SIAM J. MATH. DATA SCI.
Vol. 3, No. 1, pp. 225–252

Error Bounds for Dynamical Spectral Estimation*

Robert J. Webber†, Erik H. Thiede‡, Douglas Dow§, Aaron R. Dinner‡, and Jonathan Weare†

Abstract. Dynamical spectral estimation is a well-established numerical approach for estimating eigenvalues and eigenfunctions of the Markov transition operator from trajectory data. Although the approach has been widely applied in biomolecular simulations, its error properties remain poorly understood. Here we analyze the error of a dynamical spectral estimation method called "the variational approach to conformational dynamics" (VAC). We bound the approximation error and estimation error for VAC estimates. Our analysis establishes VAC's convergence properties and suggests new strategies for tuning VAC to improve accuracy.

Key words. transition operator, Rayleigh–Ritz method, Markov state models, computational statistical mechanics, conformation dynamics

AMS subject classifications. 65C05, 60J35, 65N30

DOI. 10.1137/20M1335984

1. Introduction.

An essential goal in simulation studies is to identify functions that decorrelate slowly in time. Since the values of these functions can be forecast far into the future, they are used for dimensionality reduction and prediction. Moreover, slowly decorrelating functions describe many scientifically significant processes. For example, in biomolecular systems, large-scale arrangements that control biological activity decorrelate slowly, compared to quickly fluctuating bond lengths and angles.

Dynamical spectral estimation is a numerical method that identifies slowly decorrelating functions by estimating the eigenfunctions and eigenvalues of the Markov transition operator of a system. Under appropriate assumptions, a small number of eigenfunctions span the most slowly decorrelating functions of the system, and the associated eigenvalues determine the slowest decorrelation rates. Dynamical spectral estimation uses simulated trajectories to estimate these quantities of interest.

*Received by the editors May 5, 2020; accepted for publication (in revised form) December 1, 2020; published electronically February 11, 2021.
https://doi.org/10.1137/20M1335984
Funding: The first author was supported by the National Science Foundation award DMS-1646339. The second author was supported by the Molecular Software Sciences Institute Software Fellows program. The second, fourth, and fifth authors were supported by the National Institutes of Health award R01GM109455. The fourth and fifth authors were supported by the National Institutes of Health award R35GM136381. The fifth author was supported by the Advanced Scientific Computing Research Program within the DOE Office of Science through award DE-SC0020427. Computing resources were provided by the University of Chicago Research Computing Center.
†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 USA (rw2515@nyu.edu, weare@nyu.edu).
‡Department of Chemistry, University of Chicago, Chicago, IL 60637 USA (thiede@uchicago.edu, dinner@uchicago.edu).
§Department of Mathematics, University of Chicago, Chicago, IL 60637 USA (ddow11@uchicago.edu).

Despite the wide acceptance of dynamical spectral estimation, estimated eigenfunctions and eigenvalues can have substantial error [45], and the cause of this error is not yet fully understood. Our goal here is to identify and bound the major error sources, thereby identifying opportunities where dynamical spectral estimation can produce accurate results.

Dynamical spectral estimation has been used in fields as diverse as biomolecular simulation [34], fluid mechanics [47], and geophysical analysis [9]. The approach goes by many names in the literature, including Markov state models [41], time-lagged independent component analysis [42], Ulam's method [51], dynamical mode decomposition [40], and extended dynamical mode decomposition [54]. The methods are all closely related, so an error analysis for any one of the methods can shed useful light on the others. Here, for concreteness, we focus on a dynamical spectral estimation method that is well established in chemistry called "the variational approach to conformational dynamics" (VAC) [28, 5, 27, 13].

VAC can be applied to any Markov process X_t that is ergodic and reversible with respect to a distribution μ. Starting from a data set of simulated trajectories, VAC is applied in two steps. First, the data set is used to estimate expectations $C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_\tau)]$ involving a set of basis functions $(\phi_i)_{1\leq i\leq n}$. Then, the spectral decomposition of the matrix $C(0)^{-1}C(\tau)$ is used to estimate eigenvalues and eigenfunctions of the transition operator of X_t.

Our mathematical analysis establishes bounds on VAC's approximation error and estimation error. *Approximation error* is the error in eigenvalue and eigenfunction estimates if the expectations $C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_\tau)]$ are computed perfectly. *Estimation error* is the additional error incurred in VAC estimates because matrices $C(0)$ and $C(\tau)$ are computed imperfectly using a finite data set.

We are not the first authors to mathematically examine VAC's error. Djurdjevac and coauthors [7] bounded the approximation error of VAC eigenvalues. We extend their work by bounding the approximation error for VAC eigenfunctions, which are the chief objects of interest in most applications of dynamical spectral estimation. Additionally, we provide the first analysis of estimation error both for VAC eigenvalues and for eigenfunctions.

Our analysis of VAC also requires proving original error bounds. Standard bounds for the approximation of eigenspaces (e.g., [38, p. 103] or [18, p. 990]) depend on the inverse gap between eigenvalues. However, the gap between any two nontrivial eigenvalues of the transition operator vanishes exponentially fast with the lag time parameter τ. Therefore, the standard bounds increase exponentially as $\tau \to \infty$. In contrast to this asymptotic scaling, we contribute a sharp new perturbation bound that depends only on the inverse *relative* gap between eigenvalues. This new bound reaches its minimal value in the large τ limit, demonstrating the benefit of long lag times for reducing approximation error. In contrast, our asymptotic expressions for the estimation error do depend on the inverse spectral gap and grow in the large τ limit. Therefore, it is best to select an intermediate lag time.

While there is no single ideal lag time dictated by our analysis, we offer new tools for tuning VAC to reduce the estimation error. One such tool, the *VAC condition number*, identifies the subspaces of VAC eigenfunctions most sensitive to estimation error. A second diagnostic, the *mean squared estimation error*, identifies the typical size of the estimation error at different lag times. We provide data-driven formulas for calculating these quantities, enabling VAC

users to identify and avoid the most problematic subspaces and lag times. Our experiments confirm that using these diagnostic tools leads to improved accuracy for VAC estimates.

The paper is organized as follows. Background material is given in section 2, theoretical results are in section 3, numerical experiments are in section 4, mathematical derivations are in section 5, and the conclusions follow in section 6.

2. Background.

This section presents background material explaining the VAC algorithm and the dynamical quantities VAC approximates.

2.1. VAC.

We begin by introducing the steps of VAC when the algorithm is applied to trajectory data from a Markov process X_t with an ergodic, reversible distribution μ. The algorithm starts by estimating expectations involving a set of basis functions $(\phi_i)_{1\leq i\leq n}$. Subsequently, VAC solves an eigenvalue problem involving matrices of expectations.

Algorithm 2.1 VAC algorithm at lag time τ.

1. Form matrix $\hat{C}(0)$ with entries $\hat{C}_{ij}(0) \approx C_{ij}(0) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_0)]$.
2. Form matrix $\hat{C}(\tau)$ with entries $\hat{C}_{ij}(\tau) \approx C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_\tau)]$.
3. Solve eigenvalue problem $\hat{\lambda}_i^\tau \hat{v}^i(\tau) = \hat{C}(0)^{-1}\hat{C}(\tau)\hat{v}^i(\tau)$.
4. Return VAC eigenvalues $\hat{\lambda}_i^\tau$ and VAC eigenfunctions $\hat{\gamma}_i^\tau = \sum_j \hat{v}_j^i(\tau)\phi_j$.

In Algorithm 2.1, we are purposefully vague about the exact method for obtaining trajectory data to estimate

$$\hat{C}_{ij}(\tau) \approx C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_\tau)]. \tag{2.1}$$

One common approach involves simulating long trajectories of X_t and removing the start of each trajectory to limit equilibration bias [43]. A second common approach ("importance sampling" [19]) involves simulating short trajectories and addressing bias through an appropriate reweighting procedure [31, 55]. Since there are no restrictions on how the data set is generated, enhanced sampling techniques can be used to generate the trajectory initial conditions or even the trajectories themselves [3, 32].

In addition to collecting a data set, another key design feature affecting VAC is the choice of basis functions. In the mid-1990s, early versions of VAC used the coordinate axes as basis functions [48, 11], a choice that remains common in molecular dynamics simulations [26, 42, 33]. Independently, in the late 1990s and early 2000s, researchers began constructing spectral estimates using "Markov state models" [41, 45, 46], a procedure mathematically equivalent to performing VAC using a basis of indicator functions on a partition of state space. This idea of using a basis of indicator functions can be traced back to a publication by Stanislaw Ulam in 1960 [51, pp. 74–75] and leads to simplifications in the eigenvalue problem in Algorithm 2.1. In the 2010s, it was observed that these schemes shared a common mathematical framework that could be extended to arbitrary basis sets [28]. Subsequent work led to the development of new families of basis functions [29, 53, 2, 30].

The name "variational approach to conformational dynamics" is inspired by the min-max principle for self-adjoint operators [28, 36]. This variational principle demonstrates that the

top eigenfunctions $\eta_1, \ldots, \eta_k$ of the transition operator maximize the value of the autocorrelation function

$$\rho_\eta(\tau) = \mathrm{corr}_\mu[\eta(X_0), \eta(X_\tau)] \tag{2.2}$$

at any lag time $\tau \geq 0$. Thus, when η is a linear combination of the top k eigenfunctions and u is uncorrelated with the top k eigenfunctions, the autocorrelation functions are related by

$$\rho_\eta(\tau) \geq \rho_u(\tau), \quad \tau \geq 0. \tag{2.3}$$

Consistent with this variational principle, VAC constructs linear combinations of basis functions that maximize autocorrelations. A recent approach due to [21] and [4] extends the linear fitting procedure in VAC by using artificial neural networks to maximize autocorrelations. However, in the present analysis we focus on the linear VAC algorithm as described in Algorithm 2.1, and we leave analysis of the nonlinear fitting procedure to future work.

To help clarify the relationship between VAC and other algorithms, we observe that the computational steps in Algorithm 2.1 can be used for many purposes. For example, AMUSE [50, 25] uses the same computational procedure as Algorithm 2.1, but the goal is to solve the blind-source separation problem in signal processing. Likewise, dynamic mode decomposition [37] and extended dynamic mode decomposition [54] use the same computational procedure as Algorithm 2.1, but the goal is to analyze nonreversible processes, particularly deterministic fluid flows. While the underlying computations are similar in all these cases, VAC refers specifically to the spectral estimation of time-reversible processes. To learn more about the connections between VAC and other related algorithms, we refer the reader to the helpful review paper by Klus and coauthors [16].

