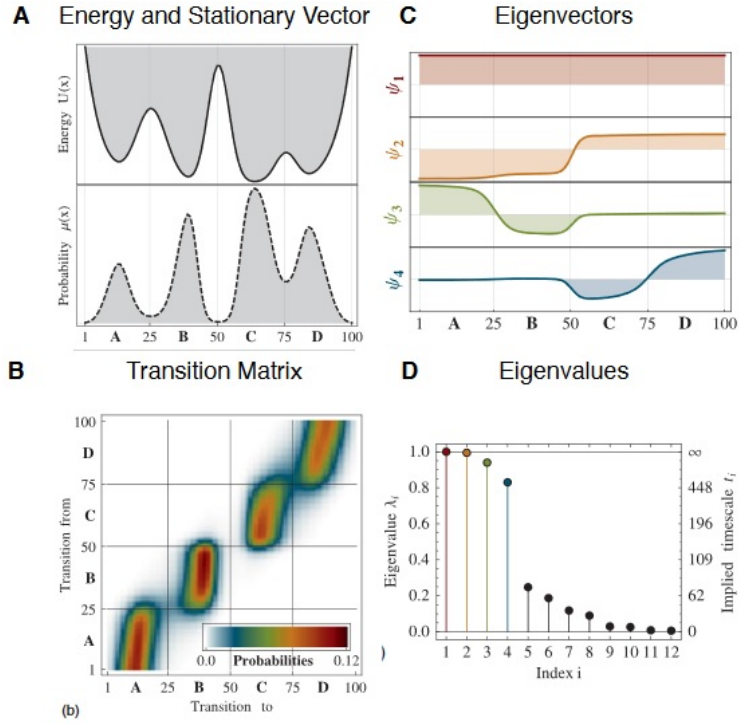


Figure 2.2: Relation of eigenvalues and eigenvectors to metastability of a process

to get real eigenvalues. 2: eigenvector(?) problem of the transfer operator has only global solutions.

## 2.3 Fuzzy Clustering

The above considerations (metastable decomposition induced by zeros of eigenfunctions) result in a metastable full decomposition of the state space, i.e. each state is assigned to exactly one of the partition 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.

via Membership Functions

### Set-based vs. Function-based Approach

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

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

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

### Membership functions (=almost characteristic function)

Assuming we have already determined that our process consists of  $n$  metastable sets (by knowing its  $n$  dominant eigenvalues). We will follow the approach of [16] by defining macro states as *overlapping partial densities*. They can be identified by membership functions  $\chi_1, \dots, \chi_n : \mathbb{X} \rightarrow [0, 1]$ . Each state of the original state space shall be assigned to the different macro states with a certain *degree of membership*.

**Definition 2.3.** *membership function  $\chi_i$  = partition of unity and nonnegative*

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 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 1.3. That gives us a Markov State Model where each (macro) state is a metastable set of the original process.

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

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

so far: membership fct indep. of metast. set

### Statistical Weights

For each macrostate we can define a statistical weight

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

i.e. ....

$D = \text{diag}(w_1, \dots, w_n)$  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 1.15.

what for?

## 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 Deuffhard et al[1] which used the sign structure of the dominant eigenvalues of the transition matrix. That approach has been improved by Deuffhard and Weber[2] 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[17] extended it even on continuous processes.

set-based approach?

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

eigenvector matrix?

membership function matrix  $\chi$

$$\chi_j(q) = \sum_{i=1}^n \mathcal{A}_{ij} \mathcal{X}_i(q), \quad j = 1, \dots, n. \quad (2.1)$$

Here,  $\mathcal{A} = \{\mathcal{A}_{ij}\}_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$  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  $\chi$  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[16].

Weber shows in [17] that the choice  $\chi = \mathcal{X}\mathcal{A}$  (any linear combination of the eigenfunctions  $\mathcal{X}$ ) preserves the Markov property of the process.

or only conv. lin.comb.? even: discretiz. error vanishes

## 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 2.5** (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  $\mu$  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

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

why only finite?

see ..

**Definition 2.6** (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  $\mu$ .

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 2.7** (Cycle of a process). A  $k$ -cycle  $\gamma$  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[6].

