MULTISCALE MODEL. SIMUL.
Vol. 10, No. 1, pp. 61–81

ESTIMATING THE EIGENVALUE ERROR OF MARKOV STATE MODELS*

NATASA DJURDJEVAC†, MARCO SARICH†, AND CHRISTOF SCHÜTTE†

Abstract. We consider a continuous-time, ergodic Markov process on a large continuous or discrete state space. The process is assumed to exhibit a number of metastable sets. Markov state models (MSMs) are designed to represent the effective dynamics of such a process by a Markov chain that jumps between the metastable sets with the transition rates of the original process. MSMs have been used for a number of applications, including molecular dynamics (cf. [F. Noé et al., *Proc. Natl. Acad. Sci. USA*, 106 (2009), pp. 19011–19016]), for more than a decade. The rigorous and fully general (no zero temperature limit or comparable restrictions) analysis of their approximation quality, however, has only recently begun. Our first article on this topics [M. Sarich, F. Noé, and Ch. Schütte, *Multiscale Model. Simul.*, 8 (2010), pp. 1154–1177] introduces an error bound for the difference in propagation of probability densities between the MSM and the original process on long timescales. Herein we provide upper bounds for the error in the eigenvalues between the MSM and the original process, which means that we analyze how well the longest timescales in the original process are approximated by the MSM. Our findings are illustrated by numerical experiments.

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

AMS subject classifications. 65C50, 60J35

DOI. 10.1137/100798910

1. Introduction.

Recent years have seen the advance of so-called Markov state models (MSMs) as low-dimensional models for ergodic Markov processes on very large, mostly continuous state spaces exhibiting metastable dynamics [1, 2, 3, 4, 5]. Recently, interest in MSMs has drastically increased since it could be demonstrated that MSMs can be constructed even for very high-dimensional systems [3] and have been especially useful for modeling the interesting slow dynamics of biomolecules [6, 7, 8, 9, 10, 11] and materials [12] (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 aggregated statistics of the original process.

In this contribution we will use the approach to MSMs via Galerkin discretization of the *transfer operator* of the original Markov process as developed in [4, 3, 2, 1] and recently addressed in detail in [13, 14]; 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.

*Received by the editors June 15, 2010; accepted for publication (in revised form) December 19, 2011; published electronically February 28, 2012. This work was supported by the DFG research center MATHEON "Mathematics for Key Technologies" in Berlin.
http://www.siam.org/journals/mms/10-1/79891.html

†Institut für Mathematik II, Freie Universität Berlin, Arnimallee 2-6, 14195 Berlin, Germany (djurdjev@math.fu-berlin.de, sarich@math.fu-berlin.de, schuette@math.fu-berlin.de).

For so-called *full partition MSMs* 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 aggregated statistics as mentioned above. However, in many cases indicator ansatz spaces do not allow good approximation quality to be obtained for reasonably small numbers of sets [11]. Therefore other ansatz spaces, e.g., fuzzy ansatz spaces, have also been discussed [15, 13, 14].

MSMs are aiming at capturing the essential dynamics of the underlying Markov process on its longest timescales. These longest timescales are endowed in the dominant eigenvalues of the transfer operator T of the underlying process. Therefore the eigenvalues of the transfer operator associated with some MSMs have to be good approximations of the dominant eigenvalues of T. Despite the growing interest in MSMs there still are only a very few rather limited rigorous results on the eigenvalue error associated with an MSM (one finds some asymptotic results in [5, 16, 17, 4], but these are of very limited algorithmic use since they depend on a smallness parameter and are valid in the limit of this parameter going to zero, i.e., in the asymptotic regime). Herein we will give rigorous results on the eigenvalue error in the form of upper bounds that hold beyond the asymptotic regime, do not assume the presence of a spectral gap, and even have an interesting consequence for the algorithmic construction of MSMs.

The remainder of the paper is organized as follows. In section 2 we introduce the setting, define transfer operators, introduce full-partition MSMs, and relate them to Galerkin projections. Then in section 3 we introduce the milestoning process, relate it to transition path theory, and analyze its transition statistics. Section 4 then discusses Galerkin projection in general and gives rigorous approximation results for eigenvalues and related timescales. Finally, the results are illustrated by numerical experiments in section 5.

2. Setting the scene. We consider a *reversible* 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

$$p_t(x,y) = \mathbb{P}[X_t = y | X_0 = x]. \tag{2.1}$$

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). In the following we always assume that (X_t) has a positive and unique invariant measure μ given by

$$\sum_x p_t(x,y)\mu(x) = \mu(y). \tag{2.2}$$

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

$$(T_t f)(y)\mu(y) = \sum_x f(x)p_t(x,y)\mu(x) \tag{2.3}$$

and set $T := T_1$ for discrete time.

In analogy, we define on L^2_μ

$$(\mathcal{L} f)(y)\mu(y) = \sum_x l(x,y)f(x)\mu(x), \tag{2.4}$$

where

$$l(x,y) = \lim_{t\to 0} \frac{p_t(x,y) - \delta_{x,y}}{t}, \tag{2.5}$$

and for the discrete case

$$\mathcal{L}_d = T - \mathrm{Id}. \tag{2.6}$$

Since (X_t) is a reversible process, it means that the detailed balance holds, i.e.,

$$\mu(x)l(x,y) = \mu(y)l(y,x), \tag{2.7}$$

and that T and $\mathcal{L}$ are self-adjoint operators in L^2_μ.

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

$$\langle f,g\rangle = \sum_x f(x)g(x)\mu(x), \qquad \|f\|^2 = \langle f,f\rangle, \qquad \|f\|_1 = \sum_x |f(x)|\mu(x), \tag{2.8}$$

and we will also call a function $f \geq 0$ a density in L^2_μ if $\|f\|_1 = 1$.

In the theory of building standard MSMs one chooses a partitioning of state space, i.e., sets $A_1, \ldots, A_n$, such that

$$A_i \cap A_j = \emptyset, \quad i \neq j, \qquad \bigcup_{i=1}^{n} A_i = S, \tag{2.9}$$

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

$$\hat{p}(i,j) = \mathbb{P}[X_\tau \in A_j | X_0 \in A_i] \tag{2.10}$$

and use the corresponding Markov chain on the index space $\{1,\ldots,n\}$ to approximate the switching behavior of the original dynamics. The conditional probability in (2.10) is taken with respect to (w.r.t.) equilibrium paths; i.e., the conditioning by $X_0 \in A_i$ simply means that X_0 is distributed according to the invariant measure restricted to A_i. The approximation quality of such MSMs is discussed in [11]. A key feature is that the transition matrix with entries $\hat{p}(i,j)$ comes out to be the matrix representation of the projection $QT_\tau Q$ of the transfer operator where Q is the orthogonal projection onto

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

As outlined above, we will not restrict our attention to full partitioning of state space. Instead, we will analyze general Galerkin projections $QT_\tau Q$ of the transfer operator where projections Q onto step-function spaces are a special case.

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 in addition to 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 [4].

3. Milestoning and transition path theory. We will now follow the approach first introduced in [14] and define sets $C_1, \dots, C_n \subset S$, which we will call *core sets*, such that

$$C_i \cap C_j = \emptyset, \quad i \neq j. \tag{3.1}$$

This means that, unlike in the standard MSM, we now relax the full partition constraint (2.9). 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* $(\hat{X}_t)$,

$$\hat{X}_t = i \Leftrightarrow X_{\sigma(t)} \in C_i, \text{ with } \sigma(t) = \sup_{s \leq t} \left\{ X_s \in \bigcup_{k=1}^{n} C_k \right\}; \tag{3.2}$$

i.e., the milestoning process is in state i if the original process came last from core set C_i; cf. [18].

