Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

On Markov State Models for Metastable Processes

Natasa Djurdjevac, Marco Sarich, and Christof Schütte*

Abstract

We consider Markov processes on large state spaces and want to find low-dimensional structure-preserving approximations of the process in the sense that the longest timescales of the dynamics of the original process are reproduced well. Recent years have seen the advance of so-called Markov state models (MSM) for processes on very large state spaces exhibiting metastable dynamics. It has been demonstrated that MSMs are especially useful for modelling the interesting slow dynamics of biomolecules (cf. Noe et al, PNAS(106) 2009) and materials. From the mathematical perspective, MSMs result from Galerkin projection of the transfer operator underlying the original process onto some low-dimensional subspace which leads to an approximation of the dominant eigenvalues of the transfer operators and thus of the longest timescales of the original dynamics. Until now, most articles on MSMs have been based on full subdivisions of state space, i.e., Galerkin projections onto subspaces spanned by indicator functions. We show how to generalize MSMs to alternative low-dimensional subspaces with superior approximation properties, and how to analyse the approximation quality (dominant eigenvalues, propagation of functions) of the resulting MSMs. To this end, we give an overview of the construction of MSMs, the associated stochastics and functional-analysis background, and its algorithmic consequences. Furthermore, we illustrate the mathematical construction with numerical examples.

Mathematics Subject Classification (2000). Primary 65C50; Secondary 60J35.

Keywords. Markov process, metastability, transition path theory, milestoning, eigenvalue problem, transfer operator, propagation error, Markov state models, committor, Galerkin approximation

*Supported by the DFG research center MATHEON "Mathematics for key technologies" in Berlin.

Fachbereich Mathematik und Informatik, Institut für Mathematik, Freie Universität Berlin. E-mails: djurdjev@mi.fu-berlin.de, sarich@mi.fu-berlin.de, and schuette@mi.fu-berlin.de.

1. Introduction

We consider Markov processes on large state spaces that have a unique invariant measure. We are interested in the question of whether we can find a low-dimensional approximation of the process in the sense that the longest timescales of the dynamics of the original process are reproduced well and the low-dimensional model inherits the essential structural properties of the original process: the dynamics transports probability distribution (or densities, respectively) into probability distributions (or densities), i.e., non-negativity and normalization are preserved. This is an rather old problem that has been answered in many different ways some belonging to classical themes in the literature [1, 2] that have been discussed in hundreds of articles, e.g., Markov chain decomposition for nearly reducible chains (for example, by aggregation-disaggregation techniques [3, 2, 4], stochastic complementation [5, 6], Perron Cluster Cluster Analysis (PCCA) [7, 8]), or network partition problems [9]. In these classical topics most contributions consider finite state spaces and have been based on linear algebra and associated stochastic analysis approaches.

Recent years have seen the advance of so-called Markov state models (MSM) as low-dimensional models for processes on very large, mostly on continuous state spaces exhibiting metastable dynamics [10, 7, 11, 12, 13]. Recently the interest in MSMs has drastically increased since it could be demonstrated that MSMs can be constructed even for very high dimensional systems [11] and have been especially useful for modelling the interesting slow dynamics of biomolecules [14, 15, 16, 17, 18, 19] and materials [20] (there under the name "kinetic Monte Carlo"). Metastable dynamics means that one can subdivide state space into metastable sets in which the system remains for *long* periods of time before it exits *quickly* to another metastable set; here the words "long" and "quickly" mainly state that the typical residence time has to be much longer than the typical transition time so that the jump process between the metastable sets is approximately Markovian. An MSM then just describes the Markov process that jumps between the sets with the aggegrated statistics of the original process.

The approximation quality of a MSM on large time scales has been rigorously studied for many different systems, e.g., for diffusion processes, or Glauber dynamics and Ising models in the limit of vanishing smallness parameters (noise itensity, temperature) where the analysis can be based on large deviation estimates and variational principles [21, 22] and/or potential theory and capacities [23, 24]. In these cases the effective dynamics is governed by some MSM with exponentially small transition probabilities and its states label the different attracting sets of the underlying Markov process. Other, quite general, rigorous approaches to the construction of MSM involve the exploitation of spectral properties, where the relation between dominant eigenvalues and eigenvectors, exit times and rates, and metastable sets has been studied

in detail, in some cases even without assumptions about smallness parameters [25, 26, 12, 13, 7, 19].

In this contribution we will use the approach via Galerkin discretization of the *transfer operator* of the original Markov process as developed in [12, 11, 7, 10]; here "transfer operator" just refers to a generalization of the transition matrix on finite discrete state spaces to general, e.g., continuous state spaces. In this approach the low-dimensional approximation results from orthogonal projection of the transfer operator onto some low-dimensional subspace. For so-called *full partition MSM* this subspace is spanned by indicator functions of n sets that partition state space. Then, the Galerkin approach has a direct stochastic interpretation since the resulting n-dimensional approximation simply exhibits jumps between the sets with aggegrated statistics as mentioned above.

However in many cases indicator ansatz spaces do not allow to achieve good approximation quality for reasonably small numbers of sets [19]. Therefore other ansatz spaces, e.g., fuzzy ansatz spaces, have also been discussed [27]. This obviously raises the questions of (a) how to find good ansatz functions, (b) what may be the associated stochastic interpretation, and (c) what can be said about the resulting approximation quality. Let $D = \mathrm{span}\{q_1, \ldots, q_n\}$ denote the low-dimensional ansatz space in state space S. We will be interested in ansatz functions q_i that are non-negative functions with $\sum_{x \in S} q_i(x) = 1$ or $\int_S q_i(x)dx = 1$ for any i so that $q_i(x)$ can be interpreted as the probability (density) that state x belongs to MSM state i. We will herein discuss an approach that allows to identify such ansatz functions and answers the above three questions jointly for full partition and fuzzy ansatz functions. The key idea will be that we consider n sets $C_1, \ldots, C_n$ that (in general) do *not* partition the state space but are just the very cores of the different attracting sets of the underlying Markov process. These *core sets* are then used as *milestones* in the sense of the milestoning approach as introduced in [28]: The approximating m-dimensional milestoning process is assigned to state i whenever the last core entered by the original process has been C_i. We will see that we can relate the milestoning process to transition path theory [29, 30, 31, 14] and use it to construct good fuzzy ansatz functions. The resulting low-dimensional MSM will prove to have very good approximation properties whenever the core sets have been chosen appropriately.

The remainder of the paper is organized as follows. In Section 2 we introduce the setting, define transfer operators, introduce full-partition MSM and relate them to Galerkin projections. Then, in Sec. 3 we introduce the milestoning process, relate it to transition path theory, and analyse its transition statistics. Section 4 then discusses Galerkin projection of the transfer operator in general, gives rigorous approximation results for long-term behavior and for eigenvalues and related timescales, and then shows how to use the milestoning process to compute the resulting MSMs efficiently. Finally, the results are illustrated with numerical experiments in Section 5.

2. Setting the Scene

We consider a Markov process $(X_t)_{t\in T}$ on a discrete state space S and its associated family of transition matrices $(P_t)_{t\in\mathbb{N}}$ with entries

$$p_t(x,y) = \mathbb{P}[X_t = y | X_0 = x]. \tag{1}$$

We restrict our considerations to discrete state spaces just for simplicity of presentation; all statements made in the following can be generalized to continuous state spaces as well (see Remark 2.1):

Because (X_t) is a Markov process, the transition matrices have the *semigroup property*

$$P_t P_s = P_{t+s}. \tag{2}$$

If (X_t) is a time-discrete process, i.e. a Markov chain with $T = \mathbb{N}$, we will only consider $P := P_1$, because (2) implies

$$P_t = P^t. \tag{3}$$

If (X_t) is time-continuous, it is usually referred to as a Markov jump process. In this case, the dynamics of the process is given by its *generator* L with entries $l(x,y)$ such that

$$P_t = e^{Lt}. \tag{4}$$

The generator L can also be defined explicitly

$$L = \lim_{t\to\infty} \frac{P_t - Id}{t}, \tag{5}$$

so that its entries form a rate matrix

$$l(x,y) \geq 0,\ x \neq y, \qquad l(x,x) = -\sum_{y\neq x} l(x,y). \tag{6}$$

For the time-discrete case, we define an analog for the rate matrix by setting $L_d = P - Id$ which we call *discrete generator*.