2.2. Spectral theory. In this subsection, we take a closer look at the transition operator of the process X_t and its eigenfunctions. We assume X_t is either a continuous-time Feller process [14] or a discrete-time process restricted to even times $t = 0, 2, 4, \ldots$. We assume X_t is ergodic and time-reversible with respect to a distribution μ. We use $\langle \cdot, \cdot \rangle$ to denote the inner product on the Hilbert space $L^2(\mu)$, and we set $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. Lastly, we use $P_{\mathcal{U}}$ to denote the orthogonal projection [35, p. 187] onto the closed linear subspace $\mathcal{U}$ and P_f to denote the orthogonal projection onto the one-dimensional subspace spanned by the function f.

The transition operator [14], also called the Koopman operator, is defined as the conditional expectation operator satisfying

$$T_t[f](x) = \mathrm{E}[f(X_t) | X_0 = x]. \tag{2.4}$$

There are three main properties of the transition operator that determine information about its eigenfunctions.

1. The transition operator T_t is self-adjoint in $L^2(\mu)$. The self-adjointness follows from the time-reversibility condition

$$\mu(dx)\, p_t(x, dy) = \mu(dy)\, p_t(y, dx), \tag{2.5}$$

where $p_t(x, dy)$ denotes the transition probabilities for the process X_t. By integrating over (2.5), we verify the self-adjointness property

$$\langle f, T_t g \rangle = \langle T_t f, g \rangle, \quad f, g \in L^2(\mu). \tag{2.6}$$

2. The transition operator satisfies the semigroup property

$$T_{t+s} = T_t T_s. \tag{2.7}$$

For discrete-time processes, the semigroup property guarantees a decomposition

$$T_t = (T_1)^t, \quad t = 0, 1, 2, \ldots. \tag{2.8}$$

For continuous-time Feller processes, the decomposition can be extended even further, leading to the formula

$$T_t = e^{tA}, \quad t \geq 0, \tag{2.9}$$

which relates the semigroup T_t to its infinitesimal generator A [14].

3. The transition operator T_t is nonnegative, that is,

$$\langle f, T_t f \rangle = \left\langle T_{t/2} f, T_{t/2} f \right\rangle \geq 0, \quad f \in L^2(\mu), \tag{2.10}$$

for all $t \geq 0$ if X_t is a continuous-time process and for $t = 0, 2, 4, \ldots$ if X_t is a discrete-time process.

Using the spectral theorem for self-adjoint operators [35], we obtain a decomposition of either A or T_2. Then, we extend this decomposition to the transition operator at all lag times $t \geq 0$ or $t = 0, 2, 4, \ldots$. The spectral decomposition takes the form

$$T_t = \int_0^\infty e^{-\sigma t} \Pi(d\sigma), \tag{2.11}$$

where $\Pi(d\sigma)$ is a projection-valued measure.

The spectral decomposition completely determines the time correlations of the process X_t. If the spectrum is discrete, then a finite set of orthonormal eigenfunctions is responsible for all the slowest decorrelations of the process. However, if there is an essential spectrum containing $\sigma = 0$, then an infinite set of orthonormal functions decorrelates arbitrarily slowly [35, p. 236].

To avoid the possibility of having an essential spectrum containing $\sigma = 0$, it is sufficient to assume that T_t is compact. Under compactness, the spectral decomposition takes the form

$$T_t = \sum_{i=1}^\infty e^{-\sigma_i t} P_{\eta_i}, \tag{2.12}$$

where $e^{-\sigma_1 t} > e^{-\sigma_2 t} \geq e^{-\sigma_3 t} \geq \cdots$ are eigenvalues and $\eta_1, \eta_2, \eta_3, \ldots$ are the associated eigenfunctions. Since the process is ergodic, $e^{-\sigma_1 t} = 1$ is a simple eigenvalue of T_t corresponding to the eigenfunction $\eta_1 = 1$. Figure 1 shows additional examples of eigenfunctions for a compact transition operator T_t.

While the compactness assumption is enough to facilitate a rigorous analysis of VAC, the compactness assumption can be overly restrictive. In the Monte Carlo literature, there are numerous examples of transition operators that are not compact, such as the transition operator for the Metropolis–Hastings sampler [23, 1]. Therefore, we prefer to use the quasi-compactness assumption, a weaker assumption satisfied by a broader class of processes.

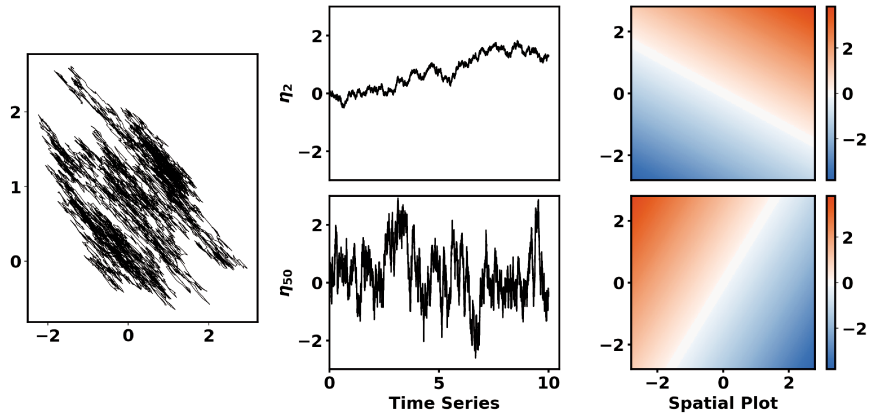

Figure 1. *Eigenfunctions of a compact transition operator, corresponding to dynamics* $d\binom{X_t}{Y_t} = \begin{pmatrix} -0.4 & 0.17 \\ 0.17 & -0.2 \end{pmatrix}\binom{X_t}{Y_t}\,dt + \sqrt{2}\,d\binom{W_t^1}{W_t^2}$. *Left: Typical trajectory of* (X_t, Y_t). *Upper middle: Time series for eigenfunction* η_2 *with slow decorrelation timescale* $\sigma_2^{-1} = 5$. *Lower middle: Time series for eigenfunction* η_{50} *with fast decorrelation timescale* $\sigma_{50}^{-1} = 0.1$. *Right: Spatial structure of* η_2 *and* η_{50}.

Assumption 2.1 (quasi-compactness). The spectral decomposition for the transition operator T_t takes the form

$$T_t = \sum_{i=1}^{r} e^{-\sigma_i t} P_{\eta_i} + R_t, \qquad R_t = \int_{\sigma_{r+1}}^{\infty} e^{-\sigma t} \Pi\left(d\sigma\right). \tag{2.13}$$

Here, $e^{-\sigma_1 t} > e^{-\sigma_2 t} \geq \cdots \geq e^{-\sigma_r t}$ are eigenvalues, $\eta_1, \eta_2, \ldots, \eta_r$ are the associated eigenfunctions, and $e^{-\sigma_{r+1} t}$ is not necessarily an eigenvalue but it bounds the operator norm of the residual operator, that is, $\|R_t\|_2 \leq e^{-\sigma_{r+1} t}$.

Remark 2.2. In the analysis to follow, an "eigenspace" of T_t denotes the closed linear subspace of eigenfunctions with a given eigenvalue. An "invariant subspace" $\mathcal{U}$ is any closed linear subspace satisfying $T_t \mathcal{U} \subseteq \mathcal{U}$.

Remark 2.3. There is a common modification of Algorithm 2.1 where the estimated mean $\hat{\mu}_i \approx \mu_i = \mathrm{E}_\mu\left[\phi_i\left(X_0\right)\right]$ is subtracted from each one of the basis functions ϕ_i before performing VAC (see the discussion in [16]). When the mean is removed, VAC no longer estimates the trivial eigenfunction $\eta_1 = 1$ with eigenvalue $e^{-\sigma_1 t} = 1$; however, VAC continues to estimate all other eigenvalues and eigenspaces.

2.3. Approximation of eigenspaces. It is colloquially said that VAC approximates eigenvalues and eigenfunctions, but it is more correct to say that VAC approximates eigenvalues and *eigenspaces*. Recall that $\hat{\lambda}_i^\tau$ and $\hat{\gamma}_i^\tau$ are the VAC eigenvalues and eigenfunctions, while $e^{-\sigma_i \tau}$ and η_i are the true eigenvalues and eigenfunctions of the transition operator. We assume that VAC eigenvalues are arranged from largest to smallest so that $\hat{\lambda}_1^\tau \geq \hat{\lambda}_2^\tau \geq \cdots \geq \hat{\lambda}_n^\tau$. Then VAC approximates eigenvalues

$$\hat{\lambda}_i^\tau \approx e^{-\sigma_i \tau}, \quad 1 \leq i \leq n. \tag{2.14}$$

VAC approximates eigenspaces and other invariant subspaces

$$\operatorname*{span}_{j\le i\le k} \hat{\gamma}_i^\tau \approx \operatorname*{span}_{j\le i\le k} \eta_i \tag{2.15}$$

whenever there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values.

To measure the error in VAC's invariant subspaces, we introduce two distances: the gap distance $d_2(\cdot,\cdot)$ and the projection distance $d_{\mathrm{F}}(\cdot,\cdot)$ [8].

Definition 2.4. *Consider closed subspaces $\mathcal{U}$ and $\mathcal{W}$, and let $\mathcal{W}^\perp$ indicate the orthogonal complement of $\mathcal{W}$. Then, we define the gap distance and projection distance as follows:*

$$d_2(\mathcal{U},\mathcal{W}) = \|P_{\mathcal{W}^\perp} P_{\mathcal{U}}\|_2, \qquad d_{\mathrm{F}}(\mathcal{U},\mathcal{W}) = \|P_{\mathcal{W}^\perp} P_{\mathcal{U}}\|_{\mathrm{F}}. \tag{2.16}$$

Here, $\|\cdot\|_2$ denotes the operator norm, and $\|\cdot\|_{\mathrm{F}}$ denotes the Hilbert–Schmidt norm, also known as the Frobenius norm.

The gap distance and projection distance are very flexible, and definitions (2.16) can be applied even if $\dim(\mathcal{U}) < \dim(\mathcal{W}) \le \infty$. In this case, we observe that $d_2(\mathcal{U},\mathcal{W})$ and $d_{\mathrm{F}}(\mathcal{U},\mathcal{W})$ are not technically distances. Rather, $d_2(\mathcal{U},\mathcal{W})$ and $d_{\mathrm{F}}(\mathcal{U},\mathcal{W})$ are properly interpreted as distances between $\mathcal{U}$ and the nearest $\dim(\mathcal{U})$-dimensional subspace of $\mathcal{W}$.

We end this section by introducing a useful property of the projection distance, which we apply repeatedly in the analysis.