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 at time 0. q_i is usually referred to as the *forward committor*; for reversible processes the forward committor is identical to the backward committor. As, for example, in [19], one can derive that q_i is the solution of

$$\begin{aligned} (\mathcal{L}q_i)(x) &= 0 \quad \forall x \in C, \\ q_i(x) &= 1 \quad \forall x \in C_i, \\ q_i(x) &= 0 \quad \forall x \in C_j, j \neq i. \end{aligned} \tag{3.3}$$

In the time-discrete case one has to replace $\mathcal{L}$ by the discrete generator $\mathcal{L}_d$. Moreover one can show that (3.3) has a unique solution under the assumption that the invariant measure is unique and not vanishing on all core sets.

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. Later, we will see that it is also *not* the crucial point whether the dynamics of the milestoning process is Markovian; for a discussion of whether the state-to-state dynamics of the original process can be reproduced by the coarse grained process with transition matrix $\hat{P}$, see Remark 4.2.

Note that the special case where we choose core sets $C_i = A_i$ that form a full partition of state space due to (2.9) is just a special case. Then the definition of the milestoning process as in (3.2) will reduce to the usual jump process between the sets A_i, that is,

$$\hat{X}_t = i \Leftrightarrow X_t \in A_i, \tag{3.4}$$

and the committors from (3.3) will be given by the characteristic functions $q_i = \mathbb{1}_{A_i}$.

We will now provide representations of the discrete or time-continuous generator of the milestoning process, respectively, in terms of the original generator and the committors. Subsequently the term *generator of the milestoning process* will be used to refer to $\hat{L}_d = \mathrm{Id} - \hat{P}$ or $\hat{L}$, respectively, being defined via the transition probabilities $\hat{p}(i,j)$ or corresponding transition rates of the milestoning process; see [13] for details. As mentioned above, the milestoning process itself in general does not have a true generator.

The following theorems from [13] give 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*

$$\hat{l}_d(i,j) = \frac{1}{\|q_i\|_1}\langle q_j, \mathcal{L}_d q_i\rangle. \tag{3.5}$$

THEOREM 3.2. *For a time-continuous process (X_t), the entries of a generator $\hat{\mathcal{L}}$ defined by the transition rates of the milestoning process $(\hat{X}_t)$ are given with*

$$\hat{l}(i,j) = \frac{1}{\|q_i\|_1}\langle \mathcal{L} q_i, q_j\rangle. \tag{3.6}$$

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.

4. Galerkin approximation. We will now discuss Galerkin projections of transfer operators. The goal is to derive a time-discrete Markovian approximation on finite state space. If (X_t) is a reversible, time-continuous Markov process with generator $\mathcal{L}$, we will fix a lag time $\tau > 0$ and consider the transfer operator

$$T_\tau = e^{\mathcal{L}\tau}. \tag{4.1}$$

The eigenvalues of the transfer operator T_τ will be given by

$$\lambda_{i,\tau} = e^{\Lambda_i \tau}, \tag{4.2}$$

where $\Lambda_i < 0$ is an eigenvalue of the generator $\mathcal{L}$. In the following we will just write $T := T_\tau$. Now we want to approximate the dynamics of (X_t) 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 and compare the operators T and QTQ. Subsequently we will only 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.

4.1. Generalized eigenvalue problem. In this section we consider subspaces $D = span\{q_1, \dots, q_n\}$ with $\mathbb{1} \in D$. The basis functions q_i are assumed to be linearly independent, nonnegative functions, do not need to be orthogonal w.r.t. $\langle \cdot, \cdot \rangle$, and are not necessarily identical with the committor functions discussed above. The orthogonal projection Q onto D can be written as

$$Qv = \sum_{i,j=1}^{n} S_{ij}^{-1} \langle v, q_j \rangle q_i, \tag{4.3}$$

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

The following theorem tells us more about the structure of the operator QTQ.

THEOREM 4.1. *Let $\hat{\lambda}$ be an eigenvalue of the operator QTQ. Then $\hat{\lambda}$ solves the generalized eigenvalue problem*

$$\hat{T} r = \hat{\lambda} M r, \tag{4.4}$$

with

$$\hat{T}_{ij} = \frac{\langle q_i, T q_j \rangle}{\hat{\mu}(i)}, \tag{4.5}$$

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

$$M_{ij} = \frac{\langle q_i, q_j \rangle}{\hat{\mu}(i)}. \tag{4.6}$$

Proof. Since

$$\hat{T}_{ij} = \frac{\langle q_i, T q_j \rangle}{\hat{\mu}(i)} = \frac{\langle q_i, (\mathrm{Id} + \mathcal{L}_d) q_j \rangle}{\hat{\mu}(i)} = (\hat{L}_d)_{ij} + M_{ij}, \tag{4.7}$$

equation (4.4) is equivalent to

$$\hat{L}_d r = (\hat{\lambda} - 1) M r. \tag{4.8}$$

Let ϕ be an eigenvector of QTQ w.r.t. $\hat{\lambda}$, i.e.,

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

This is equivalent to

$$\begin{aligned} & \langle Q\mathcal{L}_d Q\phi, q_i \rangle = (\hat{\lambda} - 1)\langle \phi, q_i \rangle \quad \forall i = 1, \dots, n \\ \Leftrightarrow \quad & \langle \mathcal{L}_d Q\phi, q_i \rangle = (\hat{\lambda} - 1)\langle \phi, q_i \rangle \quad \forall i = 1, \dots, n \\ \Leftrightarrow \quad & \sum_{j,k=1}^{n} S_{jk}^{-1} \langle \phi, q_k \rangle \langle \mathcal{L}_d q_j, q_i \rangle = (\hat{\lambda} - 1)\langle \phi, q_i \rangle \quad \forall i = 1, \dots, n. \end{aligned} \tag{4.9}$$

Introducing

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

equation (4.9) can be written as

$$\sum_{j=1}^{n} r_j \langle \mathcal{L}_d q_j, q_i \rangle = (\hat{\lambda}-1)\langle \phi, q_i \rangle = (\hat{\lambda}-1) \sum_{j,k=1}^{n} S_{ij} S_{jk}^{-1} \langle \phi, q_k \rangle = (\hat{\lambda}-1) \sum_{j=1}^{n} S_{ij} r_j. \tag{4.10}$$

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

Thus we can compute the eigenvalues of the projected transfer operator QTQ by solving the generalized eigenvalue problem (4.4). Whenever we choose the basis functions q_i to be the committor functions, then the entries $\hat{T}_{ij}$ and M_{ij} have a stochastic interpretation; cf. [13] for details. When the basis functions are chosen such that

$$q_i(x) = \mathbb{1}_{C_i}(x), \tag{4.11}$$

the sets C_i have to form a full subdivision of state space and (2.10) gives the matrix representation of QTQ. Moreover, because of orthogonality of the stepfunctions, we then have

$$M_{ij} = \frac{\langle q_i, q_j \rangle}{\hat{\mu}(i)} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{4.12}$$

4.2. Approximation of dominant eigenvalues. Our question is, how well do the eigenvalues of the projected transfer operator approximate the original eigenvalues of T? Because of self-adjointness of the transfer operator we can use the results from [20, Theorem 2.2] to show the following.