In the following we always assume that (X_t) has a unique invariant measure μ, that is given by

$$(P_t\mu)(y) = \sum_{x\in S} p_t(x,y)\mu(x) = \mu(y). \tag{7}$$

Now we introduce the family of *transfer operators* (T_t) that describes the propagation of densities in L^2_μ

$$(T_t f)(y)\mu(y) = \sum_x f(x) p_t(x,y)\mu(x) \tag{8}$$

and set $T := T_1$ for discrete time.
In analogy, we define on L^2_μ

$$(\mathcal{L}f)(y)\mu(y) = \sum_x l(x,y)f(x)\mu(x) \tag{9}$$

and for the discrete case

$$\mathcal{L}_d = T - Id. \tag{10}$$

In the following we will only consider the scalar product in L^2_μ, the induced 2-norm and the 1-norm

$$\langle f, g\rangle = \sum_x f(x)g(x)\mu(x), \qquad \|f\|^2 = \langle f, f\rangle, \qquad \|f\|_1 = \sum_x |f(x)|\mu(x). \tag{11}$$

In the theory of building standard Markov state models (MSM) one chooses a partitioning of state space, i.e. sets $A_1, ..., A_n$, such that

$$A_i \cap A_j = \emptyset, \quad i \neq j, \qquad \bigcup_{i=1}^{n} A_i = S \tag{12}$$

and a certain *lag time* $\tau > 0$. Then one can compute the transition probabilities

$$\mathbb{P}[X_\tau \in A_j | X_0 \in A_i]$$

and use the corresponding Markov chain on the index space $\{1, ..., n\}$ to approximate the original dynamics, switching between those sets. The approximation quality of such MSMs is discussed in [19]. A key feature is, that the Markov chain on the index space represents the dynamics of a projection of the transfer operator, that is $QT_\tau Q$, where Q is the orthogonal projection onto

$$D = \text{span}\left\{\mathbb{1}_{A_1}, ..., \mathbb{1}_{A_n}\right\}.$$

As outlined above, we will not restrict our attention to full partitionings of state space. However, we will return to the analysis of Galerkin projections of transfer operators $QT_\tau Q$, also to projections onto step-function spaces.

Remark 2.1. *On continuous state space the transfer operator $T_t : L^2_\mu \to L^2_\mu$ is defined via*

$$\int_C T_t f(y)\mu(dy) = \int_S \mathbb{P}[X_t \in C | X_0 = x] f(x)\mu(dx), \quad \text{for all measurable } C \subset S,$$

for the general case where the transition function $p(t, x, C) = \mathbb{P}[X_t \in C | X_0 = x]$ as well as the invariant measure may contain singular as well as absolutely continuous parts. Then, all of the above and subsequent sums have to be replaced by respective integrals. Further details, in particular regarding the respective generators for, e.g., diffusion processes, can be found in [12].

3. Milestoning and Transition Path Theory

We will now follow the approach first introduced in [32]. While the approach in [32] is restricted to reversible processes with generators, we will herein present the general framework for non-reversible processes.

3.1. Core sets and committors. Motivated by standard Markov state model approach we define sets $C_1, ..., C_n \subset S$, that we will call *core sets*, such that

$$C_i \cap C_j = \emptyset, \quad i \neq j. \tag{13}$$

That is, we relax the full partition constraint in (12). We denote the region that is not assigned to any core set by

$$C = S \setminus \bigcup_{k=1}^{n} C_k.$$

For analyzing the switching dynamics of the original process between the core sets we introduce the *milestoning process* $(\hat{X}_t)$

$$\hat{X}_t = i \Leftrightarrow X_{\sigma(t)} \in C_i, \text{ with } \sigma(t) = \sup_{s \leq t} \left\{ X_s \in \bigcup_{k=1}^{n} C_k \right\}, \tag{14}$$

i.e. the milestoning process is in state i, if the original process came last from core set C_i, cf. [28].

Now let $q_i^+(x)$ denote the probability that the process (X_t) will visit the core set C_i next, conditional on being in state x. q_i^+ is usually referred to as the *forward committor* and, as for example in [30], one can derive that q_i^+ is the solution of

$$\begin{aligned} (Lq_i^+)(x) &= 0, \quad \forall x \in C, \\ q_i^+(x) &= 1, \quad \forall x \in C_i, \\ q_i^+(x) &= 0, \quad \forall x \in C_j, j \neq i. \end{aligned} \tag{15}$$

In a similar way it can be shown, that the *backward committor* $q_i^-(x) = \mathbb{P}[\hat{X}_t = i | X_t = x]$, i.e. the probability that the process (X_t) came last from core set C_i, conditional on being in state x, solves

$$\begin{aligned} (\mathcal{L}q_i^-)(x) &= 0, \quad \forall x \in C, \\ q_i^-(x) &= 1, \quad \forall x \in C_i, \\ q_i^-(x) &= 0, \quad \forall x \in C_j, j \neq i. \end{aligned} \tag{16}$$

In the time-discrete case one has to replace L by the discrete generator L_d and $\mathcal{L}$ by $\mathcal{L}_d$. Moreover one can show, that (15) and (16) have a unique solution under the assumption that the invariant measure is unique and not vanishing on all core sets.

Note that $\mathcal{L}$ or in time-discrete setting $\mathcal{L}_d$ generate the family of transition matrices (P_t^b) with entries

$$p_t^b(y,x) = \frac{\mu(x)}{\mu(y)} p_t(x,y), \tag{17}$$

which describe the dynamics of the process (X_t) running backward in time.

For more details on the definition and properties of committors we refer to [29, 30, 31, 14]; the discrete setting studied herein is worked out in [30].

3.2. Jump statistics of milestoning process.

Time-discrete case: Transition probabilities. When observing a time-discrete process (X_n), we can define the transition matrix $\hat{P}$ of the milestoning process $(\hat{X}_n)$, with entries $\hat{p}(i,j) = \mathbb{P}_\mu(\hat{X}_{n+1} = j | \hat{X}_n = i)$. Since in general the milestoning process will not be a Markov process, we cannot assume that it is essentially characterized by its transition matrix $\hat{P}$; this also holds true for the generator $\hat{L}_d$ whose definition therefore should be understood as a formal one at this point. We will see that it is *not* the crucial point whether the dynamics of the milestoning process is governed by $\hat{P}$ or not.

Based on the introduced quantities we have

$$\mathbb{P}_\mu(\hat{X}_n = i, X_n = x) = \mathbb{P}_\mu(\hat{X}_n = i | X_n = x)\mathbb{P}_\mu(X_n = x) = q_i^-(x)\mu(x).$$

Therefore, the total probability that the milestoning process is assigned to state i, i.e. the invariant measure of the milestoning process is

$$\hat{\mu}(i) = \mathbb{P}_\mu(\hat{X}_n = i) = \sum_x \mathbb{P}_\mu(\hat{X}_n = i, X_n = x) = \sum_x q_i^-(x)\mu(x) = \|q_i^-\|_1.$$

The following theorem gives us the entries of the discrete generator.

Theorem 3.1. *For a time-discrete process (X_n), the entries of the discrete generator $\hat{L}_d$ of the milestoning process $(\hat{X}_n)$ are given with*

$$\hat{l}_d(i,j) = \frac{1}{\|q_i^-\|_1} \langle q_j^+, \mathcal{L}_d q_i^- \rangle. \tag{18}$$

Proof. Using that

$$\begin{aligned} &\mathbb{P}_\mu(X_{n+1} = y, \hat{X}_n = i, X_n = x) \\ &\quad = \mathbb{P}_\mu(X_{n+1} = y | \hat{X}_n = i, X_n = x)\mathbb{P}_\mu(\hat{X}_n = i, X_n = x) \\ &\quad = p(x,y) q_i^-(x)\mu(x), \end{aligned}$$

we can calculate

$$\begin{aligned} &\mathbb{P}_\mu(\hat{X}_{n+1} = j, X_{n+1} = y, \hat{X}_n = i, X_n = x) = \\ &\quad = \mathbb{P}_\mu(\hat{X}_{n+1} = j | X_{n+1} = y, \hat{X}_n = i, X_n = x)\mathbb{P}_\mu(X_{n+1} = y, \hat{X}_n = i, X_n = x) \\ &\quad = \begin{cases} \mathbb{1}_{C_j}(y) p(x,y) q_i^-(x)\mu(x), & \text{if } i \neq j \\ \mathbb{1}_{C_i \cup C}(y) p(x,y) q_i^-(x)\mu(x), & \text{if } i = j. \end{cases} \end{aligned}$$

Therefore, the one-step transition probability $\hat{p}(i,j)$ from i to $j \neq i$ is given by

$$\begin{aligned}\hat{p}(i,j) &= \mathbb{P}_\mu(\hat{X}_{n+1}=j|\hat{X}_n=i) = \frac{\mathbb{P}_\mu(\hat{X}_{n+1}=j,\hat{X}_n=i)}{\mathbb{P}_\mu(\hat{X}_n=i)} \\ &= \frac{1}{\mathbb{P}_\mu(\hat{X}_n=i)}\sum_{x,y}\mathbb{P}_\mu(\hat{X}_{n+1}=j,X_{n+1}=y,\hat{X}_n=i,X_n=x) \\ &= \frac{1}{\|q_i^-\|_1}\sum_{x,y}\mathbb{1}_{C_j}(y)p(x,y)q_i^-(x)\mu(x) = \frac{1}{\|q_i^-\|_1}\langle Tq_i^-,\mathbb{1}_{C_j}\rangle.\end{aligned}$$