Lemma 2.5. *Consider $\mathcal{U} = \operatorname{span}(\mathcal{U}_1, \mathcal{U}_2)$, where $\mathcal{U}_1$ and $\mathcal{U}_2$ are orthogonal subspaces, and $\mathcal{W} = \operatorname{span}(\mathcal{W}_1, \mathcal{W}_2)$, where $\mathcal{W}_1$ and $\mathcal{W}_2$ are orthogonal subspaces. Then,*

$$d_{\mathrm{F}}^2(\mathcal{U}_2,\mathcal{W}_2) \le d_{\mathrm{F}}^2(\mathcal{U},\mathcal{W}) + d_{\mathrm{F}}^2(\mathcal{U}_1,\mathcal{W}_1). \tag{2.17}$$

Proof. Calculate

$$d_{\mathrm{F}}^2(\mathcal{U}_2,\mathcal{W}_2) = \|P_{\mathcal{U}_2} P_{\mathcal{W}^\perp}\|_{\mathrm{F}}^2 + \|P_{\mathcal{U}_2} P_{\mathcal{W}_1}\|_{\mathrm{F}}^2 \tag{2.18}$$

$$\le \|P_{\mathcal{U}} P_{\mathcal{W}^\perp}\|_{\mathrm{F}}^2 + \left\|P_{\mathcal{U}_1^\perp} P_{\mathcal{W}_1}\right\|_{\mathrm{F}}^2 \tag{2.19}$$

$$= d_{\mathrm{F}}^2(\mathcal{U},\mathcal{W}) + d_{\mathrm{F}}^2(\mathcal{U}_1,\mathcal{W}_1). \tag{2.20}$$

■

3. Theoretical results. To describe the approach taken in the theoretical analysis, we introduce an idealized VAC algorithm where expectations $C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_\tau)]$ and $C_{ij}(0) = \mathrm{E}_\mu[\phi_i(X_0)\phi_j(X_0)]$ are computed perfectly. Notationally, we distinguish between VAC and idealized VAC by using the ˆ symbol to indicate the quantities calculated using data. For VAC, we write $\hat{C}_{ij}(\tau)$, $\hat{v}^i(\tau)$, $\hat{\lambda}_i^\tau$, and $\hat{\gamma}_i^\tau$. For idealized VAC, we write $C_{ij}(\tau)$, $v^i(\tau)$, λ_i^τ, and γ_i^τ.

In the theoretical analysis, we use idealized VAC to isolate two different sources of error. We decompose the subspace error using

$$\underbrace{d_{\mathrm{F}}\left(\operatorname*{span}_{j\le i\le k} \hat{\gamma}_i^\tau, \operatorname*{span}_{j\le i\le k} \eta_i\right)}_{\text{total error}} \le \underbrace{d_{\mathrm{F}}\left(\operatorname*{span}_{j\le i\le k} \gamma_i^\tau, \operatorname*{span}_{j\le i\le k} \eta_i\right)}_{\text{approximation error}} + \underbrace{d_{\mathrm{F}}\left(\operatorname*{span}_{j\le i\le k} \hat{\gamma}_i^\tau, \operatorname*{span}_{j\le i\le k} \gamma_i\right)}_{\text{estimation error}}. \tag{3.1}$$

Analogously, we decompose the eigenvalue error using

$$\underbrace{\left|\hat{\lambda}_i^\tau - e^{-\sigma_i \tau}\right|}_{\text{total error}} \leq \underbrace{\left|\lambda_i^\tau - e^{-\sigma_i \tau}\right|}_{\text{approximation error}} + \underbrace{\left|\hat{\lambda}_i^\tau - \lambda_i^\tau\right|}_{\text{estimation error}} . \tag{3.2}$$

Approximation error is the difference between idealized VAC estimates and the true eigenvalues and eigenspaces. Estimation error is the difference between VAC estimates and idealized VAC estimates. We first present approximation error bounds in subsection 3.1, and then we present estimation error bounds in subsection 3.2.

Remark 3.1. To illustrate the implications of our error bounds, we use numerical experiments. Thus, Figure 2 and Figure 3 demonstrate the error of VAC when applied to the Ornstein–Uhlenbeck (OU) process $dX = -X\,dt + \sqrt{2}\,dW$ using a basis of indicator functions. Details on how the figures were generated appear in the accompanying supplemental material file SMM133598.pdf [local/web 301KB].

3.1. Approximation error.

In this subsection, we first bound the approximation error by using traditional Rayleigh–Ritz approximation bounds. However, we find that the Rayleigh–Ritz bounds do not provide enough information to show how approximation error depends on the lag time parameter τ. Therefore, we derive improved bounds by using original methods. The improved bounds are asymptotically sharp at long lag times, revealing that long lag times cause the approximation error to stabilize.

3.1.1. Existing approximation bounds are inadequate.

The idealized VAC algorithm is equivalent to the *Rayleigh–Ritz method* in spectral estimation. In the Rayleigh–Ritz method [44], the eigenvalues and eigenspaces of a target operator A are estimated by introducing a subspace of functions $\mathcal{U}$ and then calculating the eigenvalues and eigenspaces of $P_{\mathcal{U}}\, A|_{\mathcal{U}}$, where $A|_{\mathcal{U}}$ denotes the restriction of A to the subspace $\mathcal{U}$. This is also exactly what is done in idealized VAC. The target operator is the transition operator T_τ, and the subspace of basis functions is $\Phi = \operatorname{span}_{1\leq i\leq n} \phi_i$. Moreover, the idealized VAC eigenfunctions γ_i^τ are eigenfunctions of $P_\Phi\, T_\tau|_\Phi$ with eigenvalues λ_i^τ.

The equivalence between the Rayleigh–Ritz method and idealized VAC is known in the VAC literature [39, 7]. However, the implications for VAC's approximation error have not yet been fully explored. Djurdjevac and coauthors [7] applied Rayleigh–Ritz error bounds to analyze idealized VAC eigenvalues. The following theorem goes a step further, by also applying Rayleigh–Ritz error bounds to analyze idealized VAC eigenspaces.

Theorem 3.2 (approximation bounds). *Fix the lag time $\tau > 0$ and the index $1 \leq k \leq r$, but allow the basis set Φ to vary. In the limit as $d_{\mathrm{F}}\left(\operatorname{span}_{1\leq i\leq k} \eta_i, \Phi\right) \to 0$, the idealized VAC estimates converge as follows:*

1. *The idealized VAC eigenvalues $1, 2, \ldots, k$ all converge as*

$$\lambda_i^\tau \to e^{-\sigma_i \tau}, \quad 1 \leq i \leq k. \tag{3.3}$$

2. *When there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values, the subspace $\operatorname{span}_{j\leq i\leq k} \gamma_i^\tau$ of idealized VAC eigenfunctions converges as*

$$\operatorname*{span}_{j\leq i\leq k} \gamma_i^\tau \to \operatorname*{span}_{j\leq i\leq k} \eta_i. \tag{3.4}$$

Additionally, error bounds are given as follows:

1. *The kth idealized VAC eigenvalue is bounded by*

$$1 - d_2^2\left(\operatorname*{span}_{1\le i\le k} \eta_i, \Phi\right) \le \frac{\lambda_k^\tau}{e^{-\sigma_k \tau}} \le 1. \tag{3.5}$$

2. *The top k idealized VAC eigenfunctions are bounded by*

$$1 \le \frac{d_{\mathrm{F}}^2\left(\operatorname*{span}_{1\le i\le k} \gamma_i^\tau, \operatorname*{span}_{1\le i\le k} \eta_i\right)}{d_{\mathrm{F}}^2\left(\operatorname*{span}_{1\le i\le k} \eta_i, \Phi\right)} \le 1 + \frac{\|P_{\Phi^\perp} T_\tau P_\Phi\|_2^2}{\left|e^{-\sigma_k \tau} - \lambda_{k+1}^\tau\right|^2}. \tag{3.6}$$

Proof. See [17, 18] for the original proofs, or see the derivations in the accompanying supplemental material file SMM133598.pdf [local/web 301KB]. ■

The main takeaway from Theorem 3.2 is that the approximation error converges to zero in the limit as

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1\le i\le k} \eta_i, \Phi\right) \to 0. \tag{3.7}$$

Condition (3.7) implies that the basis set Φ must become very rich, so that eigenfunctions η_i can be closely approximated using linear combinations of basis functions.

The Rayleigh–Ritz error bound (3.6) clearly indicates that the eigenspace approximation error must decay with an increasingly rich basis. However, the bound is not sufficiently detailed to show how the approximation error depends on the lag time τ. As seen in Figure 2, the Rayleigh–Ritz bound (3.6) asymptotes to infinity as the lag time increases, implying that approximation error can grow arbitrarily large. In contrast to this upper bound, however, experiments reveal that the approximation error decreases and then stabilizes as the lag time tends to infinity. In the next section, we will derive an improved bound that is asymptotically sharp, describing the exact behavior of the approximation error as $\tau \to \infty$.

3.1.2. New approximation bounds. To analyze the dependence on lag time, we develop a mathematical approach different from the methods applied to the Rayleigh–Ritz method in the past. We start by identifying a key stability property of idealized VAC that has not appeared in the previous literature. As $\tau \to \infty$, idealized VAC eigenspaces converge to a well-defined limit, implying that the approximation error must stabilize at long lag times.

To rigorously study the convergence of idealized VAC estimates, our first step is to introduce the orthogonalized projection functions $q_1, q_2, \ldots$. These are the natural functions to appear in the $\tau \to \infty$ limit. They are constructed from the projected eigenfunctions $P_\Phi \eta_1, P_\Phi \eta_2, \ldots$, but they are adjusted to meet the orthogonality constraints on idealized VAC eigenfunctions.