THEOREM 4.2. *Let $1 = \lambda_0 > \lambda_1 > \cdots > \lambda_{m-1}$ be the m dominant eigenvalues of T; i.e., for every other eigenvalue λ it holds that $\lambda < \lambda_{m-1}$. Let $u_0, u_1, \ldots, u_{m-1}$ be the corresponding normalized eigenvectors, $D \subset L^2_\mu$ a linear subspace with*

$$\mathbb{1} \in D, \qquad \dim(D) =: n \geq m, \tag{4.13}$$

and Q the orthogonal projection onto D.

Moreover, let $1 = \hat{\lambda}_0 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_{m-1}$ be the dominating eigenvalues of the projected operator QTQ. Then

$$E(\delta) = \max_{i=1,\ldots,m-1} |\lambda_i - \hat{\lambda}_i| \leq \lambda_1 (m-1)\, \delta^2, \tag{4.14}$$

where

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

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

Proof. The eigenvector of T w.r.t. the trivial eigenvalue $\lambda_0 = 1$ is known: $u_0 = \mathbb{1}$. Therefore

$$u_0 \in D \Rightarrow Q u_0 = u_0. \tag{4.15}$$

This implies that u_0 is also eigenvector of QTQ w.r.t. its largest eigenvalue $\hat{\lambda}_0 = 1$.

Now define

$$\Pi_0 v = \langle v, u_0 \rangle u_0, \tag{4.16}$$

set again $\Pi_0^\perp = \mathrm{Id} - \Pi_0$, and consider the operator $T\Pi_0^\perp = T - \Pi_0$. Since T is self-adjoint, its eigenvectors $u_0, u_1, \ldots$ are orthogonal, which implies that

$$T\Pi_0^\perp u_j = Tu_j - \Pi_0 u_j = Tu_j = \lambda_j u_j \qquad \forall j > 0$$

and $T\Pi_0^\perp u_0 = 0$; that is, the operator $T\Pi_0^\perp$ has the same eigenvalues with the same corresponding eigenvectors as T, only with the eigenvalue $\lambda_0 = 1$ changed to a zero eigenvalue.

Moreover,

$$\Pi_0 T\Pi_0^\perp = 0, \quad \text{and therefore} \quad T\Pi_0^\perp = \Pi_0^\perp T\Pi_0^\perp,$$

which implies self-adjointness of the operator $T\Pi_0^\perp$. Now set $U = \mathrm{span}\{u_0, \ldots, u_{m-1}\}$, and let Π be the orthogonal projection onto U. Then the operator $\Pi T\Pi_0^\perp \Pi$ has exactly the eigenvalues $\lambda_1, \ldots, \lambda_{m-1}$ and an additional eigenvalue zero, which corresponds to the eigenvector u_0.

From (4.15) it follows that $Q\Pi_0 Q = \Pi_0$ and hence

$$QT\Pi_0^\perp Q = QTQ - \Pi_0.$$

The same argument as above shows that the operator $QT\Pi_0^\perp Q$ has the same spectrum as QTQ, only with the corresponding eigenvalue of u_0 changed from $\hat{\lambda}_0 = 1$ to zero.

Using the results from [20, Theorem 2.2] we find for the error (4.14)

$$E(\delta) = \max_{i=1,\ldots,m-1} |\lambda_i - \hat{\lambda}_i| \leq (\lambda_1 - \lambda_{\min(U+D)}) \max_i \sin^2(\theta_i(U, D)), \tag{4.17}$$

with $\Theta = \Theta(U, D) = \{\theta_0, \ldots, \theta_{m-1}\}$, a vector of principal angles between the subspaces U and D. $\lambda_{\min(U+D)}$ is the smallest eigenvalue of the operator ZTZ, where Z is an orthogonal projection on the space $U + D$. In our case this means $\lambda_{\min(U+D)} = 0$. Let $\sigma_i(A)$ and $\Lambda_i(B)$ denote the ith singular value of operator A and ith eigenvalue of operator B, respectively. The principal angles are defined as $\cos(\theta_i) = \sigma_i(Q\Pi)$. Moreover, the definition of leading singular values yields

$$\sigma_i^2(Q\Pi) = \Lambda_i((Q\Pi)^* Q\Pi) = \Lambda_i(\Pi Q\Pi), \quad i = 1, \ldots, m-1, \tag{4.18}$$

where $(Q\Pi)^*$ denotes the adjoint of $(Q\Pi)$ in L^2_μ, in which sense $Q^*Q = Q$ also holds. We get

$$\sin^2(\theta_i) = 1 - \cos^2(\theta_i) = 1 - \Lambda_i(\Pi Q\Pi) = \Lambda_i(\Pi - \Pi Q\Pi) = \Lambda_i(\Pi Q^\perp \Pi). \tag{4.19}$$

As in (4.18),

$$\Lambda_i(\Pi Q^\perp \Pi) = \sigma_i^2(Q^\perp \Pi) \leq \|Q^\perp \Pi\|^2. \tag{4.20}$$

Now let v, $\|v\| = 1$ be arbitrary. If we define $\hat{v} \in \mathbb{R}^{m-1}$ as

$$\hat{v}_j = \langle v, u_j \rangle, \quad j = 1, \ldots, m-1,$$

and denote the usual 1- and 2-norms on $\mathbb{R}^{m-1}$ by $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively, we find immediately that

$$\sum_{j=1}^{m-1} |\langle v, u_j \rangle| = \|\hat{v}\|_1 \leq \sqrt{m-1}\|\hat{v}\|_2 = \sqrt{m-1} \left(\sum_{j=1}^{m-1} \langle v, u_j \rangle^2 \right)^{1/2} \leq \sqrt{m-1}. \tag{4.21}$$

Since $Q^\perp u_0 = 0$,

$$\begin{aligned}\|Q^\perp \Pi v\| &= \left\| \sum_{j=1}^{m-1} \langle v, u_j \rangle Q^\perp u_j \right\| \leq \sum_{j=1}^{m-1} |\langle v, u_j \rangle| \|Q^\perp u_j\| \\ &\leq \sum_{j=1}^{m-1} |\langle v, u_j \rangle| \delta \leq \sqrt{m-1} \cdot \delta.\end{aligned} \tag{4.22}$$

Combining (4.19), (4.20), and (4.22), we obtain

$$\sin^2(\theta_i) \leq \|Q^\perp \Pi\|^2 \leq (m-1)\,\delta^2. \tag{4.23}$$

Putting everything together gives (4.14). ∎

Remark 4.1. Inserting (4.2) into (4.14), we get the lag time–dependent eigenvalue estimate

$$E(\tau, \delta) = \max_{i=1,\ldots,m-1} |\lambda_i - \hat{\lambda}_i| \leq e^{\Lambda_1 \tau}(m-1)\,\delta^2, \tag{4.24}$$

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 $\Lambda_1 < 0$,

$$E(\tau, \delta) \to 0 \quad \text{for } \tau \to \infty, \tag{4.25}$$

which stems from the asymptotic convergence to the invariant measure. Furthermore, for the *relative eigenvalue error* we have, at least for the first nontrivial eigenvalue,

$$\frac{|\lambda_1 - \hat{\lambda}_1|}{|\lambda_1|} \leq (m-1)\,\delta^2, \tag{4.26}$$

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

Remark 4.2. In [11], we analyzed how well the state-to-state dynamics of the original process with transfer operator T can be approximated by the coarse grained process with transfer operator QTQ. An upper bound for the propagation error $E(k) = \|QT^kQ - (QTQ)^k\|$ is provided that shows that, e.g., $\max_k E(k)$ is small if the projection error δ and the additional constant $\eta = \exp(-\tau|\Lambda_m - \Lambda_1|)$ are both small. Thus, the question of whether the state-to-state of the original process can be approximated well can be discussed independently of whether the milestoning process is Markovian or not. Furthermore, whenever $E(k)$ is small, all errors in computing time-correlation functions will be small also; see [11] for details.