In addition, when $i=j$

$$\begin{aligned}\hat{p}(i,i) &= \mathbb{P}_\mu(\hat{X}_{n+1}=i|\hat{X}_n=i) == \frac{\mathbb{P}_\mu(\hat{X}_{n+1}=i,\hat{X}_n=i)}{\mathbb{P}_\mu(\hat{X}_n=i)} \\ &= \frac{1}{\mathbb{P}_\mu(\hat{X}_n=i)}\sum_{x,y}\mathbb{P}_\mu(\hat{X}_{n+1}=i,X_{n+1}=y,\hat{X}_n=i,X_n=x) \\ &= \frac{1}{\|q_i^-\|_1}\sum_{x,y}\mathbb{1}_{C_i\cup C}(y)p(x,y)q_i^-(x)\mu(x) \\ &= \frac{1}{\|q_i^-\|_1}\langle Tq_i^-,\mathbb{1}_{C_i\cup C}\rangle.\end{aligned}$$

Using the properties of committors on core sets for $i \neq j$, we get that

$$\begin{aligned}\langle Tq_i^-,\mathbb{1}_{C_j}\rangle &= \langle Tq_i^-,q_j^+\rangle - \langle Tq_i^-,q_j^+\mathbb{1}_C\rangle = \langle Tq_i^-,q_j^+\rangle - \langle q_i^-,q_j^+\mathbb{1}_C\rangle \\ &= \langle (T-Id)q_i^-,q_j^+\rangle = \langle \mathcal{L}_d q_i^-,q_j^+\rangle,\end{aligned}$$

which yields

$$\hat{l}_d(i,j) = \hat{p}(i,j) = \frac{1}{\|q_i^-\|_1}\langle q_j^+,\mathcal{L}_d q_i^-\rangle,\ i\neq j.$$

Similarly, for $i=j$, we get

$$\begin{aligned}\langle Tq_i^-,\mathbb{1}_{C_i\cup C}\rangle &= \langle Tq_i^-,\mathbb{1}_{C_i}\rangle + \langle Tq_i^-,\mathbb{1}_C\rangle \\ &= \langle Tq_i^-,q_i^+\rangle - \langle q_i^-,q_i^+\mathbb{1}_C\rangle + \langle q_i^-,\mathbb{1}_C\rangle \\ &= \langle (T-Id)q_i^-,q_i^+\rangle + \|q_i^-\|_1 = \langle \mathcal{L}_d q_i^-,q_i^+\rangle + \|q_i^-\|_1,\end{aligned}$$

and

$$\hat{l}_d(i,i) = \hat{p}(i,j) - 1 = \frac{1}{\|q_i^-\|_1}(\langle q_i^+,\mathcal{L}_d q_i^-\rangle + \|q_i^-\|_1) - 1 = \frac{1}{\|q_i^-\|_1}\langle q_i^+,\mathcal{L}_d q_i^-\rangle.$$

□

Time-continuous case: Transition rates. Now we will show that all the above identities are still valid in a time-continuous case. For a given infinitely long trajectory and $i \neq j$, we define a (i,j)-reactive trajectory as a piece of this infinite long trajectory in a time interval R_{ij}^m such that for any $t \in R_{ij}^m$ we have that the next first entry into a core set is in C_j while the last first entry into a core set happened in C_i. Then, at a certain time t we are on a (i,j)-reactive trajectory if

$$t \in R_{ij} = \cup_{m=-\infty}^{\infty} R_{ij}^m.$$

The probability current from x to y generated by (i,j)-reactive trajectories is then given by

$$f_{ij}(x,y) = \lim_{s \to 0+} \frac{1}{s} \mathbb{P}_\mu \Big(X_t = x, X_{t+s} = y, t \in R_{ij}, t+s \in R_{ij} \Big),$$

In order to compute this quantities we define $B_j = \bigcup_{k \neq j} C_k$ and denote the first hitting time of a set A by τ_A. Then $\mathbb{P}_y[\tau_{C_j} < \tau_{B_j}]$, is the probability to start at y and enter the core set C_j next rather than any other core set. Therefore

$$\begin{aligned}
&\mathbb{P}_\mu \Big(X_t = x, X_{t+s} = y, t \in R_{ij}, t+s \in R_{ij} \Big) = \\
&= \mathbb{P}_\mu \Big(X_{t+s} = y, t+s \in R_{ij} | X_t = x, t \in R_{ij} \Big) \mathbb{P}_\mu \Big(X_t = x, t \in R_{ij} \Big) \\
&= \mathbb{P}_\mu \Big(X_{t+s} = y | X_t = x \Big) \mathbb{P}_y(\tau_{C_j} < \tau_{B_j}) \mathbb{P}_\mu \Big(X_t = x, \hat{X}_t = i \Big) \\
&= p_s(x,y) q_j^+(y) q_i^-(x) \mu(x).
\end{aligned}$$

Since $i \neq j$ we have $l(x,y) = \lim_{s \to 0+} \frac{1}{s} p_s(x,y)$ and thus

$$f_{ij}(x,y) = l(x,y) q_j^+(y) q_i^-(x) \mu(x).$$

Now we can compute the rate k_{ij} of transitions from i to j, which is defined as the average number of (i,j)-reactive trajectories per unit time. This quantity is given by the total probability current through a dividing surface between C_i and C_j, i.e. by the total probability current generated by (i,j)-reactive trajectories through the boundary of C_i:

$$\begin{aligned}
k_{ij} &= \sum_{x \in C_i, y \in S \setminus C_i} f_{ij}(x,y) \\
&= \sum_{x \in C_i, y \in S \setminus C_i} q_j^+(y) l(x,y) q_j^-(x) \mu(x) \\
&= \sum_{x \in C_i, y \in S} q_j^+(y) l(x,y) q_i^-(x) \mu(x),
\end{aligned}$$

where the last identity results from $q_j^+(y) = 0$ for all $y \in C_i$. Since additionally $q_i^-(x) = 1$ for $x \in C_i$ we find

$$k_{ij} = \langle \mathcal{L} \mathbb{1}_{C_i}, q_j^+ \rangle.$$

Therefore, the off-diagonal entries $\hat{l}(i,j)$ of the generator for the milestoning process $\hat{X}_t$ result as

$$\hat{l}(i,j) = \frac{1}{\|q_i^-\|_1}\langle \mathcal{L}\mathbb{1}_{C_i}, q_j^+\rangle, \tag{19}$$

such that the diagonal entries have to be

$$\begin{aligned}\hat{l}(i,i) &= -\sum_{j\neq i}\frac{1}{\|q_i^-\|_1}\langle \mathcal{L}\mathbb{1}_{C_i}, q_j^+\rangle = -\frac{1}{\|q_i^-\|_1}\left\langle \mathcal{L}\mathbb{1}_{C_i}, \sum_{j\neq i} q_j^+\right\rangle \\ &= -\frac{1}{\|q_i^-\|_1}\langle \mathcal{L}\mathbb{1}_{C_i}, \mathbb{1}-q_i^+)\rangle = \frac{1}{\|q_i^-\|_1}\langle \mathcal{L}\mathbb{1}_{C_i}, q_i^+\rangle.\end{aligned}$$

Since $\langle \mathcal{L}\mathbb{1}_{C_i}, q_i^+\rangle = \langle \mathbb{1}_{C_i}, Lq_i^+\rangle$, we can use the same arguments as above to end up with

$$\langle \mathcal{L}\mathbb{1}_{C_i}, q_j^+\rangle = \langle \mathcal{L}q_i^-, q_j^+\rangle,$$

so that we have just proved the following theorem

Theorem 3.2. *For a time-continuous process (X_t), the entries of a generator of the milestoning process $(\hat{X}_t)$ are given with*

$$\hat{l}(i,j) = \frac{1}{\|q_i^-\|_1}\langle \mathcal{L}q_i^-, q_j^+\rangle. \tag{20}$$

3.3. Invariant measure and self-adjointness.

A Markov process (X_t) is reversible if

$$p_t(x,y)\mu(x) = p_t(y,x)\mu(y). \tag{21}$$

This condition is called *the detailed balance condition*. It obviously implies that

$$p_t^b(x,y) = p_t(x,y), \tag{22}$$

so the process running backward in time is equivalent to the process running forward in time.
Moreover, (21) implies

$$\begin{aligned}\langle Tf, g\rangle &= \sum_{x,y} p(x,y)f(x)g(y)\mu(x) \\ &\overset{(21)}{=} \sum_{x,y} p(y,x)f(x)g(y)\mu(y) = \langle f, Tg\rangle.\end{aligned} \tag{23}$$

This means that T is a self-adjoint operator. The same argument shows that also $\mathcal{L}$ is self-adjoint in the reversible case. Further, (22), (15) and (16) yield the identity of forward and backward committors, i.e.

$$q_i^- = q_i^+ \ \forall i = 1, ..., n. \tag{24}$$

Hence, in the following we will use the shorthand notation $q_i := q_i^- = q_i^+$.