**Definition 2.8** (Cycle/flow Decomposition). A collection  $\mathcal{C}_+ \subset \mathcal{C}$  of cycles  $\gamma$  with real positive weights  $w(\gamma)$  is a flow decomposition if for every edge  $(i, j) \in S^2$  we have

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

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

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

**Definition 2.9** (Weight of a Cycle).

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

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 2.10** (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 2.11** (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 triangular, but only quasi-triangular, allowing  $2 \times 2$ -blocks on its diagonal. The eigenvalues of the  $2 \times 2$ -blocks are exactly the complex conjugate eigenpairs of  $P$ .

**Definition 2.12** (Schur Vector). Let  $\tilde{R} \in \mathbb{R}^{m \times m}$  be a submatrix of  $R$  (top left part of  $R$ ). Then

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

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

## Using PCCA+

Djurdjevac Conrad et al[3] propose an algorithm in order to determine/get the desired membership vectors  $\chi$ .

- i) Compute a real Schur decomposition  $(\tilde{X}, R)$  of ...
- ii) Sort the Schur values and the  $2 \times 2$ -blocks such that they are in a descending order
- iii) Determine the submatrix  $\hat{X}$  and solve PCCA+ equation (...) in order to get the membership functions  $\chi$

## 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.

kin./dyn.?

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 1.2). This operator will be projected onto a finite-dimensional state space (section 1.3) which may spoil the Markov Property of the process (section 1.4). This chapter is mainly based on Weber and Fackeldey[19].

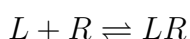
nop

**TODO: In which model are we working? Hamiltonian ( $\Gamma$ )? Hamiltonian w/ randomized momentum ( $\Omega$  w/ any momentum)? Langevin? Diffusion Dynamics?**

### 3.1 Receptor-Ligand System

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

together with reb. eff.



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

### 3.2 Rebinding Effect

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

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

In fact, a stochastic process describing a receptor-ligand molecular system is NOT necessarily Markovian. The Markovianity can be spoiled by the Rebinding Effect. If a Receptor-Ligand system dissolves, due to the favorable spatial situation (?) it is more likely to rebind again than to stay dissolved.

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

An important application of receptor-ligand processes is drug design. In short: A drug consists of ligands which should bind to the receptors of the virus. If the drug creates many bindings, the virus is "bound" and cannot attack the human (cell?) anymore. So many bindings are a favorable thing. So a high rebinding effect enhances the (overall?) binding affinity of the process/ system which is good for the efficiency of a drug. We want a high rebinding effect. So in this chapter, we examine the minimal rebinding effect for a given Kinetics. This task has been solved by Weber and Fackeldey[19] for reversible processes.

How: from bound to un-bound

### 3.3 Molecular Kinetics as a Projection

In this section, we will basically embed the mathematical concepts/results of chapter 1 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* (=microscopic) 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, a probability distribution assigning to each micro state a probability depending on the energy and temperature of the system, see McQuarrie[8]. It can be expressed in the form

$\bar{\rho}$  vs  $\pi$

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

where  $\beta = 1/(k_B T)$  is the inverse of the temperature  $T$  multiplied with the Boltzmann constant  $k_B$ . The Hamilton function denoted by  $H$  is given by  $H(q, p) = K(p) + V(q)$ , the sum of the kinetic energy  $K(p)$  and the potential energy  $V(q)$ . The Boltzmann distribution  $\pi$  can be decomposed as  $\pi = \pi_p \pi_q$ , where  $\pi_p : \mathbb{R}^{3N} \rightarrow \mathbb{R}$  is the probability function of the kinetic part in the momentum space and  $\pi_q : \Omega \rightarrow \mathbb{R}$  is the probability function of the potential part in the spatial/configurational space  $\Omega$ . Later, we will project our dynamics onto  $\Omega$  (need a *reduced density*).

see ..



## Macro States

Group/cluster a collection of the micro states having the same or similar values in one observable (e.g. bound, unbound state of receptor-ligand system). By grouping micro states, the (corresponding?) macro states yield statistical weights and entropic information (?). Macro states need not be distinct sets. Approach here: overlapping partial densities, which can be identified as membership functions  $\chi_1, \dots, \chi_n$ .

**TODO: statistical weights and entropic information**

## General Approach here

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

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

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

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

At the end, we get the transition matrix(not stoch.) and the infinitesimal generator of the finite process. This important tool will help us to solve our optimization problem at the end; maximize/minimize the rebinding effect.

## Membership functions

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

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

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

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

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

## Statistical Weights

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

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

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

and entropic information?  
what is that good for?