Remark 4.3. The above result does not require any specific assumptions about spectral gaps or comparable quantities. Since this may seem strange we want to add two comments: First, there are a variety of results for metastable processes that show that the existence of a spectral gap (that is, the existence of a group of dominant eigenvalues which are separated from all the other ones by significant interval without eigenvalues) leads to a small projection error δ, for example, for diffusion processes in multiwell potentials; see [21, 16, 17]. Second, there are also cases with small δ for original dynamics with wide spectrum without any significant spectral gaps [6].

This frequent observation is most easily illustrated by the 3-state Markov chain with transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 1/2 & 0 & 1/2 \\ 0 & \alpha & 1-\alpha \end{pmatrix},$$

with some $\alpha \in (0,1)$. Its eigenvalues are $\lambda = 1, 1-\alpha, -\alpha$ with a clear Perron cluster and a spectral gap for α close to 0 and no gap for α closer to 1. We have $\mu = 1/3 \cdot (1,1,1)$ and $u_1 = \sqrt{3/2} \cdot (1,0,-1)$. With core sets $C_1 = \{1\}$ and $C_2 = \{3\}$ we easily compute $Qu_1 = u_1$ such that $\delta = 0$ for $m = 2$ *independent* of α. In fact, QTQ has the matrix representation

$$\begin{pmatrix} 1-\alpha/2 & \alpha/2 \\ \alpha/2 & 1-\alpha/2 \end{pmatrix},$$

with eigenvalues $\hat{\lambda} = 1, 1-\alpha$.

Remark 4.4. As outlined in the introduction we are mainly interested in building MSMs for metastable processes. Thus it is of interest whether the existence of metastable sets may induce a small projection error δ. Let us first consider the simple situation of a reversible finite state space Markov chain with two disjoint metastable sets A and B. Metastability of A and B is related to the expected return time to $M = A \cup B$ if starting in the transition region between A and B,

$$R = \max_{y \notin M} \mathbb{E}(\tau(M) | X_0 = y),$$

and to the expected transition times between A and B,

$$W_{AB} = \mathbb{E}(\tau(B) | X_0 \in A), \qquad W_{BA} = \mathbb{E}(\tau(A) | X_0 \in B),$$

where $\tau(C)$ denotes the first entry time into set C. Following [5] we call A and B metastable if

$$r = \max\left\{ \frac{R}{W_{AB}}, \frac{R}{W_{BA}} \right\} \ll 1.$$

For this and more general situations [22] provides an upper bound on δ which is small as long as r is small such that the core sets are metastable sets. This shows that metastability implies small projection error as long as the core sets are chosen appropriately.

Our next question is, how well do the eigenvalues of the projected generator $Q\mathcal{L}Q$ approximate the original eigenvalues of $\mathcal{L}$? Because the generator $\mathcal{L}$ is self-adjoint and its spectrum $\sigma(\mathcal{L})$ is nonpositive, setting $A = \alpha \mathrm{Id} - \mathcal{L}$ with an arbitrary scalar $\alpha > 0$ such that $\alpha \notin \sigma(\mathcal{L})$ defines a *positive definite, self-adjoint* operator that has the same eigenvectors as $\mathcal{L}$. We will see that we need the scalar product induced by A in L^2_μ, being defined via

$$\langle u, v \rangle_A = \langle u, Av \rangle.$$

We can use different results from [20] to show the following.

THEOREM 4.3. *Let $0 = \Lambda_0 > \Lambda_1 > \cdots > \Lambda_{m-1}$ be the m largest eigenvalues of $\mathcal{L}$; i.e., for every other eigenvalue Λ it holds that $\Lambda < \Lambda_{m-1}$. Let $u_0, u_1, \ldots, u_{m-1}$ be the corresponding normalized eigenvectors, $D \subset L^2_\mu$ a linear subspace with*

$$\mathbb{1} \in D, \qquad \dim(D) =: n \geq m, \tag{4.27}$$

Q_A the orthogonal projection onto $A^{1/2}D$ w.r.t. $\langle \cdot, \cdot \rangle$ (see below for details), and Q the orthogonal projection onto D w.r.t. $\langle \cdot, \cdot \rangle$. Moreover, let $0 = \hat{\Lambda}_0 > \hat{\Lambda}_1 > \cdots > \hat{\Lambda}_{m-1}$ be the m dominant eigenvalues of the projected operator $Q\mathcal{L}Q$.

Then we have that $|\hat{\Lambda}_i| \geq |\Lambda_i|$ for $i = 0, \ldots, m-1$, and for every positive scalar ϵ the following estimate holds:

$$E_{\mathcal{L}} = \max_{i=1,\ldots,m-1} \frac{|\Lambda_i - \hat{\Lambda}_i|}{|\hat{\Lambda}_i|} \leq (1+\epsilon)\,(m-1)\,\delta_A^2, \tag{4.28}$$

where

$$\delta_A = \max_{i=1,\ldots,m-1} \|Q_A^{\perp} u_i\|$$

with $A = \epsilon|\Lambda_1|\mathrm{Id} - \mathcal{L}$, is the maximal projection error of the eigenvectors to the space D w.r.t. the scalar product induced by A.

Proof. Set $A = \alpha \mathrm{Id} - \mathcal{L}$ for some $\alpha > 0$ such that $\alpha \notin \sigma(\mathcal{L})$. The nonpositive eigenvalues $\Lambda \in \sigma(\mathcal{L})$ of $\mathcal{L}$ induce positive eigenvalues $\Lambda^A = \alpha - \Lambda$ of A with identical eigenvectors. Therefore the eigenvalues $0 < \alpha = \Lambda_0^A < \Lambda_1^A < \cdots < \Lambda_{m-1}^A$ of A are associated with the largest eigenvalues of $\mathcal{L}$, and $U = \mathrm{span}\{u_0, \ldots, u_{m-1}\}$ is an A-invariant m-dimensional subspace associated with the smallest eigenvalues of A.

Let Π be the orthogonal projection onto U w.r.t. $\langle \cdot, \cdot \rangle$, and let Q be the orthonormal projection onto D, again w.r.t. $\langle \cdot, \cdot \rangle$. Then the m smallest eigenvalues of QAQ are $\alpha = \hat{\Lambda}_0^A < \hat{\Lambda}_1^A < \cdots < \hat{\Lambda}_{m-1}^A$ with $\hat{\Lambda}_{m-1}^A = \alpha - \hat{\Lambda}_{m-1}$.

Using the results from [20, Theorem 2.5], we find

$$\max_{i=1,\ldots,m-1} \frac{|\Lambda_i^A - \hat{\Lambda}_i^A|}{\hat{\Lambda}_i^A} \leq \max_i \sin^2(\theta_{i,A}(U, D)), \tag{4.29}$$

with $\Theta_A = \Theta_A(U, D) = \{\theta_{0,A}, \ldots, \theta_{m-1,A}\}$, a vector of principal angles between the subspaces U and D w.r.t. $\langle \cdot, \cdot \rangle_A$. Furthermore one finds there that $0 \leq \Lambda_i^A \leq \hat{\Lambda}_i^A$, from which it immediately follows that $|\hat{\Lambda}_i| \geq |\Lambda_i|$.