First we note some properties of the milestoning generator $\hat{L}$.

Lemma 3.3. *Let (X_t) be a reversible Markov process with unique invariant measure μ. Then the milestoning generator $\hat{L}$ has the invariant measure*

$$\hat{\mu}(i) = \sum_x q_i(x)\mu(x)$$

and the according operator in $L^2(\hat{\mu})$

$$(\hat{\mathcal{L}}v)(j)\hat{\mu}(j) = \sum_{i=1}^{n} \hat{l}(i,j)v(i)\hat{\mu}(i)$$

is self-adjoint. Therefore it also defines a reversible jump process.

Proof. We have

$$\begin{aligned}\sum_{i=1}^{n} \hat{l}(i,j)\hat{\mu}(i) &= \sum_{i=1}^{n} \langle q_i, \mathcal{L}q_j\rangle \\ &= \langle \mathbb{1}, \mathcal{L}q_j\rangle = 0.\end{aligned}$$

Moreover,

$$\begin{aligned}\hat{l}(i,j)\hat{\mu}(i) &= \langle q_i, \mathcal{L}q_j\rangle \\ &= \langle \mathcal{L}q_i, q_j\rangle = \hat{l}(j,i)\hat{\mu}(j),\end{aligned}$$

which implies reversibility and self-adjointness. □

4. Galerkin Approximation

We will now discuss Galerkin projections of transfer operators. For the sake of simplicity we will restrict our considerations to reversible Markov processes. Before we enter into the details of Galerkin projections we will shortly address the properties of the milestoning process induced by reversible Markov processes.

4.1. Galerkin projection and eigenvalues. In this section we will only consider discrete processes (X_n). If (X_t) is time-continuous with generator $\mathcal{L}$, we will fix a lag time $\tau > 0$ and just consider the snapshot dynamics of $(X_{n\tau})$ with the semi-group of transfer operators (T_τ^n). In this case the eigenvalues of the transfer operator T_τ will be given by

$$\lambda_{i,\tau} = e^{\Lambda_i \tau}, \tag{25}$$

where $\Lambda_i < 0$ is an eigenvalue of the generator $\mathcal{L}$. Now we want to approximate the dynamics of (X_n) by its projection to some low-dimensional subspace D in terms of density propagation. Therefore we will denote the orthogonal

projection onto D by Q. Assume that the process (X_n) is initially distributed according to

$$\rho_0(x)\mu(x) = \mathbb{P}[X_0 = x], \tag{26}$$

where $\rho_0(x)$ is a distribution with respect to μ. Then at any time n the distribution of X_n is given by

$$\rho_n(y)\mu(y) = \sum_x p_n(x,y)\rho_0(x)\mu(x) \tag{27}$$

or in matrix notation

$$\rho_n = T^n\rho_0. \tag{28}$$

Next, consider

$$\tilde{\rho}_n = Q\rho_n. \tag{29}$$

If we assume that $\rho_0 = Q\rho_0 \in D$ is a consistent initial distribution, i.e. it belongs to the subspace $D \subset S$, we find

$$\tilde{\rho}_n = QT^nQ\rho_0. \tag{30}$$

So the operator QT^nQ describes the propagation of the initial density $\rho_0 \in D$ to $\tilde{\rho}_n = Q\rho_n$, but we do not have a semi-group property anymore, i.e.

$$QT^nQ \neq (QTQ)^n. \tag{31}$$

Subsequently we will consider subspaces $D \subset L^2_\mu$ such that $\mathbb{1} \in D$, i.e., the invariant measure with density $\mathbb{1}$ in L^2_μ is still contained in D. In this case we find in [19] an error bound for the approximation error from (31) $\|QT^nQ - (QTQ)^n\|$. We now cite Theorem 3.3 from section 3.4 of [19].

Theorem 4.1. *Let $T = T_\tau$ be a transfer operator of a time-continuous reversible Markov process with generator $\mathcal{L}$ for lag time $\tau > 0$, or the transfer operator of some time-discrete reversible process. Let $1 = \lambda_0 > \lambda_1 > ... > \lambda_{m-1}$ be the m dominant eigenvalues of T, i.e. for every other eigenvalue λ it holds $\lambda \leq r \leq \lambda_{m-1}$ such that r is the upper bound on the remaining spectrum. Furthermore, set $\eta = r/\lambda_1 < 1$. Whenever we have a generator, its eigenvalues Λ_i then satisfy: $\Lambda \in spec(\mathcal{L}), \Lambda \leq \Lambda_{m-1} \Rightarrow \Lambda \leq R < 0$ with $r = \exp(\tau R)$. Then, $\eta(\tau) = \exp(-\tau\Delta) < 1$ with*

$$\Delta = \Lambda_1 - R > 0, \tag{32}$$

as a τ-independent measure for the spread in the spectrum between the first nontrivial eigenvalue and the part of the spectrum that is not taken into account. Let $u_0, u_1, ..., u_{m-1}$ be the corresponding normalized eigenvectors. Let Q denote the projection onto some subspace $D \subset S$ with $\mathbb{1} \in D$ and define

$$\delta := \max_{j=1,...,m-1} \|Q^\perp u_j\| \tag{33}$$

where $Q^{\perp} = \mathrm{Id} - Q$. Finally, define the projected transfer operator $P = QTQ$. Then the error $E(k)$ satisfies $\|QT^kQ - P^k\|_1 \leq E(k) = \|QT^kQ - P^k\|$ and is bounded from above by

$$E(k) \leq \min[2\,;\, C(\delta, \eta(\tau), k)] \cdot \lambda_1^k, \tag{34}$$

with a leading constant of following form

$$\begin{aligned} C(\delta, \eta, k) &= ((m-1)\delta + \eta)[C_{sets}(\delta, k) + C_{spec}(\eta, k)] && (35)\\ C_{sets}(\delta, k) &= (m-1)^{1/2}(k-1)\,\delta && (36)\\ C_{spec}(\eta, k) &= \frac{\eta}{1-\eta}(1-\eta^{k-1}). && (37) \end{aligned}$$

The bound of Theorem 4.1 consists of two prefactors. C_{spec} depends on the lag time and the gap Δ in the spectrum of the generator. It will go to zero, if we increase the lag time τ, or, alternatively, the number m of eigenvectors that we have to approximate. The approximation or *projection error* δ of eigenvectors that we take into account governs the second part of the bound C_{sets}. More precisely, for fixed k, i.e., time span $k\tau$, the prefactor C_{sets} will be small, if the maximal projection error δ is small.

The next question is, how well the eigenvalues of the projected operator approximate the original eigenvalues of T. Because of self-adjointness of the transfer operator we can use the results from [33] to show

Theorem 4.2. *Let $1 = \lambda_0 > \lambda_1 > ... > \lambda_{m-1}$ be the m dominant eigenvalues of T, i.e. for every other eigenvalue λ it holds $\lambda < \lambda_{m-1}$. Let $u_0, u_1, ..., u_{m-1}$ be the corresponding normalized eigenvectors, $D \subset S$ a subspace with*

$$\mathbb{1} \in D \qquad \dim(D) =: n \geq m \tag{38}$$

and Q the orthogonal projection onto D.
Moreover, let $1 = \hat{\lambda}_0 > \hat{\lambda}_1 > ... > \hat{\lambda}_{m-1}$ be the dominating eigenvalues of the projected operator QTQ. Then

$$E(\delta) = \max_{i=1,...,m-1} |\lambda_i - \hat{\lambda}_i| \leq \lambda_1 (m-1)\delta^2, \tag{39}$$

where

$$\delta = \max_{i=1,...,m-1} \|Q^{\perp} u_i\|$$

is the maximal projection error of the eigenvectors to the space D.

For the proof we refer to [34].

Remark 4.1. *Inserting (25) into (39), we get the lag time depended eigenvalue estimate*

$$E(\tau, \delta) = \max_{i=1,...,m-1} |\lambda_i - \hat{\lambda}_i| \leq e^{\Lambda_1 \tau}(m-1)\delta^2, \tag{40}$$

where (λ_i) *are the dominant eigenvalues of the transfer operator* T_τ *and* $(\hat{\lambda}_i)$ *the dominant eigenvalues of the projection* $QT_\tau Q$.
Since $\Lambda_1 < 0$,

$$E(\tau, \delta) \to 0, \text{ for } \tau \to \infty. \tag{41}$$

Furthermore, for the relative eigenvalue error *we have, at least for the first non-trivial eigenvalue*

$$\frac{|\lambda_1 - \hat{\lambda}_1|}{|\lambda_1|} \leq (m-1)\delta^2, \tag{42}$$

from which we see that by decreasing the maximal projection error we will have control even over the relative eigenvalue error.