Definition 3.3. *Set $p = \min\{n, r\}$, where n is the number of basis functions $(\phi_i)_{1\le i\le n}$ and r is the number of eigenfunctions $(\eta_i)_{1\le i\le r}$ identified in Assumption 2.1. Assume that*

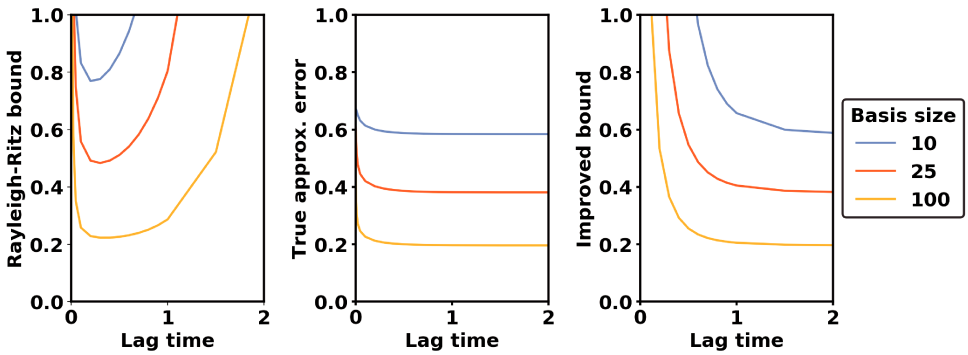

Figure 2. *Comparison between different bounds for the approximation error. Left: The Rayleigh–Ritz bound asymptotes to infinity at long and short lag times. Center: The true approximation error stabilizes at long lag times. Right: The improved bound presented in Theorem 3.4 is asymptotically sharp at long lag times.*

projections $P_\Phi \eta_i$ are linearly independent for $1 \le i \le p$. Then, define

$$\tilde{q}_1 = P_\Phi \eta_1, \qquad q_1 = \tilde{q}_1 / \|\tilde{q}_1\|, \tag{3.8}$$

$$\tilde{q}_2 = P_\Phi \eta_2 - \langle q_1, \eta_2 \rangle q_1, \qquad q_2 = \tilde{q}_2 / \|\tilde{q}_2\|, \tag{3.9}$$

$$\vdots \tag{3.10}$$

$$\tilde{q}_p = P_\Phi \eta_p - \sum_{i=1}^{p-1} \langle q_i, \eta_p \rangle q_i, \qquad q_p = \tilde{q}_p / \|\tilde{q}_p\|. \tag{3.11}$$

Our next step is to prove that idealized VAC eigenfunctions γ_i^τ converge to the orthogonalized projections q_i at long lag times.

Theorem 3.4 (the $\tau \to \infty$ limit). *Fix the basis set Φ but allow the lag time τ to vary. In the limit as $\tau \to \infty$, idealized VAC estimates converge as follows:*

1. *When there is a gap between σ_k and all other σ_i values, the kth idealized VAC eigenvalue satisfies*

$$\frac{\lambda_k^\tau}{e^{-\sigma_k \tau}} \to \langle \eta_k, q_k \rangle^2. \tag{3.12}$$

2. *When there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values, the subspace $\operatorname{span}_{j \le i \le k} \gamma_i^\tau$ of idealized VAC eigenfunctions converges as*

$$\operatorname*{span}_{j \le i \le k} \gamma_i^\tau \to \operatorname*{span}_{j \le i \le k} q_i. \tag{3.13}$$

3. *When there is a gap between σ_k and all other σ_i values and a gap between σ_{k+1} and all other σ_i values, the top k idealized VAC eigenfunctions satisfy*

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1 \le i \le k} \gamma_i^\tau, \operatorname*{span}_{1 \le i \le k} q_i \right) \frac{\lambda_k^\tau}{\lambda_{k+1}^\tau} \to \left| \frac{\langle \eta_{k+1}, q_k \rangle}{\langle \eta_{k+1}, q_{k+1} \rangle} \right|. \tag{3.14}$$

Proof. See subsection 5.2, subsection 5.3, and subsection 5.4. ■

The main message of Theorem 3.4 is that idealized VAC eigenspaces converge exponentially fast as $\tau \to \infty$. Because of this convergence, the approximation error must stabilize. As the last step of our approximation error analysis, we use the stabilization at long lag times to provide a new, asymptotically sharp bound on VAC's approximation error.

Theorem 3.5. *When $\lambda_k^\tau > e^{-\sigma_{k+1}\tau}$, the top k idealized VAC eigenfunctions are bounded by*

$$1 \leq \frac{d_{\mathrm{F}}^2\left(\underset{1\leq i\leq k}{\operatorname{span}}\, \gamma_i^\tau, \underset{1\leq i\leq k}{\operatorname{span}}\, \eta_i\right)}{d_{\mathrm{F}}^2\left(\underset{1\leq i\leq k}{\operatorname{span}}\, \eta_i, \Phi\right)} \leq 1 + \frac{1}{4}\left|\frac{e^{-\sigma_{k+1}\tau}}{\lambda_k^\tau - e^{-\sigma_{k+1}\tau}}\right|^2. \tag{3.15}$$

Proof. See subsection 5.3. ■

Interpreting the results of this subsection, we can identify concrete strategies for how best to reduce approximation error. The approximation error can be divided into two parts:

$$\underbrace{d_{\mathrm{F}}\left(\underset{j\leq i\leq k}{\operatorname{span}}\, \gamma_i^\tau, \underset{j\leq i\leq k}{\operatorname{span}}\, \eta_i\right)}_{\text{approximation error}} \leq \underbrace{d_{\mathrm{F}}\left(\underset{j\leq i\leq k}{\operatorname{span}}\, q_i, \underset{j\leq i\leq k}{\operatorname{span}}\, \eta_i\right)}_{\text{lag-time-independent error}} + \underbrace{d_{\mathrm{F}}\left(\underset{j\leq i\leq k}{\operatorname{span}}\, \gamma_i^\tau, \underset{j\leq i\leq k}{\operatorname{span}}\, q_i\right)}_{\text{lag-time-dependent error}}. \tag{3.16}$$

In this decomposition, we separate the lag-time-independent error and the lag-time-dependent error. In applications of VAC, there are separate strategies for reducing these two error sources.

To reduce the lag-time-independent error, the best strategy is to enrich the basis set as much as possible. If the basis set is rich enough to approximate the top eigenfunctions $\eta_1, \eta_2, \ldots, \eta_k$ with high accuracy, then the lag-time-independent error must be low. Assuming there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values, Lemma 2.5 guarantees

$$\underbrace{d_{\mathrm{F}}^2\left(\underset{j\leq i\leq k}{\operatorname{span}}\, q_i, \underset{j\leq i\leq k}{\operatorname{span}}\, \eta_i\right)}_{\text{squared lag-time-independent error}} \leq d_{\mathrm{F}}^2\left(\underset{1\leq i\leq j-1}{\operatorname{span}}\, \eta_i, \Phi\right) + d_{\mathrm{F}}^2\left(\underset{1\leq i\leq k}{\operatorname{span}}\, \eta_i, \Phi\right). \tag{3.17}$$

As the basis set becomes increasingly rich, the right-hand side of the inequality converges to zero, implying that the lag-time-independent error must vanish.

To reduce the lag-time-dependent error, the best strategy is simply to increase the lag time. As $\tau \to \infty$, Theorem 3.4 guarantees that the lag-time-dependent error must decay exponentially quickly.

3.2. Estimation error. In this subsection, we present formulas for the estimation error and explain how to calculate the mean squared estimation error using data.

3.2.1. Formulas for the estimation error. In applications of VAC, it is not typically possible to evaluate expectations $C_{ij}(\tau) = \mathrm{E}_\mu[\phi_i(X_0)\,\phi_j(X_\tau)]$ exactly. Instead, trajectory data is used to provide estimates $\hat{C}_{ij}(\tau) \approx C_{ij}(\tau)$. In the asymptotic limit as $\hat{C}(\tau) \to C(\tau)$ and $\hat{C}(0) \to C(0)$, the estimation error is governed by the following asymptotic formulas.

Theorem 3.6 (estimation error). *Fix the basis set Φ and the lag time $\tau > 0$, but allow matrices $\hat{C}(0)$ and $\hat{C}(\tau)$ to vary. Assume idealized VAC eigenfunctions are normalized so that $\langle \gamma_i^\tau, \gamma_j^\tau \rangle = \delta_{ij}$, and recall that $v_i(\tau)$ is the vector with $\gamma_i^\tau = \sum_{j=1}^n v_j^i(\tau)\phi_j$. Set*

$$\hat{L}_{ij}(\tau) = v^i(\tau)^T \left[\hat{C}(\tau) - \lambda_j^\tau \hat{C}(0)\right] v^j(\tau), \quad 1 \le i, j \le n. \tag{3.18}$$

Then, VAC estimates have the following behavior as $\hat{C}(\tau) \to C(\tau)$ and $\hat{C}(0) \to C(0)$:

1. *When there is a gap between λ_k^τ and all other λ_i^τ values, the kth VAC eigenvalue satisfies*

$$\hat{\lambda}_k^\tau - \lambda_k^\tau = \hat{L}_{kk}(\tau) + \mathcal{O}\left(\left\|\hat{C}(\tau) - C(\tau)\right\|_{\mathrm{F}}^2 + \left\|\hat{C}(0) - C(0)\right\|_{\mathrm{F}}^2\right). \tag{3.19}$$

2. *When there is a gap between $\{\lambda_j^\tau, \ldots, \lambda_k^\tau\}$ and all other λ_i^τ values, the subspace $\operatorname{span}_{j \le i \le k} \hat{\gamma}_i^\tau$ of VAC eigenfunctions satisfies*

$$d_{\mathrm{F}}\left(\operatorname*{span}_{j \le i \le k} \hat{\gamma}_i^\tau, \operatorname*{span}_{j \le i \le k} \gamma_i^\tau\right) = \left(\sum_{\substack{l<j \\ \text{or } l>k}} \sum_{m=j}^{k} \left|\frac{\hat{L}_{lm}^\tau}{\lambda_l^\tau - \lambda_m^\tau}\right|^2\right)^{1/2} + \mathcal{O}\left(\left\|\hat{C}(\tau) - C(\tau)\right\|_{\mathrm{F}}^2 + \left\|\hat{C}(0) - C(0)\right\|_{\mathrm{F}}^2\right). \tag{3.20}$$

Moreover, the condition number for the subspace $\operatorname{span}_{j \le i \le k} \hat{\gamma}_i^\tau$ is given by

$$\limsup_{\substack{\hat{C}(\tau) \to C(\tau) \\ \hat{C}(0) \to C(0)}} \frac{d_{\mathrm{F}}\left(\operatorname*{span}_{j \le i \le k} \hat{\gamma}_i^\tau, \operatorname*{span}_{j \le i \le k} \gamma_i^\tau\right)}{\left\|\hat{L}(\tau)\right\|_{\mathrm{F}}} = \frac{1}{\min\left\{\lambda_{j-1}^\tau - \lambda_j^\tau, \lambda_k^\tau - \lambda_{k+1}^\tau\right\}}, \tag{3.21}$$

with the conventions $\lambda_0^\tau = \infty$ and $\lambda_{n+1}^\tau = -\infty$.