Let us again assume that the subspace D is given by $D = \mathrm{span}\{q_1, \ldots, q_n\}$, where the q_i are linearly independent and not necessarily orthogonal functions.

According to [23, Theorem 2.9] the values $\sin^2 \theta_{i,A}(U, D)$ can be computed as follows: Let $A^{1/2}$ denote the square root of A, and consider the subspaces $A^{1/2}U = \mathrm{span}\{u_1, \ldots, u_{m-1}\} = U$ and $A^{1/2}D = \mathrm{span}\{A^{1/2}q_1, \ldots, A^{1/2}q_n\}$. Then

$$\sin^2 \theta_{i,A}(U, D) = \sin^2 \theta_i(A^{1/2}U, A^{1/2}D),$$

where the angles $\theta_i(A^{1/2}U, A^{1/2}D)$ are defined via the original scalar product $\langle \cdot, \cdot \rangle$ and can be computed as in the previous proof.

Using the same tricks as in the previous proof and analogous arguments, we thus get

$$\sin^2(\theta_{i,A}) \leq (m-1) \cdot \delta_A^2, \tag{4.30}$$

where $\delta_A = \max_j \|Q_A^\perp u_j\|$, with Q_A denoting the orthogonal projection w.r.t. the original scalar product onto $A^{1/2}D$, i.e.,

$$Q_A v = \sum_{ij} (S_A^{-1})_{ij} \langle A^{1/2} q_j, v \rangle A^{1/2} q_i, \quad S_{A,ij} = \langle A^{1/2} q_i, A^{1/2} q_j \rangle = \langle q_i, q_j \rangle_A.$$

Putting everything together gives

$$\max_{i=1,\dots,m-1} \frac{|\Lambda_i^A - \hat{\Lambda}_i^A|}{\hat{\Lambda}_i^A} \leq (m-1) \cdot \delta_A^2. \tag{4.31}$$

Furthermore we have

$$\frac{1}{\hat{\Lambda}_i^A} = \frac{1}{|\hat{\Lambda}_i|} \frac{1}{|1 - \alpha/\hat{\Lambda}_i|}.$$

However, all the while our positive scalar α has been arbitrary, so that

$$\epsilon = \frac{\alpha}{|\Lambda_1|} \leq -\frac{\alpha}{\hat{\Lambda}_1} < -\frac{\alpha}{\hat{\Lambda}_i}$$

is some arbitrarily small positive scalar with

$$\frac{1}{\hat{\Lambda}_i^A} = \frac{1}{|\hat{\Lambda}_i|} \frac{1}{1+\epsilon}.$$

Putting this and $|\Lambda_i^A - \hat{\Lambda}_i^A| = |\Lambda_i - \hat{\Lambda}_i|$ into (4.31) finally yields (4.28). ∎

Remark 4.5. Starting with

$$Q_A v = A^{1/2} \sum_{ij} (S_A^{-1})_{ij} \langle q_j, A^{1/2} v \rangle q_i, \quad S_{A,ij} = \langle q_i, q_j \rangle_A,$$

we can use the orthonormal projection onto D w.r.t. $\langle \cdot, \cdot \rangle_A$,

$$\mathcal{Q}_A v = \sum_{ij} (S_A^{-1})_{ij} \langle q_j, v \rangle_A q_i,$$

to get $Q_A u_j = |\Lambda_j^A|^{-1/2} A^{1/2} \mathcal{Q}_A u_j$, and thus

$$\|Q_A u_j\| = (\Lambda_j^A)^{-1/2} \|\mathcal{Q}_A u_j\|_A.$$

Therefore, with the A-orthonormal basis $\phi_j = u_j / \sqrt{\Lambda_j^A}$ of U we find $\|Q_A u_j\| = \|\mathcal{Q}_A \phi_j\|_A$, and therefore

$$\delta_A = \|Q_A^\perp u_j\| = \|\mathcal{Q}_A^\perp \phi_j\|_A,$$

since $\|Q_A^\perp u_j\|^2 = \|u_j\|^2 - \|Q_A u_j\|^2 = \|\phi_j\|_A^2 - \|\mathcal{Q}_A \phi_j\|_A^2 = \|\mathcal{Q}_A^\perp \phi_j\|_A^2$.

5. Illustrative examples.

5.1. Double well potential with diffusive transition region. We consider the diffusion process

$$\gamma dX_t = -\nabla V(X_t) dt + \sqrt{2\beta^{-1}\gamma} dB_t, \tag{5.1}$$

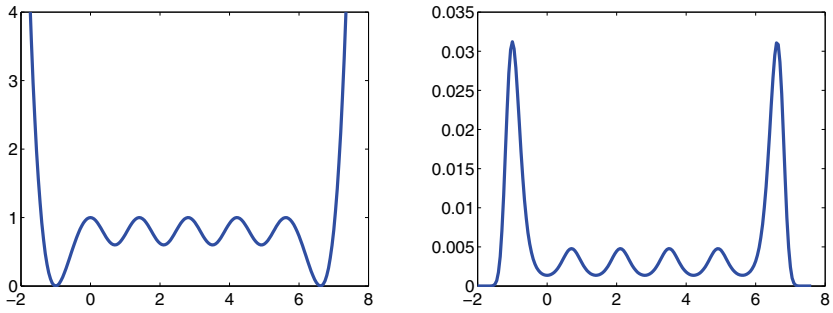

FIG. 5.1. *The potential V with extended transition region and the associated invariant measure for $\sigma = 0.8$.*

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 5.1; 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 each have their own, less pronounced metastability. 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 w.r.t. x and Δ_x the associated Laplacian. Thus, the transfer operator reads

$$T_t = \exp(t\mathcal{L}). \tag{5.2}$$

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 MSMs. More precisely, we first choose a lag time τ and consider the transfer operator T_τ. Because of (4.2) we can compute the ITS

$$|1/\Lambda_1| = -\frac{\tau}{\ln(\lambda_{1,\tau})}, \tag{5.3}$$

where $\lambda_{1,\tau} < 1$ is the largest nontrivial 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; it should be obvious that the sets $\tilde{C}_i^s = [x_i - s, x_i + s]$ would define exactly the same milestoning process such that we can talk of *small* core sets for small values of s. In what follows, we consider the subspace D that is spanned by the committor functions defined by the core sets C_i^s, and we denote by Q the associated orthogonal projection.

Next we compare the ITS from (5.3) to the one that corresponds to the largest nontrivial eigenvalue $\hat{\lambda}_{i,\tau}$ of the projected operator $QT_\tau Q$,

$$|1/\hat{\Lambda}_1| = -\frac{\tau}{\ln(\hat{\lambda}_{1,\tau})}. \tag{5.4}$$

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 us to compute the eigenvalues of $QT_\tau Q$ as in Theorem 4.1. Figure 5.2 shows the dependence of the nontrivial eigenvalue on the core set size s for different values of the lag time τ.