4.2. Estimating the eigenvalues from trajectories. In this section we choose the special subspace D that is spanned by the committors associated with some core sets $C_1 \ldots, C_n$, i.e. $D = span\{q_1, ..., q_n\}$. Because $\mathbb{1} \in D$, Theorem 4.1 and 4.2 apply. Moreover we will see that this subspace allows us to compute the projected operator QTQ, its eigenvalues and all other related quantities from a trajectory. The first step is

Theorem 4.3. *Let* D *be a the subspace spanned by the committors*

$$D = span\{q_1, ..., q_n\} \tag{43}$$

and let $\hat{\lambda}$ *be an eigenvalue of the operator* QTQ. *Then* $\hat{\lambda}$ *solves the generalized eigenvalue problem*

$$\hat{T}r = \hat{\lambda} M r, \tag{44}$$

with

$$\hat{T}_{ij} = \frac{\langle q_i, Tq_j \rangle}{\hat{\mu}(i)}, \tag{45}$$

$\hat{\mu}(i) = \|q_i\|$, *and the mass matrix*

$$M_{ij} = \frac{\langle q_i, q_j \rangle}{\hat{\mu}(i)}. \tag{46}$$

Proof. Let D be as in (43). Then the orthogonal projection Q can be written as

$$(Qv)(y) = \sum_{i,j=1}^{n} S_{ij}^{-1} \langle v, q_j \rangle q_i, \tag{47}$$

with $S_{ij} = \langle q_i, q_j \rangle$.
Since

$$\hat{T}_{ij} = \frac{\langle q_i, Tq_j \rangle}{\hat{\mu}(i)} = \frac{\langle q_i, (Id + \mathcal{L}_d) q_j \rangle}{\hat{\mu}(i)} = (\hat{L}_d)_{ij} + M_{ij}, \tag{48}$$

(44) is equivalent to

$$\hat{L}_d r = (\hat{\lambda} - 1) M r. \tag{49}$$

Let ϕ be an eigenvector of QTQ with respect to $\hat{\lambda}$, i.e.

$$\begin{aligned} QTQ\phi = \hat{\lambda}\phi \Leftrightarrow Q(\mathcal{L}_d + Id)Q\phi &= \hat{\lambda}\phi \\ \Leftrightarrow Q\mathcal{L}_d Q\phi = (\hat{\lambda} - 1)\phi. & \end{aligned}$$

This is equivalent to

$$\begin{aligned} & \langle Q\mathcal{L}_d Q\phi, q_i\rangle = (\hat{\lambda}-1)\langle\phi, q_i\rangle \quad \forall i = 1, ..., n \\ \Leftrightarrow \quad & \langle \mathcal{L}_d Q\phi, q_i\rangle = (\hat{\lambda}-1)\langle\phi, q_i\rangle \quad \forall i = 1, ..., n \\ \Leftrightarrow \quad & \sum_{j,k=1}^{n} S^{-1}_{jk}\langle\phi, q_k\rangle\langle\mathcal{L}_d q_j, q_i\rangle = (\hat{\lambda}-1)\langle\phi, q_i\rangle \quad \forall i = 1, ..., n. \end{aligned} \tag{50}$$

Introducing

$$r_j = \sum_{k=1}^{n} S^{-1}_{jk}\langle\phi, q_k\rangle$$

(50) can be written as

$$\sum_{j=1}^{n} r_j\langle\mathcal{L}_d q_j, q_i\rangle = (\hat{\lambda}-1)\langle\phi, q_i\rangle = (\hat{\lambda}-1)\sum_{j,k=1}^{n} S_{ij}S^{-1}_{jk}\langle\phi, q_k\rangle = (\hat{\lambda}-1)\sum_{j=1}^{n} S_{ij}r_j. \tag{51}$$

Deviding both sides by $\hat{\mu}(i)$ completes the proof. □

Theorem 4.3 states, that we can compute the eigenvalues of the projected transfer operator QTQ by solving the generalized eigenvalue problem (44). In general, Theorem 4.3 does not depend on the special choice of D being the subspace spanned by the committors. The advantage is, that for D as in (43) the entries $\hat{l}_d(i,j)$ and M_{ij} have a stochastic interpretation: We have already seen that $\hat{L}_d = \hat{P} - Id$ with

$$\hat{p}(i,j) = \mathbb{P}[\hat{X}_{n+1} = j | \hat{X}_n = i].$$

As well-known, we can approximate the transition probabilities $\hat{p}(i,j)$ of the process $(\hat{X}_n)$ from a (long enough) realization via the maximum likelihood estimator $\hat{p}^*_N$ with entries

$$\hat{p}^*_N(i,j) = \frac{n_{ij}(N)}{N_i(N)},$$

where $n_{ij}(N)$ is the number of transition from i to j observed in the finite trajectory $\hat{X}_n$, $n = 0, \ldots, N$, while $N_i(N) = \sum_j n_{ij}(N)$ is the total number of visits to state i in the trajectory. Since we are dealing with ergodic processes, we know that $\hat{P}^*_N \to \hat{P}$ in the limit of arbitrarily long trajectories, i.e., for $N \to \infty$ (law of large numbers).

Similarly, we can approximate the mass matrix M. We find

Lemma 4.4. *Let i, j be arbitrary and, as above, let $B_j = \bigcup_{k \neq j} C_k$ and let τ_A denote the first hitting time into set A. Then M_{ij} can be written as*

$$M_{ij} = \mathbb{P}[X_n \in C, \tau_{C_j} < \tau_{B_j} | \hat{X}_n = i],$$

that is, the probability to be outside of the core sets and enter the core set C_j next rather than any other core set, under the condition, that the last core set hit was C_i.

Thus, the entry M_{ij} of the mass matrix includes only those parts of (i, j)-reactive trajectories that are *outside* of core i and go to core j next, that is, in general, a typical (i, j) reactive trajectory will be much longer than those parts of it which enter into the mass matrix.

Proof. By definition

$$\begin{aligned}
\mathbb{P}[X_n \in C, \tau_{C_j} < \tau_{B_j} | \hat{X}_n = i] &= \frac{\mathbb{P}[\hat{X}_n = i, X_n \in C, \tau_{C_j} < \tau_{B_j}]}{\mathbb{P}[\hat{X}_n = i]} \\
&= \sum_{x \in C} \mathbb{P}[\hat{X}_n = i, X_n = x] \mathbb{P}_x[\tau_{C_j} < \tau_{B_j}] \frac{1}{\hat{\mu}(i)} \\
&= \sum_{x \in C} \mathbb{P}[\hat{X}_n = i | X_n = x] \mathbb{P}[X_n = x] \mathbb{P}_x[\tau_{C_j} < \tau_{B_j}] \frac{1}{\hat{\mu}(i)} \\
&= \sum_{x \in C} q_i(x) \mu(x) q_j(x) \frac{1}{\hat{\mu}(i)} = \frac{\langle q_i, q_j \rangle}{\hat{\mu}(i)} = M_{ij}.
\end{aligned}$$

□

Lemma 4.4 implies that we can estimate the mass matrix M by

$$M_N^*(i, j) = \frac{r_{ij}(N)}{R_i(N)}, \quad i \neq j$$

where $r_{ij}(N)$ is the total number of time steps during which the finite trajectory $\hat{X}_n, n = 0, \ldots, N$ is reactive from i to j, i.e. the number of time steps the process spend in C coming from C_i and going to C_j, while $R_i(N)$ is the total number of time steps during which the finite trajectory resides in i, i.e., $\hat{X}_n = i$.

So we can estimate $\hat{P}$ and M from a realization, i.e. a trajectory of the process (X_n), compute $\hat{T}$ by

$$\hat{T} \overset{(48)}{=} L_d + M = \hat{P} - Id + M \tag{52}$$

and solve the generalized eigenvalue problem in order to estimate the eigenvalues of the projected transfer operator QTQ.

Special case: full partition. When the core sets are chosen such that they form a full partition of state space (12), the definition of the committors directly yield

$$q_i(x) = \mathbb{1}_{C_i}(x). \tag{53}$$

That is, the committors are given by the characteristic functions on the coresets. This is exactly the standard MSM setting, such that the operator QTQ has a special interpretation, because

$$\hat{p}(i,j) = \mathbb{P}[X_n \in A_j | X_0 \in A_i] \tag{54}$$

is a matrix representation of the operator. Because of orthogonality of the stepfunctions we have

$$M_{ij} = \frac{\langle q_i, q_j \rangle}{\hat{\mu}(i)} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}. \tag{55}$$

Now Theorem 4.2 states, that the eigenvalues of the matrix (54) are close to the eigenvalues of the transfer operator T, if the corresponding eigenvectors are well approximated by step-functions on the partitioning sets.