Proof. See subsection 5.5. ∎

A useful quantity identified in Theorem 3.6 is the condition number (3.21), which quantifies VAC's sensitivity to small errors in the matrices $\hat{C}(\tau)$ and $\hat{C}(0)$. In experiments, we find the condition number is a useful heuristic for judging whether a VAC estimation problem is easy or hard—more specifically, whether a large or small data set is required for accurate estimation. When the condition number for a subspace of VAC eigenfunctions is higher than 5 at all lag times, the numerical experiments in section 4 show that VAC is prone to experiencing large amounts of estimation error. Empirically, we can estimate the minimum condition number across all lag times by using

$$\min_{\tau \ge 0} \frac{1}{\min\left\{\hat{\lambda}_{j-1}^\tau - \hat{\lambda}_j^\tau, \hat{\lambda}_k^\tau - \hat{\lambda}_{k+1}^\tau\right\}}. \tag{3.22}$$

We recommend that VAC users identify the minimum condition number for various subspaces and focus on estimating the well-conditioned subspaces whenever possible. Additionally, we recommend that authors report the minimum condition number along with their VAC results, helping readers to assess whether the results could be affected by estimation error.

3.2.2. Calculating the asymptotic mean squared estimation error using data. Here, we explain how to calculate the mean squared estimation error using trajectory data. We assume for simplicity that the data consists of a single long stationary trajectory of the process X_t. However, the estimation procedure described here could be generalized to other types of trajectory data.

Our approach for calculating the mean squared estimation error is based on the following convergence in distribution result.

Theorem 3.7. *Fix the basis set Φ and the lag time $\tau > 0$, but allow the data set used in VAC to vary. Assume $\mathrm{E}_\mu \left|\phi_i\left(X_0\right)\right|^4 < \infty$ for $1 \leq i \leq n$. Assume that a stationary trajectory $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{T-\Delta}$ is simulated and estimates $\hat{C}_{ij}\left(t\right) \approx C_{ij}\left(t\right)$ are formed using*

$$\hat{C}_{ij}\left(t\right) = \frac{\Delta}{T-t} \sum_{s=0}^{\frac{T-t}{\Delta}-1} \frac{\phi_i\left(X_{s\Delta}\right)\phi_j\left(X_{s\Delta+t}\right) + \phi_j\left(X_{s\Delta}\right)\phi_i\left(X_{s\Delta+t}\right)}{2}. \tag{3.23}$$

Then, VAC estimates have the following behavior as $T \to \infty$:

1. *When there is a gap between λ_k^τ and all other λ_i^τ values, the kth VAC eigenvalue satisfies*

$$\sqrt{T}\left(\hat{\lambda}_k^\tau - \lambda_k^\tau\right) \xrightarrow{\mathcal{D}} Z_{kk}^\tau. \tag{3.24}$$

2. *When there is a gap between $\{\lambda_j^\tau, \ldots, \lambda_k^\tau\}$ and all other λ_i^τ values, the subspace $\operatorname{span}_{j\leq i\leq k} \hat{\gamma}_i^\tau$ of VAC eigenfunctions satisfies*

$$T d_{\mathrm{F}}\left(\operatorname*{span}_{j\leq i\leq k} \hat{\gamma}_i^\tau, \operatorname*{span}_{j\leq i\leq k} \gamma_i^\tau\right)^2 \xrightarrow{\mathcal{D}} \sum_{\substack{l<j \\ or\ l>k}} \sum_{m=j}^{k} \left|\frac{Z_{lm}^\tau}{\lambda_l^\tau - \lambda_m^\tau}\right|^2. \tag{3.25}$$

Here, $\left(Z_{lm}^\tau\right)_{1\leq l,m\leq n}$ is a mean-zero multivariate Gaussian random variable with variance terms

$$\mathrm{E}\left|Z_{lm}^\tau\right|^2 = \Delta \sum_{s=-\infty}^{\infty} \mathrm{Cov}_\mu\left[F_{lm}^\tau\left(X_0, X_\tau\right), F_{lm}^\tau\left(X_{s\Delta}, X_{s\Delta+\tau}\right)\right], \tag{3.26}$$

where we have defined

$$F_{lm}^\tau\left(x,y\right) = \frac{\gamma_l^\tau\left(x\right)\gamma_m^\tau\left(y\right) + \gamma_l^\tau\left(y\right)\gamma_m^\tau\left(x\right)}{2} - \lambda_m^\tau \frac{\gamma_l^\tau\left(x\right)\gamma_m^\tau\left(x\right) + \gamma_l^\tau\left(y\right)\gamma_m^\tau\left(y\right)}{2}. \tag{3.27}$$

Proof. See subsection 5.6. ∎

The great value of Theorem 3.7 is that it suggests a data-driven approach for calculating the mean squared estimation error in the asymptotic limit as $T \to \infty$. First, we can use the data set to estimate the $\mathrm{E}\left|Z_{lm}^\tau\right|^2$ terms by means of (3.26). Second, we can substitute the $\mathrm{E}\left|Z_{lm}^\tau\right|^2$ estimates into (3.24) or (3.25) to compute the mean squared estimation error for eigenvalues or invariant subspaces. In the accompanying supplemental material file SMM133598.pdf [local/web 301KB], we provide a step-by-step description of this estimation procedure.

In Figure 3, we calculate the mean squared estimation error by using a single trajectory of data. We find that it is possible to accurately identify the lag times at which the mean squared estimation error exceeds a critical threshold, such as 0.2. Moreover, in the numerical experiments in section 4, 0.2 is a typical level at which the estimation error begins to contribute significantly to VAC's overall error. Therefore, in VAC applications we recommend calculating the mean squared error and avoiding lag times where the error exceeds such a threshold.

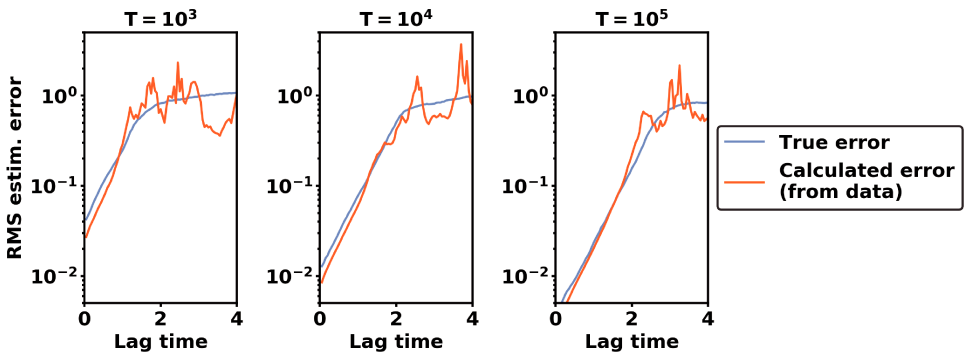

Figure 3. *Root mean squared estimation error for different trajectory lengths T. The calculated error is obtained from a single trajectory of data using formulas in Theorem 3.7. The true error is obtained through* 100 *independent trials.*

We conclude this section by considering three additional strategies to reduce the estimation error of VAC. The first strategy is to increase trajectory length. By increasing the length T of the trajectory, the estimation error consistently decreases at a $1/\sqrt{T}$ rate as shown in Figure 3.

The second strategy for reducing the estimation error is to prune the size of the basis set. We find in Theorem 3.6 that the squared estimation error increases linearly with the number of basis functions. Therefore, it is best to include only those basis functions that have the potential to overlap with the eigenfunctions of the transition operator.

The final strategy for reducing the estimation error is to select basis functions with favorable integrability properties. In Theorem 3.7, it is seen that the mean squared estimation error depends on the fourth moments of the idealized VAC coordinates. If the basis functions themselves have large kurtosis

$$\frac{\mathrm{E}_\mu \left|\phi_i\left(X_0\right) - \mathrm{E}_\mu\left[\phi_i\left(X_0\right)\right]\right|^4}{\left(\mathrm{E}_\mu \left|\phi_i\left(X_0\right) - \mathrm{E}_\mu\left[\phi_i\left(X_0\right)\right]\right|^2\right)^2}, \tag{3.28}$$

this can increase the estimation error in VAC calculations. Favorable integrability properties may be one factor that helps explain the success of Markov state models, in which the basis consists of indicator functions on a partition of the state space. The fourth moments of indicator functions are often well controlled, compared to, e.g., higher-order polynomials of the coordinate axes.

4. Numerical experiments.

In this section, we report on two numerical experiments that illustrate the major factors impacting VAC accuracy. These experiments show how computing

the VAC condition number and the mean squared estimation error can help to improve VAC's results.

4.1. Varying the basis size and trajectory length. First, we apply VAC to estimate the span of eigenfunctions η_1, η_2, and η_3 for the Ornstein–Uhlenbeck (OU) process

$$dX = -X\,dt + \sqrt{2}\,dW\,. \tag{4.1}$$

In two different trials, we show how VAC's accuracy depends on the size of the basis set and the length of the simulated trajectory. The number of basis functions and the trajectory length are varied as follows:

	Trial 1	Trial 2
Basis functions	$n = 20$	$n = 50$
Trajectory length	$T = 10^4$	$T = 500$

In both trials, the basis functions are indicator functions on disjoint intervals.

The two different trials demonstrate that the breakdown of approximation error and estimation error is sensitive to context, as seen in Figure 4. The approximation error is higher in trial 1 because of the smaller set of basis functions, whereas the estimation error is higher in trial 2 because of the smaller data set. In trial 1, it is optimal to use a comparatively long lag time of $\tau = 0.7$ to reduce the approximation error. In contrast, in trial 2 it is optimal to use a comparatively short lag time of $\tau = 0.1$ to avoid the increase in estimation error at longer lag times.

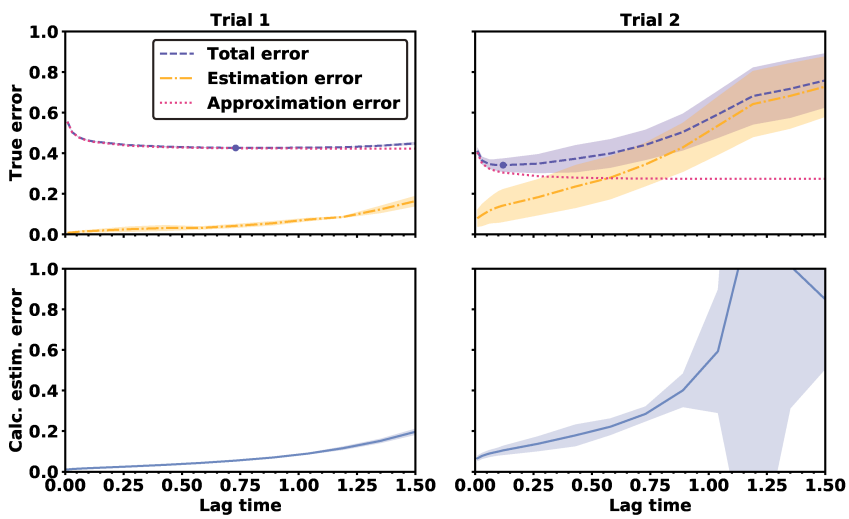

Figure 4. *True error and calculated error for the OU process. Top: The bold line indicates the true root mean squared error over* 30 *independent trajectories, while the shaded region indicates the mean* $\pm$ 1 *standard deviation. The purple dot shows the optimal lag time. Bottom: The calculated root mean squared estimation error obtained from each of the* 30 *trajectories.*

In addition to showing the true error levels, Figure 4 shows the root mean squared estimation error calculated directly from the data. In trial 1, the calculated estimation error remains

below 0.2 for lag times up to 1.5. Therefore, the VAC practitioner can infer that the data set is rich enough to take high lag times without experiencing large estimation error. However, in trial 2, the calculated estimation rises more rapidly, reaching a level of 0.2 when the lag time is just $\tau \approx 0.5$. In this case, the VAC practitioner can infer that the data set is not rich enough to take $\tau > 0.5$.