## Transfer Operator

For reversible(!) processes Weber [17] defines a transfer operator as follows:

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

This transfer operator is self-adjoint! due to the detailed balance condition. So for the nonreversible approach we cannot use this operator? Or: we can use this operator but it is not self-adjoint anymore? (so complex eigenvalues)..

trivial?

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

?

## Relation between Transfer Operator and Markov Operator

A transfer operator  $\mathcal{P}(\tau)$  also defines a (corresponding) projected Markov operator  $\bar{\mathcal{P}}(\tau)$  acting in configurational space  $\Omega$ , see Weber[17];

def

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

## Discrete Markov State Model

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

As we have seen in chapter 2, the number of metastable sets of a process can be determined by the number of dominant eigenvalues; i.e. we want 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  $P(\tau)$  (i.e. row-stochastic matrix).

We want to get from our continuous operator  $\mathcal{P}(\tau)$  to a discrete matrix  $P(\tau)$ , while preserving the most important properties of the process. For doing this, we have at first to determine the size and shape of the membership function  $\chi_i$ . Markovianity?

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

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

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

### Galerkin Projection

We can reduce our continuous stochastic process to a discrete process by the Galerkin projection

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

If we use the transfer operator  $\mathcal{P}$  from (3.4) we have NO discretization error under this projection. It means that  $(\mathcal{P}(\tau))^k = (P(\tau))^k$ . In particular, the Galerkin projection of this particular transfer operator preserves Markovianity!

For the projected transfer operator, we can use the matrix representation from theorem 1.15  $P = S^{-1}T$ , where  $S$  is the matrix corresponding to the statistical weights  $w_i$ .

### Infinitesimal Generator

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

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

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

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

Weber shows in [17] that such an infinitesimal generator exists for a discretization in terms of membership functions.

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

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

## 3.4 Minimizing the Rebinding Effect

For reversible processes, this problem is solved by Weber and Fackeldey [19] with the spectral approach.

A nontrivial rebinding effect can be estimated only if the kinetics  $Q_c$  of a system is nonreversible!!

## 3.5 Approach for nonreversible processes

With the tools from section 2.4 (Schur Decomposition) we give an approach how this problem can be solved for nonreversible processes (NESS processes) using Schur Decomposition to get rid of the possibly nonreal eigenvalues, see Djurdevac et al [3](2016).

### Transfer Operator

Have the transfer operator  $\mathcal{P}$  from (3.4) given, but in the case of a nonreversible process it 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?  
?

### Galerkin Projection

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

## 4 Illustrative Examples

### 4.1 An artificial (bivalent) binding Process

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

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

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

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

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

# Bibliography

- [1] 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.
- [2] P. DEUFLHARD AND M. WEBER, *Robust perron cluster analysis in conformation dynamics*, Linear algebra and its applications, 398 (2005), pp. 161–184.
- [3] N. DJURDJEVAC CONRAD, M. WEBER, AND C. SCHÜTTE, *Finding dominant structures of nonreversible markov processes*, Multiscale Modeling & Simulation, 14 (2016), pp. 1319–1340.
- [4] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, The John Hopkins University Press, Baltimore and London, 1996.
- [5] W. HUISINGA, *Metastability of markovian systems*, Ph.D. thesis, Freie Universität Berlin, (2001).
- [6] S. L. KALPAZIDOU, *Cycle representations of Markov processes*, vol. 28, Springer Science & Business Media, 2007.
- [7] T. KATO, *Perturbation Theory for Linear Operators*, Classics in Mathematics. Springer, 1995.
- [8] D. A. MCQUARRIE, *Statistical Mechanics*, University Science Books, California, 2000.
- [9] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Communications and Control Engineering Series. Springer, 1993.
- [10] A. NIELSEN, *Computation schemes for transfer operators*, Ph.D. thesis, Freie Universität Berlin, (2015).
- [11] 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.

- [12] M. SARICH, *Projected transfer operators*, PhD thesis, Freie Universität Berlin, 2011.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] M. WEBER, *Meshless methods in conformation dynamics*, Ph.D. thesis, Freie Universität Berlin, (2006).
- [17] M. WEBER, *A subspace approach to molecular markov state models via a new infinitesimal generator*, Habilitation thesis, Freie Universität Berlin, (2011).
- [18] 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.
- [19] M. WEBER AND K. FACKELDEY, *Computing the minimal rebinding effect included in a given kinetics*, Multiscale Modeling & Simulation, 12 (2014), pp. 318–334.