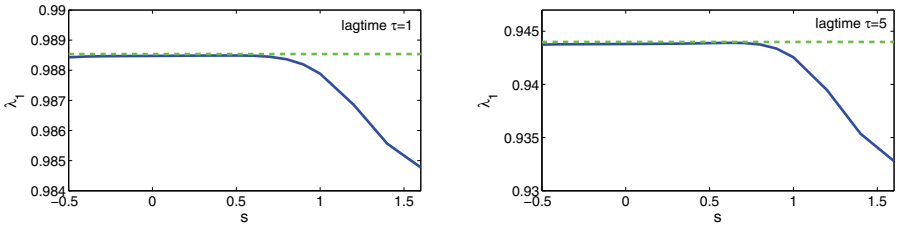

FIG. 5.2. *Nontrivial eigenvalues $\lambda_{1,\tau}^s < 1$ of the generalized eigenvalue problem $\hat{T}_\tau r = \hat{\lambda} M r$ versus core set size parameter s for lag times $\tau = 1$ (left) and $\tau = 5$ (right) in comparison to the exact first nontrivial eigenvalue* $\exp(\tau\Lambda_1)$.

We observe that for small enough core sets the approximation of the exact first nontrivial 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.

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

$$E_{rel}(\tau, \delta) = \frac{|\lambda_{1,\tau} - \hat{\lambda}_{1,\tau}|}{\lambda_{1,\tau}} \tag{5.5}$$

for different core set sizes s; see Figure 5.3. We observe that for small lag times 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 lag times the upper bound and the real error are very close.

Last but not least, Figure 5.4 presents the comparison between relative eigenvalue error and upper bound as of Theorem 4.3 based on generators instead of transfer operators. Again we observe impressively small deviations, which shows that the upper bound incorporates the main aspects of the underlying error. In addition we again see that the relative error increases significantly with increasing core set size s.

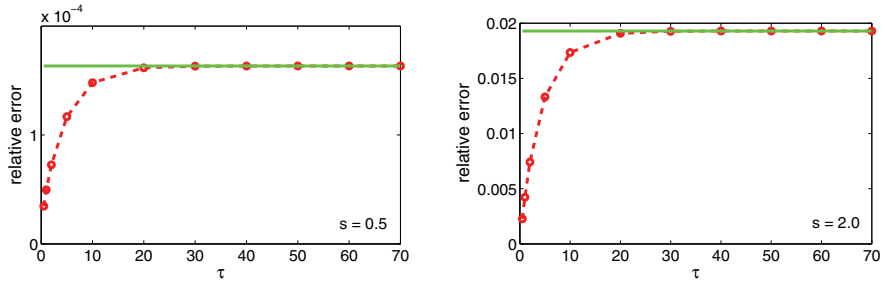

FIG. 5.3. *Relative error $E_{rel}(\tau, s)$ versus lag time τ (dashed line) compared to the upper bound δ^2 given by Theorem 4.2 (solid line) for $s = 0.5$ (left) and $s = 2$ (right).*

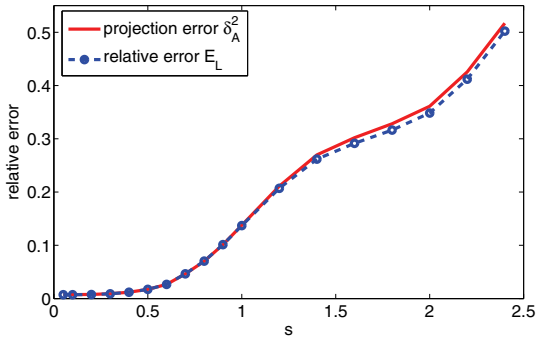

FIG. 5.4. *Projection error $\|Q_A^{\perp} u_1\|^2$ (solid line) and relative eigenvalue error E_L for the generator eigenvalues (dashed line) versus size of core sets, i.e., the parameter s. (Results are insensitive to changes in the parameter ϵ in Theorem 4.3 for small enough values of ϵ.)*

Despite this observation, Figure 5.4 demonstrates that the approximation quality is rather *robust* w.r.t. changes of the core sets as long as the core sets are not extended beyond the vicinity of the respective main wells of the energy landscape. This last observation can be made for many similar systems (cf. [5], or [14] for high dimensions): Core sets that are part of the attractive basin around the main wells in the energy landscape lead to rather good approximation of the longest timescales.

5.3. Full partition of state space. Let us fix $m = 2$ and observe how the relative eigenvalue error E_{rel} as defined in (5.5) above behaves in this case, especially 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 5.5 and 5.6 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 the respective value of δ^2 for increasing τ. Also, a better choice of the subdivision results not only in a smaller relative error, but also in its faster convergence to the bound. This is important because in practice one would like to use the smallest possible lag times in order to limit the simulation length needed to parametrize the coarse grained model. That is, for simulation purposes good approximation quality is required for rather small lag times, despite the analytical insight that longer lag times will always improve the approximation quality.

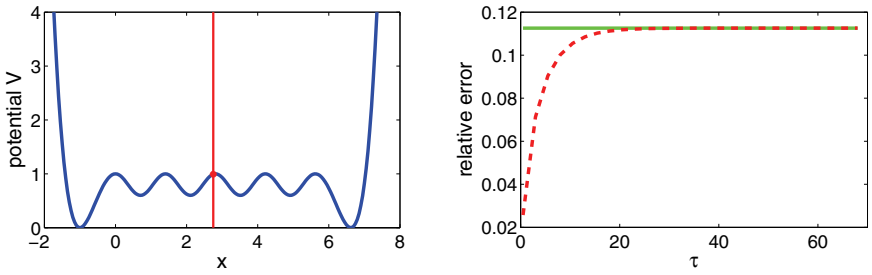

FIG. 5.5. *Relative error for eigenvalues and bound for* $\tau = 0.5$, $n = 2$, *and* $x = 2.75$.

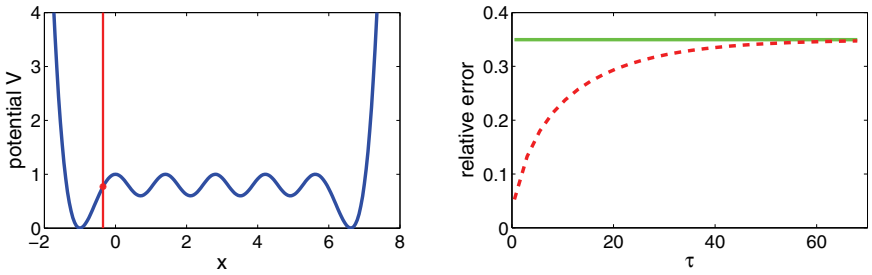

FIG. 5.6. *Relative error for eigenvalues and bound for* $\tau = 0.5$, $n = 2$, *and* $x = -0.35$.

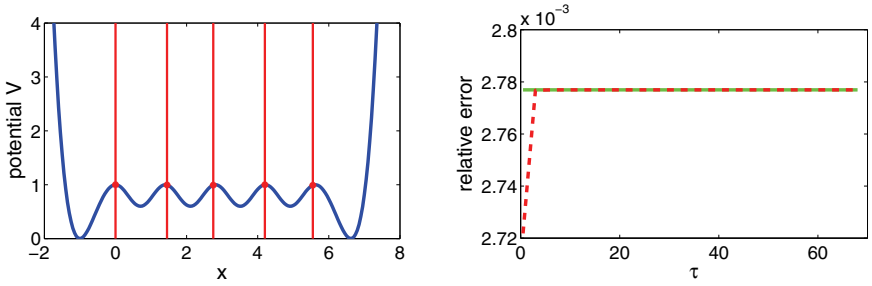

FIG. 5.7. *Relative error for eigenvalues and bound for* $\tau = 0.5$ *and* $n = 6$.