An alternative lag time selection strategy called *implied timescale analysis* has been advocated in the past by VAC researchers [45]. In this strategy, the VAC eigenvalues are used to compute the implied timescales

$$\frac{-\tau}{\log\left(\hat{\lambda}_i^{\tau}\right)}. \tag{4.2}$$

If VAC eigenvalues were perfect estimates of the true eigenvalues, then implied timescales would be perfectly flat and they would equal σ_i^{-1}. In practice, however, implied timescales are not flat. They increase quickly at short lag times and then increase more slowly at long lag times. To cut down on VAC's error, Swope and coauthors [45] proposed selecting a long enough lag time so that the implied timescales for the eigenfunctions of interest are approximately level.

Figure 5 presents the implied timescales for the OU process. From the figure it is clear that the implied timescales cannot be used to assess the estimation error. The estimation error is much higher in the second trial, yet the implied timescales for trial 1 and trial 2 are similar. However, implied timescales may help assess the approximation error. As the lag time is increased from $\tau = 0$ to $\tau = 0.1$, the second and third implied timescales become much flatter, which provides an accurate indication that the approximation error is decreasing and beginning to settle.

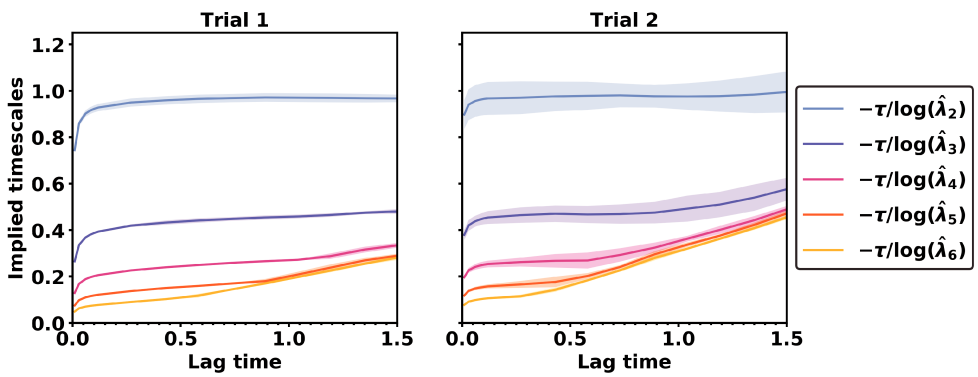

Figure 5. *Implied timescales of the OU process.*

We conclude that implied timescale analysis may provide an approach for assessing approximation error that is complementary to our approach for assessing the estimation error. Whereas our approach is useful for identifying and avoiding the error that is prevalent at long lag times, implied timescale analysis may be useful for identifying and avoiding the error that is prevalent at short lag times. However, while our approach for computing the mean

squared estimation error is rigorously justified, it remains an open research problem to rigorously justify this proposed relationship between the implied timescales and the approximation error.

4.2. Varying the size of the subspace. In our second experiment, we apply VAC to estimate the eigenfunctions of the diffusion process

$$dX = -\frac{1}{2}\sigma\sigma^T \nabla U(X)\, dt + \sigma\, dW, \tag{4.3}$$

where the potential U and the diffusion matrix σ are given by

$$U(x_1, x_2) = 4x_1^4 - 8x_1^2 + x_1 + 0.5x_2^2, \qquad \sigma = \begin{pmatrix} 2 & 0 \\ -1 & \sqrt{3} \end{pmatrix}. \tag{4.4}$$

We simulate a stationary trajectory of length $T = 500$ and then apply VAC using the basis set $\{1, x_1, x_2, x_1^2, x_1x_2, x_2^2\}$.

We investigate how the accuracy changes when VAC is used to estimate two subspaces of different sizes: span $\{\eta_1, \eta_2\}$ and span $\{\eta_1, \eta_2, \eta_3\}$. When estimating span $\{\eta_1, \eta_2\}$, there is a wide range of lag times that all lead to low error levels. As seen in Figure 6, the total error decreases between lag times of $\tau = 0$ and $\tau = 0.2$, but it is nearly constant for all lag times between $\tau = 0.2$ and $\tau = 1.5$. In contrast, when estimating span $\{\eta_1, \eta_2, \eta_3\}$, the total error is V-shaped with a distinct minimum at the lag time $\tau = 0.2$. The error rises rapidly as the lag time is increased beyond $\tau = 0.2$ due to an upsurge in the estimation error.

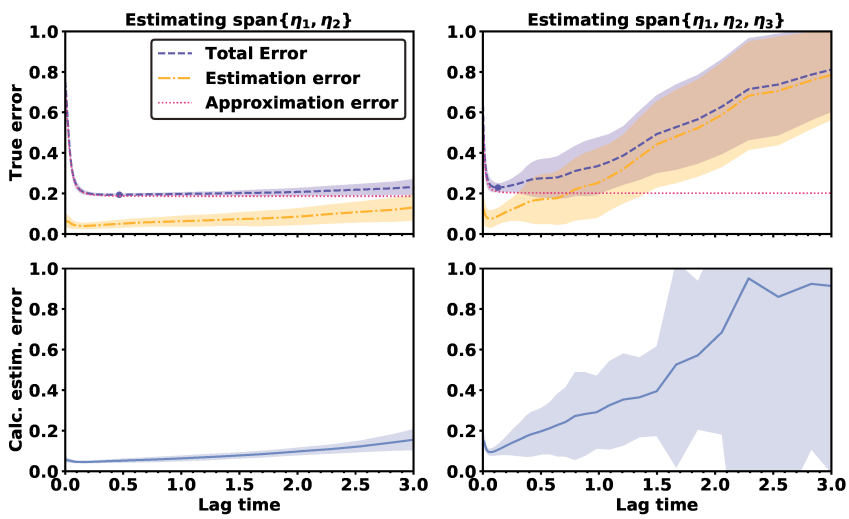

Figure 6. *When the minimum condition number is* 2.0*, the estimation error is low (left). When the minimum condition number is* 9.5*, the estimation error is much higher (right).*

What explains the different error profiles when estimating the subspace span $\{\eta_1, \eta_2\}$ versus span $\{\eta_1, \eta_2, \eta_3\}$? The explanation is not a difference in the data set or the basis set, since these factors remain the same when estimating the two subspaces. Rather, the increase in

estimation error is due to the much higher condition number for the subspace span $\{\eta_1, \eta_2, \eta_3\}$. No matter how the lag time is selected, the inverse spectral gap $(\lambda_4^\tau - \lambda_3^\tau)^{-1}$ is at least as high as 9.5. In contrast, when estimating the subspace span $\{\eta_1, \eta_2\}$, the minimum condition number $\min_\tau (\lambda_3^\tau - \lambda_2^\tau)^{-1}$ is just 2.0. Here we see that a high condition number is associated with increased levels of estimation error and a stronger relationship between estimation error and lag time.

To avoid situations where the estimation error is uncontrollably high, VAC users should identify well-conditioned subspaces and focus on estimating these subspaces whenever possible. As shown in the VAC eigenvalue plot in Figure 7, eigenvalues for well-conditioned subspaces visually stand apart from the rest of the eigenvalues. The large gap between the second and third VAC eigenvalues indicates a natural separation in timescales, which implies that span $\{\eta_1, \eta_2\}$ is a well-conditioned subspace.

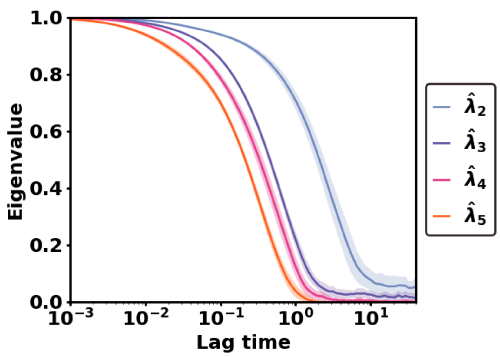

Figure 7. *VAC eigenvalues.*

5. Mathematical derivations.

In this section, we prove the mathematical results presented in Theorem 3.4, Theorem 3.5, Theorem 3.6, and Theorem 3.7.

5.1. Building mathematical intuition.

Before proving Theorem 3.4, we identify the intuitive mathematical reason why idealized VAC estimates converge at long lag times. Applying the spectral decomposition (2.13), we find that each matrix entry $C_{ij}(\tau)$ has an exponentially decaying structure.

$$C_{ij}(\tau) = \langle \phi_i, T_\tau \phi_j \rangle = \sum_{l=1}^{r} e^{-\sigma_l \tau} \langle \eta_l, \phi_i \rangle \langle \eta_l, \phi_j \rangle + \mathcal{O}\left(e^{-\sigma_{r+1}\tau}\right), \quad \tau \to \infty. \tag{5.1}$$

Thus, the matrix $C(\tau)$ is the sum of exponentially decaying rank-one matrices

$$C(\tau) = \sum_{l=1}^{r} e^{-\sigma_l \tau} \left\langle \eta_l, \vec{\phi} \right\rangle \left\langle \eta_l, \vec{\phi} \right\rangle^T + \mathcal{O}\left(e^{-\sigma_{r+1}\tau}\right), \quad \tau \to \infty, \tag{5.2}$$

where we have used the shorthand $\langle \eta_l, \vec{\phi} \rangle$ to denote the vector with entries $\langle \eta_l, \phi_i \rangle$.