5. Illustrative Examples

5.1. Double well potential with diffusive transition region.

We consider the diffusion process

$$\gamma dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}\gamma}dB_t \tag{56}$$

with B_t denoting Brownian motion in a potential V with two wells that are connected by an extended transition region. The potential V and its unique invariant measure μ are shown in Figure 1, we set the noise intensity $\sigma = \sqrt{2\beta^{-1}\gamma} = 0.8$ with $\gamma = 1$. We observe that the transition region between the two main wells contains four smaller wells that will have their own, less pronounced metastability each. The minima in the two main wells are located at $x_0 = -1$ and $x_1 = 6.62$, the respective saddle points that separate the main wells from the rest of the landscape at $x_0^{\pm} = x_0 \pm 1$, and $x_1^{\pm} = x_1 \pm 1$, respectively.

In order to find the transfer operator for this process we start with the Fokker-Planck equation $\partial_t u = \mathcal{L}u$, $u(t = 0, x) = f(x)$ that governs the propagation of a function f by the diffusion process. In the weighted Hilbert space L^2_μ the generator in the Fokker-Planck equation reads $\mathcal{L} = -\nabla V(x) \cdot \nabla_x + \beta^{-1}\Delta_x$, where ∇_x denotes the first derivative wrt. x and Δ_x the associated Laplacian. Thus, the transfer operator reads

$$T_t = \exp(t\mathcal{L}) \tag{57}$$

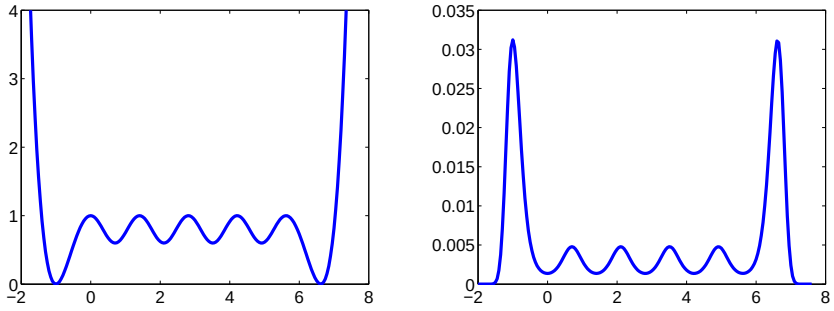

Figure 1. The potential V with extended transition region and the associated invariant measure for $\sigma = 0.8$.

This operator is self-adjoint since the diffusion process is reversible. The dominant eigenvalues of $\mathcal{L}$ take the following values:

Λ_0	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6	Λ_7
$+0.0000$	-0.0115	-0.0784	-0.2347	-0.4640	-0.7017	-2.9652	-3.2861

The main metastability has a corresponding implied timescale (ITS) $|1/\Lambda_1| \approx 88$ related to the transitions from one of the main wells to the other. Four other, minor metastable timescales related to the interwell switches between the main and the four additional small wells exist in addition. The eigenvalues have been computed by solving the eigenvalue problem for the partial differential operator $\mathcal{L}$ by an adaptive finite element (FE) discretization with an accuracy requirement of tol $= 1e-8$.

5.2. Two core sets.

In the following paragraphs we will compare the eigenvalues and ITS of the original process to the ones resulting from different MSM. More precisely, we first choose a lagtime τ and consider the transfer operator T_τ. Because of (25) we can compute the implied timescale

$$|1/\Lambda_1| = -\frac{\tau}{\ln(\lambda_{1,\tau})}, \tag{58}$$

where $\lambda_{1,\tau} < 1$ is the largest non-trivial eigenvalue of T_τ.
Next we choose two core sets of the form $C_0^s = (-\infty, x_0+s]$ and $C_1^s = [x_1-s, \infty)$ for some parameter s. Then we compare the ITS from (58) to the one, which corresponds to the largest non-trivial eigenvalue $\hat{\lambda}_{i,\tau}$ of the projected operator $QT_\tau Q$

$$|1/\hat{\Lambda}_1| = -\frac{\tau}{\ln(\hat{\lambda}_{1,\tau})}. \tag{59}$$

Since the process under investigation is just one-dimensional, we can compute the committor functions from the already mentioned FE discretization of $\mathcal{L}$

and just compute very accurate FE approximations of $\hat{T}_\tau$ and M, which allows to compute the eigenvalues of $QT_\tau Q$ as in Theorem 4.3. Figure 2 shows the dependence of the non-trivial eigenvalue on the core set size s for different values of the lagtime τ.

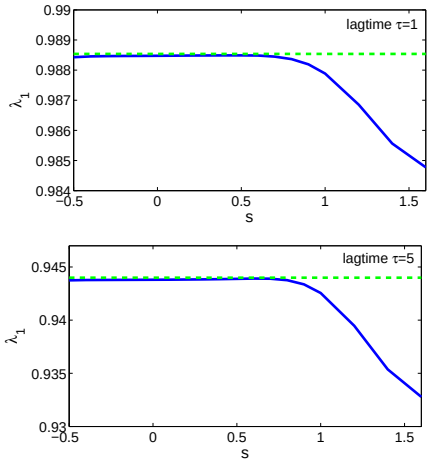

Figure 2. Non-trivial eigenvalues $\lambda_{1,\tau}^s < 1$ of the generalized eigenvalue problem $\hat{T}_\tau r = \hat{\lambda} M r$ versus cores set size parameter s for lagtimes $\tau = 1$ (left) and $\tau = 5$ (right) in comparison to the exact first non-trivial eigenvalue $\exp(\tau\Lambda_1)$.

We observe that the for small enough core sets the approximation of the exact first non-trivial eigenvalue of T_τ, $\exp(\tau\Lambda_1)$, is good, while for too large core sets the approximation quality decreases. This can be understood since for $s > 1$ the core sets contain parts of the transition regions of the process where recrossing events lead to an overestimation of the transition probability between the cores. Moreover, Theorem 4.2 connected this error to the projection error $\|Q^\perp u_1\|$ and Figure 3 shows that this error behaves exactly like the approximation quality of the eigenvalues.

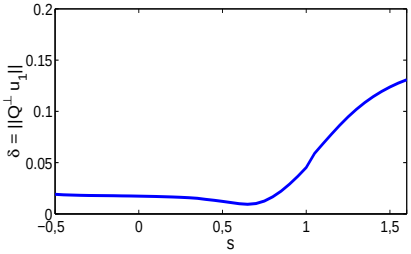

Figure 3. Projection error $\|Q^\perp u_1\|$ versus size of core sets, i.e., the parameter s.

Let us finally compare the effect of our choice of (two) core sets on the approximation error of dominant eigenvalues with the statements of Theorem 4.2

(with $m = 2$). To this end we will study the relative error

$$E_{rel}(\tau, \delta) = \frac{|\lambda_{1,\tau} - \hat{\lambda}_{1,\tau}|}{\lambda_{1,\tau}} \tag{60}$$

for different core set sizes s, see Figure 4. We observe that for small lagtimes the real relative error is significantly smaller than the upper bound (here given by the τ-independent square of the projection error $\delta = \|Q^{\perp}u_1\|$) but for larger lagtimes the upper bound and the real error are very close. As to be expected from Figure 3 the error for good core sets ($s = 0.5$) is two orders of magnitude smaller than the "not so good" core sets for $s = 2$.

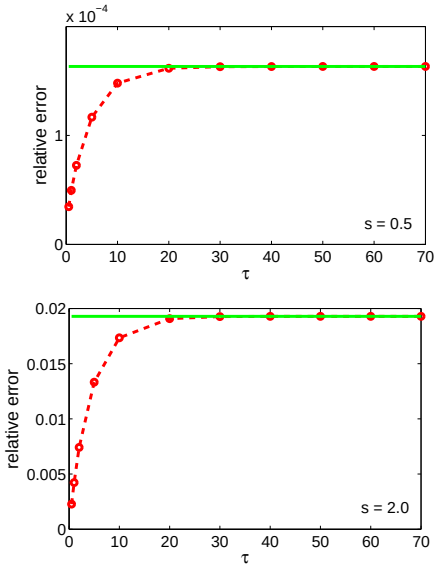

Figure 4. Relative error $E_{rel}(\tau, s)$ versus lagtime τ (dashed red line) compared to the upper bound δ^2 given by Theorem 4.2 (green solid line), for $s = 0.5$ (left hand panel) and $s = 2$ (right).