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 5.7.

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 a better approximation quality than comparable full subdivision MSMs, especially in the presence of extended transition regions. Particularly if one is interested in a small coarse grained model for the dynamics on the

slowest timescales, core set MSMs seem to provide the opportunity to construct such models without having to add sets inside of the transition region, which would cause an increase in size of the resulting Markov model.

5.4. Three well potential. In this example we will study the influence of noise in (5.1) on the choice of core sets and the approximation quality of slow timescales. Moreover, we will now consider a two-dimensional diffusion process as in (5.1) with $\gamma = 1$ and $\beta = 6.67$. The potential and its invariant measure are illustrated in Figure 5.8.

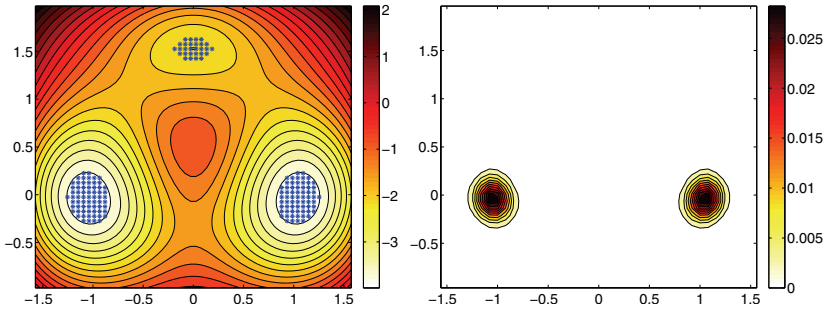

FIG. 5.8. *Left: Levelsets of potential V and indication of chosen core sets (small grids in the wells of the energy landscape). Right: Invariant measure (the peak in the third well of the energy landscape is below the threshold of visibility in this map).*

The eigenvalues of the corresponding generator are given by

Λ_0	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6
$+0.0000$	-0.0000003	-0.0463	-2.793	-4.939	-5.3301	-6.5049

Motivated by the results above and the visualization of the second and third eigenvectors in Figure 5.9, three core sets have been chosen around the local minima of the potential, as illustrated in Figure 5.8.

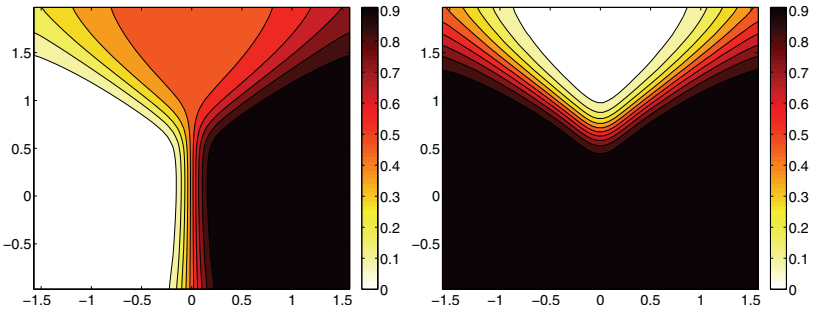

FIG. 5.9. *Left: Second eigenvalue u_1. Right: Third eigenvalue u_2. The map of the third eigenvector seems to show that it is nonnegative in the region colored white; this is not true since the eigenvalue has small negative values there.*

One should note that the second and third eigenvalues differ by a factor of 10^5. Together with the image of the invariant measure in Figure 5.8 being concentrated around the two main wells for small noise, one would typically choose only two core sets. Nevertheless we introduce a small third core set around the third

minimum, such that the eigenvectors are almost constant on the chosen core sets and the projection errors to the space spanned by the committors are small, i.e., $\|Q^{\perp}u_1\| \leq 0.00002, \|Q^{\perp}u_2\| \leq 0.005$. Therefore, we can even approximate the third slowest timescale corresponding to Λ_2. If we were interested in the slowest timescale only, it would be possible to choose two rather than three core sets, and we would get an insignificantly better approximation. Now Figure 5.10 shows the two slowest timescales of the original process: the approximation by the timescales from (5.4) and the bound from Theorem 4.2.

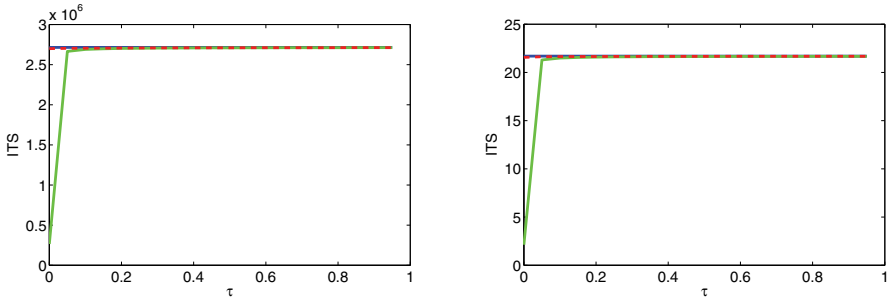

FIG. 5.10. *ITS of original generator L (dark solid line), ITS estimate (dashed line) as in* (5.4)*, and bound from Theorem* 4.2 *(light solid line). Left: ITS* $1/\Lambda_1$*. Right: ITS* $1/\Lambda_2$*.*

Increasing the noise. Finally we perform the same experiment for the three well potential as above, but we increase the noise intensity (and thus the temperature) by setting $\sigma = 1.1$ ($\beta = 1.67$). The eigenvalues of the corresponding generator now take the form

Λ_0	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5	Λ_6
$+0.0000$	-0.0818	-0.7809	-3.9230	-5.4286	-6.7504	-7.001

That is, the gap between the slowest timescales has closed, such that Λ_1 and Λ_2 differ only by a factor of 10^1 now. In this situation one could be interested in an approximation of the third timescale as well. Moreover, the invariant measure (Figure 5.11) is not completely concentrated in the two main wells anymore, but the regions around the wells have grown and the third well also carries significant invariant measure.

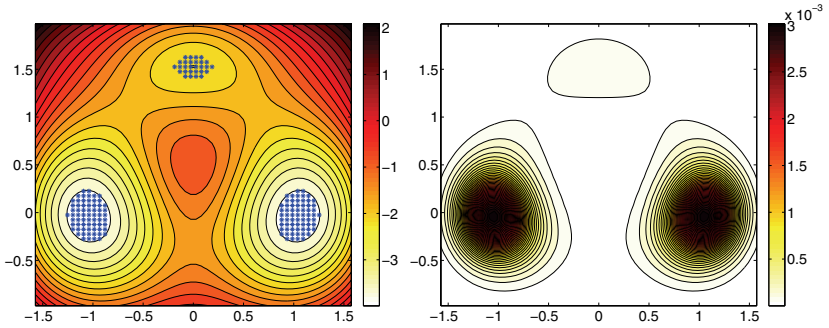

FIG. 5.11. *Left: Levelsets of potential V. Right: Invariant measure.*

On the other hand Figure 5.12 shows that one has to be more careful with the introduction of a third core set, because the variation of the second eigenvector u_1 increases in the region around the third local minimum. That is, we have to keep this third core set small in order to avoid introducing a large projection error of the second eigenvector to the committors, which would yield a worse approximation of the slowest timescale $1/\Lambda_1$. Nevertheless the projection errors to the space spanned by the committors increase, i.e., $\|Q^{\perp}u_1\| \leq 0.0086, \|Q^{\perp}u_2\| \leq 0.0911$.