To approximate the behavior of idealized VAC at long lag times, we remove the smallest terms in the expansion (5.2) and replace $C(\tau)$ by the sum of k rank-one matrices

$$\sum_{l=1}^{k} e^{-\sigma_l \tau} \left\langle \eta_l, \vec{\phi} \right\rangle \left\langle \eta_l, \vec{\phi} \right\rangle^T. \tag{5.3}$$

When the rank-k approximation is used in place of $C(\tau)$, it results that the top k idealized VAC eigenfunctions $\gamma_1^\tau, \ldots, \gamma_k^\tau$ span the subspace

$$\operatorname*{span}_{1 \le i \le k} q_i = P_\Phi \operatorname*{span}_{1 \le i \le k} \eta_i. \tag{5.4}$$

Therefore, the truncation argument helps intuitively explain the convergence of idealized eigenfunctions γ_i^τ to orthogonalized projections q_i as $\tau \to \infty$. Our proofs in subsection 5.2, subsection 5.3, and subsection 5.4 essentially serve to justify the truncation argument and to provide rigorous bounds on the convergence.

5.2. Convergence of eigenvalues. In this section, we verify the statement in Theorem 3.4 that the kth idealized eigenvalue converges

$$\frac{\lambda_k^\tau}{e^{-\sigma_k \tau}} \to \langle \eta_k, q_k \rangle^2 \tag{3.12}$$

in the limit $\tau \to \infty$, provided there is a gap between σ_k and all other σ_i values. To prove this result, our main tool is the min-max principle for self-adjoint operators [36].

Lemma 5.1. *Consider a quasi-compact self-adjoint operator*

$$A = \sum\nolimits_{i=1}^{r} \lambda_i P_{\eta_i} + R. \tag{5.5}$$

Here, $\eta_1, \eta_2, \ldots, \eta_r$ are orthonormal eigenfunctions of A with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r$, and the spectrum of R lies in $(-\infty, \lambda_r)$. Then, for each $1 \le k \le r$,

$$\lambda_k(A) = \max_{\dim(\mathrm{H})=k} \min_{\eta \in \mathrm{H}} \frac{\langle \eta, A\eta \rangle}{\langle \eta, \eta \rangle}. \tag{5.6}$$

Before applying the min-max principle, we derive two estimates.

Proposition 5.2. *For any $\phi \in \Phi \cap \left(\operatorname{span}_{1 \le i \le k-1} q_i\right)^\perp$,*

$$\frac{\langle \phi, T_\tau \phi \rangle}{\langle \phi, \phi \rangle} \le e^{-\sigma_k \tau} \langle \eta_k, q_k \rangle^2 + e^{-\sigma_{k+1} \tau}. \tag{5.7}$$

Proof. Calculate

$$\langle \phi, T_\tau \phi \rangle = \left\langle \phi, \left(\sum\nolimits_{i=k}^{r} e^{-\sigma_i \tau} P_{\eta_i} + R_\tau \right) \phi \right\rangle \tag{5.8}$$

$$= e^{-\sigma_k \tau} \langle \eta_k, \phi \rangle^2 + \left\langle \phi, \left(\sum\nolimits_{i=k+1}^{r} e^{-\sigma_i \tau} P_{\eta_i} + R_\tau \right) \phi \right\rangle \tag{5.9}$$

$$\le e^{-\sigma_k \tau} \langle \eta_k, q_k \rangle^2 \langle q_k, \phi \rangle^2 + e^{-\sigma_{k+1} \tau} \langle \phi, \phi \rangle \tag{5.10}$$

$$\le e^{-\sigma_k \tau} \langle \eta_k, q_k \rangle^2 \langle \phi, \phi \rangle + e^{-\sigma_{k+1} \tau} \langle \phi, \phi \rangle. \tag{5.11}$$

■

Proposition 5.3. *Set* $\mathrm{H}_{1:k-1} = \operatorname{span}_{1\le i\le k-1} \eta_i$. *Then for any* $q \in Q_{1:k} = \operatorname{span}_{1\le i\le k} q_i$,

$$\frac{\langle q, T_\tau q\rangle}{\langle q, q\rangle} \ge e^{-\sigma_k \tau}\left(\langle \eta_k, q_k\rangle^2 - \frac{1}{e^{(\sigma_k-\sigma_{k-1})\tau}\left(1 - d_2^2\left(\mathrm{H}_{1:k-1}, \Phi\right)\right) - \langle \eta_k, q_k\rangle^2}\right), \tag{5.12}$$

provided the denominator term is positive.

Proof. It suffices to consider the $\|q\| = 1$ case. Then, q can be decomposed as $q = aq' + bq_k$, where $a^2 + b^2 = 1$, $q' \in Q_{1:k-1}$, and $\|q'\| = 1$. It follows that

$$\langle q, T_\tau q\rangle \ge \left\langle q, \left(e^{-\sigma_{k-1}\tau} P_{\mathrm{H}_{1:k-1}} + e^{-\sigma_k\tau} P_{\eta_k}\right) q\right\rangle \tag{5.13}$$

$$= a^2 e^{-\sigma_{k-1}\tau} \left\|P_{\mathrm{H}_{1:k-1}} q'\right\|^2 + e^{-\sigma_k\tau} \left\langle \eta_k, aq' + bq_k\right\rangle^2. \tag{5.14}$$

Thus, $\langle q, T_\tau q\rangle$ is bounded from below by the lowest eigenvalue of

$$M = e^{-\sigma_{k-1}\tau} \left\|P_{\mathrm{H}_{1:k-1}} q'\right\|^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + e^{-\sigma_k\tau} \begin{pmatrix} \langle \eta_k, q'\rangle^2 & \langle \eta_k, q'\rangle \langle \eta_k, q_k\rangle \\ \langle \eta_k, q'\rangle \langle \eta_k, q_k\rangle & \langle \eta_k, q_k\rangle^2 \end{pmatrix}. \tag{5.15}$$

For any symmetric real-valued matrix $M = \left(\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right)$ with $a > c$, the lowest eigenvalue is at least as large as $c - b^2/(a-c)$ [22]. We can check that

$$\left\|P_{\mathrm{H}_{1:k-1}} q'\right\|^2 = 1 - \left\|P_{\mathrm{H}_{1:k-1}^\perp} q'\right\|^2 \ge 1 - d_2^2\left(\mathrm{H}_{1:k-1}, \Phi\right). \tag{5.16}$$

Therefore, the lowest eigenvalue of the matrix M is at least as large as

$$e^{-\sigma_k\tau} \langle \eta_k, q_k\rangle^2 - \frac{e^{-2\sigma_k\tau}}{e^{-\sigma_{k-1}\tau}\left(1 - d_2^2\left(\mathrm{H}_{1:k-1}, \Phi\right)\right) - e^{-\sigma_k\tau}\langle \eta_k, q_k\rangle^2}. \tag{5.17}$$

■

Proof of (3.12). Using the min-max principle and Proposition 5.2, we obtain

$$\lambda_k^\tau = \max_{\dim(\mathrm{H})=k, \mathrm{H}\subseteq\Phi} \min_{\eta\in\mathrm{H}} \frac{\langle \eta, T_\tau \eta\rangle}{\langle \eta, \eta\rangle} \le e^{-\sigma_k\tau} \langle \eta_k, q_k\rangle^2 (1 + o(1)) \tag{5.18}$$

as $\tau \to \infty$. Using the min-max principle and Proposition 5.3, we obtain

$$\lambda_k^\tau = \max_{\dim(\mathrm{H})=k, \mathrm{H}\subseteq\Phi} \min_{\eta\in\mathrm{H}} \frac{\langle \eta, T_\tau \eta\rangle}{\langle \eta, \eta\rangle} \ge e^{-\sigma_k\tau} \langle \eta_k, q_k\rangle^2 (1 + o(1)). \tag{5.19}$$

■

5.3. Convergence of invariant subspaces. In this section, we verify the statement in Theorem 3.4 that the subspace $\operatorname{span}_{j\le i\le k} \gamma_i^\tau$ of idealized VAC eigenfunctions converges as

$$\operatorname*{span}_{j\le i\le k} \gamma_i^\tau \to \operatorname*{span}_{j\le i\le k} q_i \tag{3.13}$$

in the limit $\tau \to \infty$, provided there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values. To prove this result, our main tool is a well-known lemma due to Davis and Kahan [6].

Lemma 5.4. *Suppose A and B are self-adjoint operators and $\mathcal{U}$ and $\mathcal{W}$ are closed subspaces. If the spectrum of $P_{\mathcal{U}}\, A|_{\mathcal{U}}$ lies in the interval $[a, b]$ and the spectrum of $P_{\mathcal{W}}\, B|_{\mathcal{W}}$ lies in $(-\infty, a-\delta] \cup [b+\delta, \infty)$, then*

$$\delta \left\| P_{\mathcal{W}} P_{\mathcal{U}} \right\|_{\mathrm{F}} \leq \left\| P_{\mathcal{W}} P_{\mathcal{U}} A P_{\mathcal{U}} - P_{\mathcal{W}} B P_{\mathcal{W}} P_{\mathcal{U}} \right\|_{\mathrm{F}}. \tag{5.20}$$

The Davis and Kahan lemma leads to the following error bound.

Proposition 5.5. *When $\lambda_k^\tau > e^{-\sigma_{k+1}}$, the distance between subspaces $\Gamma_{1:k}^\tau = \operatorname{span}_{1\leq i\leq k} \gamma_i^\tau$ and $Q_{1:k} = \operatorname{span}_{1\leq i\leq k} q_i$ is bounded by*

$$d_{\mathrm{F}}\left(\Gamma_{1:k}^\tau, Q_{1:k}\right) \leq \frac{e^{-\sigma_{k+1}\tau}}{2\left(\lambda_k^\tau - e^{-\sigma_{k+1}\tau}\right)} d_{\mathrm{F}}\left(\operatorname*{span}_{1\leq i\leq k} \eta_i, \Phi\right). \tag{5.21}$$