5.3. Estimation from data.

The computation of the committor functions will only be possible via FE discretization of the generator, which is infeasible in higher dimensions. This motivates to follow the instructions of Sec. 4.2 to estimate the eigenvalues from a trajectory.

We study the milestoning process $(\hat{X}_{n\tau})$ on state space $\{0, 1\}$ induced by the time-discrete process given by T_τ and the cores sets C_i^s, $i = 0, 1$.

Therefore we compute a very long trajectory $x(t)$, $t \in [0, t_{max}]$ of the diffusion process (for example based on Euler-Maruyama discretization of the SDE (56)). From this, we get discrete trajectories of the process $X_{n\tau}$ and of the

milestoning process $\hat{X}_{n\tau}$, $n = 0, \ldots, N_\tau$ with $N_\tau = \lfloor t_{max}/\tau \rfloor$. This was done based on a trajectory $x(t)$ in the time interval $[0, t_{max}]$ with $t_{max} = 50000$. Then we can estimate $\hat{T}$ and M by $\hat{T}^*_{N_\tau}$ and $M^*_{N_\tau}$ respectively as described in Sec. 4.2. The resulting non-trivial eigenvalues $\hat{\lambda}^*_1$ of the generlized eigenvalues problem $\hat{T}^*_{N_\tau} r = \hat{\lambda}^* M^*_{N_\tau} r$ are compared to the ones of $\hat{T}r = \hat{\lambda} M r$ and to the exact first non-trivial eigenvalue $\lambda_1 = \exp(\tau\Lambda_1)$ in Figure 5.

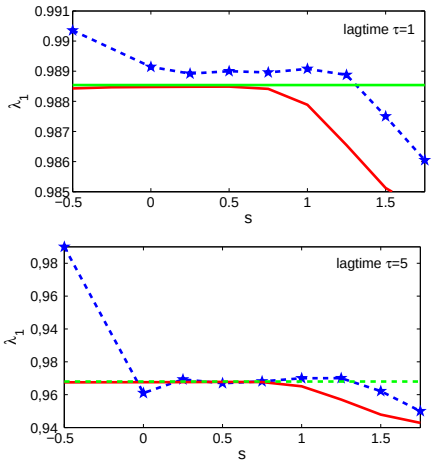

Figure 5. Comparison of the non-trivial eigenvalues λ^*_1 of the trajectory-based generlized eigenvalues problem $\hat{T}^*_{N_\tau} r = \hat{\lambda} M^*_{N_\tau} r$ (blue, dashed, stars), the ones of $\hat{T}r = \hat{\lambda} M r$ (red, solid line) and the exact first non-trivial eigenvalue $\lambda_1 = \exp(\tau\Lambda_1)$ (green, flat line) in dependence on the core size parameter s for different lagtime $\tau = 1$ (left) and $\tau = 5$ (right).

We observe that the trajectory-based eigenvalues are overestimating the "exact" eigenvalues of the generalized eigenvalue problem, and that the approximation is getting worse for small values of s, especially for larger lagtimes. This is not surprising since for $s < 0$ and sparse undersampling of the trajectory for large lagtimes, we will miss events in which the process stays close to the minima x_i without entering the cores for some time which is not long compared to the lagtime.

Despite the good approximation quality of the trajectory-based generalized eigenvalues we should not forget that they are subject to an unknown statistical sampling error resulting from the finiteness of the trajectory. Assuming that the process $(\hat{X}_n)$ is Markov and under additional assumptions on the prior [35, 36] one can show that the probability (density) that the given observation $\hat{X}_{n\tau}$, $n = 0, \ldots, N_\tau$ results from the 2×2 stochastic transition matrix $P = p_{ij}$ is given by

$$\mathbb{P}(P|\hat{X}_{n\tau}) = p_{12}^{n_{12}}(1-p_{12})^{n_{11}} p_{21}^{n_{21}}(1-p_{21})^{n_{22}},$$

with $n_{ij} = n_{ij}(N_\tau)$ as defined in Sec.4.2. We have

$$\hat{P}^*_{N_\tau} = \text{argmax}_{P \text{ stochastic matrix}} \, \mathbb{P}(P|\hat{X}_{n\tau}),$$

and for $N_\tau \to \infty$ this distribution is singularly supported in the "exact" transition matrix $\hat{P}_\tau$ of the milestoning process.

Now, let $\nu = \nu(P)$ denote an arbitrary observable that is defined in terms of the transition matrix P, e.g., the first non-trivial eigenvalue $\nu(P) = \lambda_1(P)$ or the corresponding implied timescale $\text{ITS}(P) = -\tau/\ln(\lambda_1(P))$. Then the pdf $\mathbb{P}(P|\hat{X}_{n\tau})$ on the transition matrix space and the corresponding pdf on the mass matrix space induce a pdf $\mathbb{P}(\nu|\hat{X}_{n\tau})$ on the state space of the observable. From this pdf we can compute a posteriori error indicators for the observable, e.g., the confidence intervals $I_\alpha(N_\tau)$ defined via

$$\mathbb{P}\Big(\nu \in I_\alpha(N_\tau)|\hat{X}_{n\tau}\Big) \geq \alpha.$$

In Figure 6, these confidence intervals are shown for the ITS for different values of s and $\tau = 1$.

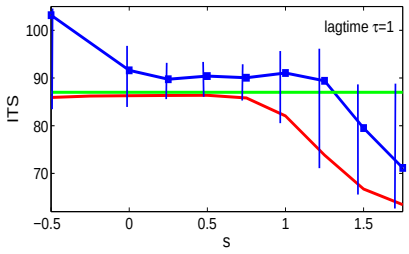

Figure 6. Implied timescale ITS and associated confidence interval I_α for $\alpha = 0.9$ of the trajectory-based generalized eigenvalue problem (blue, solid) in comparison to the ITS of "exact" transition matrices $\hat{P}_\tau$ and to the ITS of the original process versus the cores size parameter s. Lagtime $\tau = 1$.

5.4. Full partition of state space. Let us fix $m = 2$ and observe how the relative eigenvalue error E_{rel} as defined in (60) above behaves in this case, especially how does it change for different full subdivisions of the state space and different lag times. From Theorem 4.2 we know that, as above, the bound on the relative eigenvalue error is given by the square of the projection error δ. First we choose $n = 2$ and the subdivision $A_1 = (-\infty, x]$ and $A_2 = (x, \infty)$. Figures 7 and 8 show the bound δ^2 compared to the relative error $E_{rel}(\tau, \delta)$, for two different subdivisions, i.e., different values of x. We can see that the error converges to δ^2 when increasing τ. Also, a better choice of the subdivision results not only in a smaller relative error, but in its faster convergence to the bound.

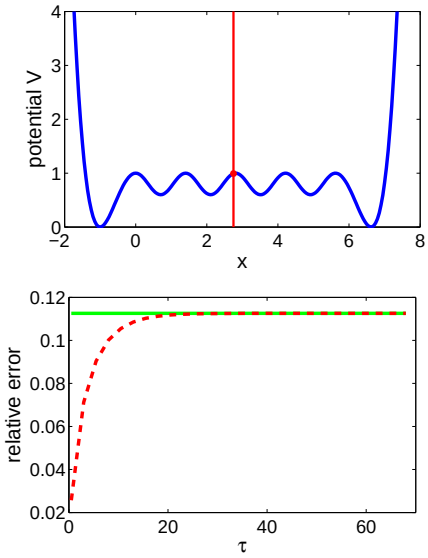

Figure 7. Relative error for eigenvalues and bound for $\tau = 0.5$, $n = 2$ and $x = 2.75$

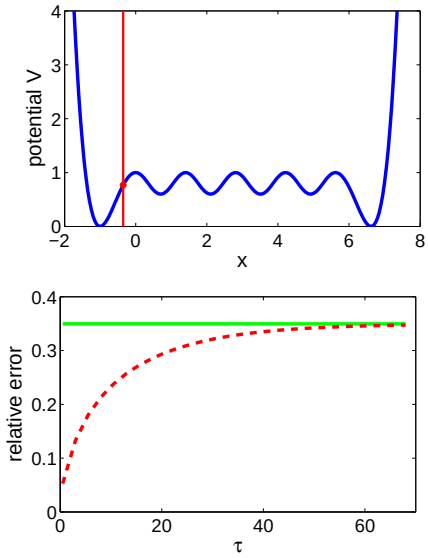

Figure 8. Relative error for eigenvalues and bound for $\tau = 0.5$, $n = 2$ and $x = -0.35$

Now we consider the full partition of a state space into $n = 6$ sets. The sets are chosen in such a way that every well belongs to one set. This choice of sets results in a smaller bound and faster convergence of the relative error to this bound, which can be seen in Figure 9.