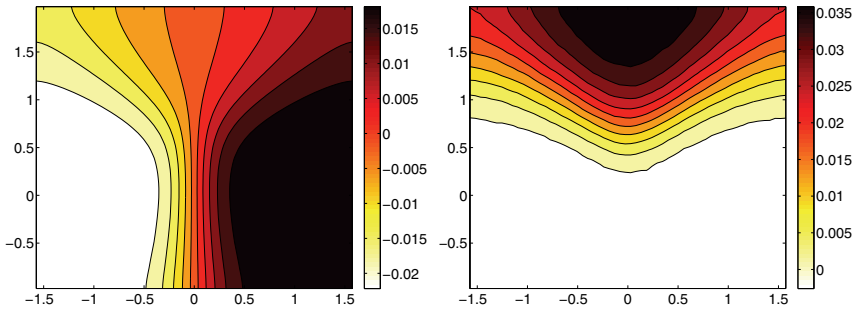

FIG. 5.12. *Left: Second eigenvalue* u_1. *Right: Third eigenvalue* u_2. $\sigma = 1.1$.

This results in a good, but slightly worse, approximation quality of the timescales compared to the small noise situation, as one can see by comparing Figures 5.13 and 5.10.

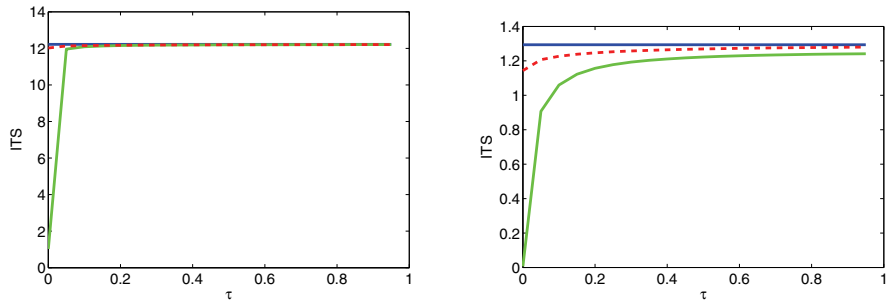

FIG. 5.13. *ITS of original generator* L *(dark solid line), ITS estimate (dashed line) as in* (5.4), *and bound from Theorem* 4.2 *(light solid line). Left: ITS* $1/\Lambda_1$. *Right: ITS* $1/\Lambda_2$.

Conclusion. We presented a quite general estimate for the approximation quality of the dominant eigenvalues of an ergodic, metastable Markov process by Markov state models (MSMs). We employed the approach via Galerkin projections to low-dimensional subspaces, and particularly considered subspaces D spanned by the committor functions defined by some core sets via the milestoning process. Our interpretation suggests that the associated MSM will approximate the dominant eigenvalues well if the space spanned by the corresponding eigenvectors of the transfer operator T_t (or low-lying eigenvalues of the respective generator L) is well approximated by the ansatz space D. In this case, the Galerkin projection QTQ of the transfer operator (or of the generator, respectively) onto D captures the long-time behavior of the original process well.

Technically, our theorems *do not* require that the transfer operator/generator of the original dynamics T possess a *spectral gap*, i.e., a group of dominant eigenvalues which are separated from all the other ones by a 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. What we need instead is that our committor functions be good approximations of the dominant eigenvectors. Since the committors depend on the choice of the core sets, smallness of the projection error can be achieved only for appropriately chosen core sets.

What our approximation theorems do not tell, however, is *how to choose* the core sets, because in general we will not be able to compute the dominant eigenvectors and committor functions (such that we cannot compute the respective projection errors δ or δ_A) that would be needed to identify the sets based on the above insight. The results presented herein can thus only guide the investigation of how to choose core sets optimally. Algorithmic research will therefore have to concentrate on estimating the projection error based on trajectories of the underlying dynamics. Some of these issues are discussed in [14], where the interested reader may also find an application to a high-dimensional example from molecular dynamics.

Acknowledgment. We are indebted to P. Metzner for providing us with the three well potential discretization tool.

REFERENCES

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

[2] P. Deuflhard, W. Huisinga, A. Fischer, and Ch. Schuette, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra Appl., 315 (2000), pp. 39–59.

[3] Ch. Schuette, A. Fischer, W. Huisinga, and P. Deuflhard, *A direct approach to conformational dynamics based on hybrid Monte Carlo*, J. Comput. Phys. (Special Issue on Computational Biophysics), 151 (1999), pp. 146–168.

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

[5] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, *Metastability and low lying spectra in reversible Markov chains*, Comm. Math. Phys., 228 (2002), pp. 219–255.

[6] F. Noé, Ch. Schütte, E. Vanden-Eijnden, L. Reich, and T. Weikl, *Constructing the full ensemble of folding pathways from short off-equilibrium trajectories*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 19011–19016.

[7] 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 (2007), 155102.

[8] 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*, J. Chem. Phys., 126 (2007), 155101.

[9] N. V. Buchete and G. Hummer, *Coarse master equations for peptide folding dynamics*, J. Phys. Chem. B, 112 (2008), pp. 6057–6069.

[10] A. C. Pan and B. Roux, *Building Markov state models along pathways to determine free energies and rates of transitions*, J. Chem. Phys., 129 (2008), 064107.

[11] M. Sarich, F. Noé, and Ch. Schütte, *On the approximation quality of Markov state models*, Multiscale Model. Simul., 8 (2010), pp. 1154–1177.

[12] A. Voter, *Introduction to the kinetic Monte Carlo method*, in Radiation Effects in Solids, K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga, eds., Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2005, pp. 1–23.

[13] N. Djurdjevac, M. Sarich, and Ch. Schütte, *On Markov state models for metastable processes*, in Proceedings of the International Congress of Mathematicians 2010 (ICM 2010), Vol. 4, World Scientific, 2010, pp. 3105–3131.

[14] Ch. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden, *Markov state models based on milestoning*, J. Chem. Phys., 134 (2011), 204105.

[15] M. Weber, S. Kube, L. Walter, and P. Deuflhard, *Stable computation of probability densities for metastable dynamical systems*, Multiscale Model. Simul., 6 (2007), pp. 396–416.

[16] 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 (2004), pp. 399–424.

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

[18] A. K. Faradjian and R. Elber, *Computing time scales from reaction coordinates by milestoning*, J. Chem. Phys., 120 (2004), pp. 10880–10889.

[19] P. Metzner, Ch. Schütte, and E. Vanden-Eijnden, *Transition path theory for Markov jump processes*, Multiscale Model. Simul., 7 (2009), pp. 1192–1219.

[20] A. V. Knyazev and M. E. Argentati, *Rayleigh–Ritz majorization error bounds with applications to FEM*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1521–1537.

[21] M. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 1998.

[22] M. Sarich and Ch. Schütte, *Approximating selected non-dominant timescales by Markov state models*, submitted to CMS, 2011; preprint download via http://publications.mi.fu-berlin.de/1113/.

[23] M. E. Argentati, *Principal Angles between Subspaces as Related to Rayleigh Quotient and Rayleigh Ritz Inequalities with Applications to Eigenvalue Accuracy and an Eigenvalue Solver*, Ph.D. thesis, University of Colorado, Denver, 2003.