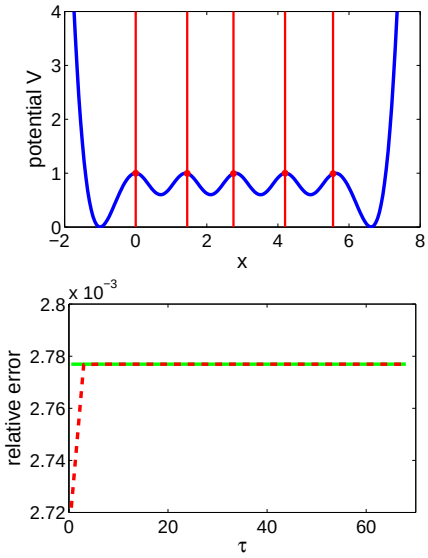

Figure 9. Relative error for eigenvalues and bound for $\tau = 0.5$ and $n = 6$

Let us finally compare the results for full subdivisions to the approximation via two core sets. We observe the following: Even the optimal full subdivision into $n = 2$ sets cannot compete with the approximation quality of the approximation based on two "reasonable/good" core sets. Good core sets result in an approximation error that is even better than the one for the optimal full subdivision into $n = 6$ sets which already resolves the well structure of the energy landscape. Thus, MSMs based on fuzzy ansatz spaces resulting from appropriate core sets and associated committor ansatz functions seem to lead to superior approximation quality than comparable full subdivision MSMs.

Conclusion

We presented a quite general approach to Markov State Models (MSM) via Galerkin projections to low-dimensional subspaces. We particularly considered the subspace spanned by the committor functions $q_1, \ldots, q_n$ defined by some core sets via the milestoning process. Our interpretation suggests that the method will work well if the space spanned by the eigenvectors corresponding to the dominant eigenvalues of the transfer operator T_t (or low-lying eigenvalues the respective generator L) is well approximated by the subspace spanned by the committor functions. In this case, the Galerkin projection QTQ of the transfer operator $T = T_\tau$ associated with the lagtime chosen will approximate well the dominant eigenvalues of T, so that the long-time behavior will be captured,

see Theorem 4.2 as well as the propagation of functions by the dynamics, see Theorem 4.1.

Technically, these theorems *do not* require that the transfer operator of the original dynamics T possesses a *spectral gap*, i.e., a group of dominant eigenvalues which are separated from all the other ones by significant interval without eigenvalues. This is in partial contrast to the usual belief: The existence of a cluster of eigenvalues close to the largest eigenvalue $\lambda = 1$ and a spectral gap is often thought of as the fundamental condition under which MSMs can have good approximation quality.

Theorems 4.2 and 4.1 need a cluster of eigenvalues close to $\lambda = 1$ since this indicates that slow processes are taking place in the original state-space. These slow processes are what the generalized eigenvalue problem is meant to capture, in the sense that the generalized eigenvalues should be close to the small eigenvalues of the original process.

However, we do *not* need the existence of a spectral gap, at least not explicitly. What we need instead is that our committor functions are good approximations of the dominant eigenvectors, i.e., that the projection error δ is small. Since the committors depend on the choice of the core sets, smallness of the projection error can only be achieved for appropriately chosen core sets.

What our approximation theorems do not tell, however, is *how to choose* the core sets, because in in general we will not be able to compute the dominant eigenvectors and committor functions that would be needed to identify the sets based on the above insight. If we assume that the original process has a cluster of eigenvalues eigenvalues close to 1 and a spectral gap, then general results guarantee the existence of a good collection of good core sets. What these sets are, however, is not given explicitly, except for the rather vague property that the process should oscillate inside and around each for a long time before visiting another and transitions to other core sets are significantly faster. How to use this criterion in a constructive way and whether a spectral gap is a necessary requirement here is the subject of current research, so we shall not dwell on these issues further here.

References

[1] W.J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, NJ, 1994.

[2] N. Madras and D. Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, 12, 2002.

[3] G. E. Cho and C. D. Meyer. Aggregation/disaggregation methods for nearly uncoupled Markov chains. *Technical Report NCSU no. 041600-0400, North Carolina State University*, 1999.

[4] E. Meerbach, C. Schuette, and A. Fischer. Eigenvalue bounds on restrictions of reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 398, 2005.

[5] C. D. Meyer. Stochastic complementation, uncoupling markov chains, and the theory of nearly reducible systems. *SIAM Rev.*, 31, 1989.

[6] R.B. Mattingly. A revised stochastic complementation algorithm for nearly completely decomposable markov chains. *ORSA Journal on Computing*, 7(2), 1995.

[7] P. Deuflhard, W. Huisinga, A. Fischer, and Ch. Schuette. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra and its Applications*, 315:39–59, 2000.

[8] P. Deuflhard and M. Weber. Robust perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398 Special issue on matrices and mathematical biology:161–184, 2005.

[9] Tijun Li, Weinan E, and Eric Vanden Eijnden. Optimal partition and effective dynamics of complex networks. *Proc. Nat. Acad. Sci.*, 105, 2008.

[10] Ch. Schuette. *Conformational Dynamics: Modelling, Theory, Algorithm, and Applications to Biomolecules*. Habilitation thesis, Fachbereich Mathematik und Informatik, FU Berlin, 1998.

[11] Ch. Schuette, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comp. Physics Special Issue on Computational Biophysics*, 151:146–168, 1999.

[12] Ch. Schuette and W. Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. In *Handbook of Numerical Analysis*, pages 699–744. Elsevier, 2003.

[13] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible markov chains. *Comm. Math. Phys.*, 228:219–255, 2002.

[14] F. Noe, Ch. Schütte, E. Vanden-Eijnden, L. Reich, and T. Weikl. Constructing the full ensemble of folding pathways from short off-equilibrium trajectories. *PNAS*, 106(45):19011–19016, 2009.

[15] F. Noé, I. Horenko, Ch. Schuette, and J. Smith. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.*, 126:155102, 2007.

[16] J. Chodera, N. Singhal, V. S. Pande, K. Dill, and W. Swope. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics*, 126, 2007.

[17] Nicaolae V. Buchete and Gerhard Hummer. Coarse master equations for peptide folding dynamics. *Journal of Physical Chemistry B*, 112:6057–6069, 2008.

[18] A. C. Pan and B. Roux. Building Markov state models along pathways to determine free energies and rates of transitions. *Journal of Chemical Physics*, 129, 2008.

[19] M. Sarich, F. Noé, and Ch. Schuette. On the approximation quality of markov state models. *to appear in Multiscale Modeling and Simulation*, 2010.

[20] A. Voter. Introduction to the kinetic Monte Carlo method. In *Radiation Effects in Solids*. Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2005.

[21] M. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*. Springer, New York, 1998.

[22] Weinan E and E. Vanden Eijnden. Metastability, conformation dynamics, and transition pathways in complex systems. In *Multiscale Modelling and Simulation*, pages 38–65. Springer, 2004.

[23] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. I. sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6:399–424, 2004.

[24] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7:69–99, 2005.

[25] W. Huisinga, S. Meyn, and Ch. Schuette. Phase transitions and metastability for Markovian and molecular systems. *Ann. Appl. Probab.*, 14:419–58, 2004.

[26] E. B. Davies. Spectral properties of metastable markov semigroups. *J. Funct. Anal.*, 52:315–329, 1983.

[27] M. Weber, S. Kube, L. Walter, and P. Deuflhard. Stable computation of probability densities for metastable dynamical systems. *Mult. Mod. Sim.*, 6(2):396–416, 2007.

[28] Anton K. Faradjian and Ron Elber. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.*, 120:10880–10889, 2004.

[29] Weinan E and E. Vanden-Eijnden. Towards a theory of transition paths. *Journal of statistical physics*, 123:503–523, 2006.

[30] P. Metzner, Ch. Schuette, and E. Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Modeling and Simulation*, 7(3):1192–1219, 2009.

[31] P. Metzner, Ch. Schuette, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.*, 125, 2006.

[32] Ch. Schütte, F. Noe, J. Lu, M. Sarich, and E. Vanden-Eijnden. Markov state model building using milestoning. *submitted to J. Chem. Phys.*, 2010. Preprint download via http://www.math.fu-berlin.de/groups/biocomputing/publications/index.html.

[33] A.V. Knyazev and M. E. Argentati. Rayleigh-ritz majorization error bounds with applications to fem. *SIAM Journal on Matrix Analysis and Applications*, 31:1521, 2010.

[34] N. Djurdjevac, M. Sarich, and Ch. Schütte. Estimating the eigenvalue error of Markov state models. *submitted to Mult. Mod. Sim.*, 2010. Preprint download via http://www.math.fu-berlin.de/groups/biocomputing/publications/index.html.

[35] Ph. Metzner, F. Noé, and Ch. Schütte. Estimating the sampling error: Distribution of transition matrices and functions of transition matrices for given trajectory data. *Phys. Rev. E*, 2008.

[36] F. Noé. Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.*, 128:244103, 2008.