Proof. The spectrum of $P_{\Phi\cap Q_{1:k}^\perp}\, T_\tau|_{\Phi\cap Q_{1:k}^\perp}$ lies in the interval $[0, e^{-\sigma_{k+1}\tau}]$, while the spectrum of $P_{\Gamma_{1:k}^\tau}\, T_\tau|_{\Gamma_{1:k}^\tau}$ lies in the interval $[\lambda_k^\tau, 1]$. Therefore, the spectral gap is at least $\lambda_k^\tau - e^{-\sigma_{k+1}\tau}$. We calculate

$$\left(\lambda_k^\tau - e^{-\sigma_{k+1}\tau}\right) \left\| P_{\Phi\cap Q_{1:k}^\perp} P_{\Gamma_{1:k}^\tau} \right\|_{\mathrm{F}} \leq \left\| P_{\Phi\cap Q_{1:k}^\perp} P_{\Gamma_{1:k}^\tau} T_\tau P_{\Gamma_{1:k}^\tau} - P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{\Phi\cap Q_{1:k}^\perp} P_{\Gamma_{1:k}^\tau} \right\|_{\mathrm{F}} \tag{5.22}$$

$$= \left\| P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{\Gamma_{1:k}^\tau} - P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{\Phi\cap Q_{1:k}^\perp} P_{\Gamma_{1:k}^\tau} \right\|_{\mathrm{F}} \tag{5.23}$$

$$= \left\| P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{Q_{1:k}} P_{\Gamma_{1:k}^\tau} \right\|_{\mathrm{F}} \tag{5.24}$$

$$\leq \left\| P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{Q_{1:k}} \right\|_{\mathrm{F}}, \tag{5.25}$$

where we have used the fact that $\Gamma_{1:k}^\tau$ is an invariant subspace of $P_\Phi T_\tau P_\Phi$. Next, we introduce the subspace $\mathrm{H}_{1:k} = \operatorname{span}_{1\leq i\leq k} \eta_i$, which is orthogonal to $\Phi \cap Q_{1:k}^\perp$. Then,

$$\left\| P_{\mathrm{H}_{1:k}^\perp} P_{Q_{1:k}} \right\|_{\mathrm{F}} = \left\| P_{\mathrm{H}_{1:k}} P_{Q_{1:k}^\perp} \right\|_{\mathrm{F}} = \left\| P_{\mathrm{H}_{1:k}} P_{\Phi^\perp} \right\|_{\mathrm{F}} = d_{\mathrm{F}}\left(\operatorname*{span}_{1\leq i\leq k} \eta_i, \Phi\right). \tag{5.26}$$

To complete the theorem, it is enough to show that

$$\left\| P_{\Phi\cap Q_{1:k}^\perp} T_\tau P_{Q_{1:k}} \right\|_{\mathrm{F}} \leq \frac{e^{-\sigma_{k+1}\tau}}{2} \left\| P_{\mathrm{H}_{1:k}^\perp} P_{Q_{1:k}} \right\|_{\mathrm{F}}. \tag{5.27}$$

To prove (5.27), we apply a useful property of the Frobenius norm. For bounded linear operators A and B, if it is true that $\|Au\| \leq \|Bu\|$ for all u, then it follows that $\|A\|_{\mathrm{F}} \leq \|B\|_{\mathrm{F}}$ [12]. Using this property, it is sufficient to prove that

$$\left\| P_{\Phi\cap Q_{1:k}^\perp} T_\tau q \right\| \leq \frac{e^{-\sigma_{k+1}\tau}}{2} \left\| P_{\mathrm{H}_{1:k}^\perp} q \right\|, \quad q \in Q_{1:k}. \tag{5.28}$$

Moreover, it is sufficient to prove that

$$|\langle \phi, T_t q \rangle| \leq \frac{e^{-\sigma_{k+1}\tau}}{2} \tag{5.29}$$

for all $\phi \in \Phi \cap Q_{1:k}^{\perp}$ and $q \in Q_{1:k}$ with $\|\phi\| = \|P_{\mathrm{H}_{1:k}^{\perp}} q\| = 1$. We observe that

$$\left\|P_{\mathrm{H}_{1:k}^{\perp}}(\phi \pm q)\right\|^2 = \left\|\phi \pm P_{\mathrm{H}_{1:k}^{\perp}} q\right\|^2 = 2 \pm 2\left\langle \phi, P_{\mathrm{H}_{1:k}^{\perp}} q\right\rangle = 2 \pm 2\langle \phi, q\rangle = 2. \tag{5.30}$$

Using the polarization identity and the fact that $\mathrm{H}_{1:k}^{\perp}$ is an invariant subspace of T_τ, we conclude that

$$|\langle \phi, T_\tau q\rangle| = \left|\left\langle P_{\mathrm{H}_{1:k}^{\perp}} \phi, T_\tau P_{\mathrm{H}_{1:k}^{\perp}} q\right\rangle\right| \tag{5.31}$$

$$= \left|\frac{1}{4}\left\langle P_{\mathrm{H}_{1:k}^{\perp}}(\phi+q), T_\tau P_{\mathrm{H}_{1:k}^{\perp}}(\phi+q)\right\rangle - \frac{1}{4}\left\langle P_{\mathrm{H}_{1:k}^{\perp}}(\phi-q), T_\tau P_{\mathrm{H}_{1:k}^{\perp}}(\phi-q)\right\rangle\right| \tag{5.32}$$

$$\leq \frac{1}{2}\left\|P_{\mathrm{H}_{1:k}^{\perp}} T_\tau P_{\mathrm{H}_{1:k}^{\perp}}\right\|_2 \tag{5.33}$$

$$\leq \frac{e^{-\sigma_{k+1}\tau}}{2}. \tag{5.34}$$

■

Proof of Theorem 3.5. When $\lambda_k^\tau > e^{-\sigma_{k+1}}$, Proposition 5.5 allows us to calculate

$$d_{\mathrm{F}}^2\left(\operatorname*{span}_{1\leq i\leq k} \eta_i, \Phi\right) \leq d_{\mathrm{F}}^2\left(\operatorname*{span}_{1\leq i\leq k} \gamma_i^\tau, \operatorname*{span}_{1\leq i\leq k} \eta_i\right) \tag{5.35}$$

$$= \|P_{\Phi^\perp} P_{\mathrm{H}_{1:k}}\|_{\mathrm{F}}^2 + \left\|P_{(\Gamma_{1:k}^\tau)^\perp} P_\Phi P_{\mathrm{H}_{1:k}}\right\|_{\mathrm{F}}^2 \tag{5.36}$$

$$= \|P_{\Phi^\perp} P_{\mathrm{H}_{1:k}}\|_{\mathrm{F}}^2 + \left\|P_{(\Gamma_{1:k}^\tau)^\perp} P_{Q_{1:k}} P_{\mathrm{H}_{1:k}}\right\|_{\mathrm{F}}^2 \tag{5.37}$$

$$\leq \left(1 + \frac{1}{4}\left|\frac{e^{-\sigma_{k+1}\tau}}{\lambda_k^\tau - e^{-\sigma_{k+1}\tau}}\right|^2\right) d_{\mathrm{F}}^2\left(\operatorname*{span}_{1\leq i\leq k} \eta_i, \Phi\right). \tag{5.38}$$

■

Proof of (3.13). When there is a gap between $\{\sigma_j, \ldots, \sigma_k\}$ and all other σ_i values, Proposition 5.5 shows that $\operatorname{span}_{1\leq i\leq j} \gamma_i^\tau \to \operatorname{span}_{1\leq i\leq k} q_i$ and $\operatorname{span}_{1\leq i\leq k} \gamma_i^\tau \to \operatorname{span}_{1\leq i\leq k} q_i$ in the limit $\tau \to \infty$. By applying Lemma 2.5, we verify $\operatorname{span}_{j\leq i\leq k} \gamma_i^\tau \to \operatorname{span}_{j\leq i\leq k} q_i$. ■

5.4. Exponential speed of convergence. In this section, we verify the last statement in Theorem 3.4 that the top k idealized VAC eigenfunctions satisfy

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1\leq i\leq k} \gamma_i^\tau, \operatorname*{span}_{1\leq i\leq k} q_i\right) \frac{\lambda_k^\tau}{\lambda_{k+1}^\tau} \to \left|\frac{\langle \eta_{k+1}, q_k\rangle}{\langle \eta_{k+1}, q_{k+1}\rangle}\right| \tag{3.14}$$

in the limit $\tau \to \infty$, provided there is a gap between σ_k and all other σ_i values and a gap between σ_{k+1} and all other σ_i values.

Proof of (3.14). If there are fewer orthogonalized projection functions $(q_i)_{1\leq i\leq p}$ compared to basis functions $(\phi_i)_{1\leq i\leq n}$, we select additional functions $(q_i)_{p+1\leq i\leq n}$ so that $(q_i)_{1\leq i\leq n}$ is a

complete orthonormal basis for Φ. Then,

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1\le i\le k}\gamma_i^\tau, \operatorname*{span}_{1\le i\le k} q_i\right) = \left\|\left(\sum\nolimits_{i=k+1}^{n} q_i\left\langle q_i,\cdot\right\rangle\right)\left(\sum\nolimits_{j=1}^{k}\gamma_j^\tau\left\langle\gamma_j^\tau,\cdot\right\rangle\right)\right\|_{\mathrm{F}} \tag{5.39}$$

$$= \left\|\sum\nolimits_{i=k+1}^{n}\sum\nolimits_{j=1}^{k} q_i\left\langle q_i,\gamma_j^\tau\right\rangle\left\langle\gamma_i^\tau,\cdot\right\rangle\right\|_{\mathrm{F}} \tag{5.40}$$

$$= \left(\sum\nolimits_{i=k+1}^{n}\sum\nolimits_{j=1}^{k}\left\langle q_i,\gamma_j^\tau\right\rangle^2\right)^{1/2}. \tag{5.41}$$

The terms $\langle q_i,\gamma_j^\tau\rangle$ are determined by the eigenvalue equation

$$\lambda_j^\tau\left\langle q_i,\gamma_j^\tau\right\rangle = \left\langle q_i, T_\tau\gamma_j^\tau\right\rangle = \sum\nolimits_{l=1}^{n}\left\langle q_i, T_\tau q_l\right\rangle\left\langle q_l,\gamma_j^\tau\right\rangle. \tag{5.42}$$

As $\tau\to\infty$, Theorem 3.2 and the calculations in (5.31)–(5.33) imply that

$$1/\lambda_j^\tau = \mathcal{O}\left(e^{\sigma_j\tau}\right) \quad\text{and}\quad \left|\langle q_i, T_\tau q_l\rangle\right| \le \min\left\{e^{-\sigma_i\tau}, e^{-\sigma_l\tau}\right\}. \tag{5.43}$$

Setting $\epsilon = \min\{\sigma_{k+2}-\sigma_k, \sigma_{k+1}-\sigma_{k-1}\}$ and sending $\tau\to\infty$, we find that

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1\le i\le k}\gamma_i^\tau, \operatorname*{span}_{1\le i\le k} q_i\right) = \left|\langle q_{k+1},\gamma_k\rangle\right| + \mathcal{O}\left(e^{-\epsilon\tau}\right) \tag{5.44}$$

$$= \left|\frac{1}{\lambda_k^\tau}\sum\nolimits_{l=1}^{n}\left\langle q_{k+1}, T_\tau q_l\right\rangle\left\langle q_l,\gamma_k^\tau\right\rangle\right| + \mathcal{O}\left(e^{-\epsilon\tau}\right). \tag{5.45}$$

Lastly, using the facts that $\gamma_k^\tau\to q_k$ and $\lambda_{k+1}^\tau e^{\sigma_{k+1}\tau}\to\langle\eta_{k+1},q_{k+1}\rangle^2$, we obtain

$$d_{\mathrm{F}}\left(\operatorname*{span}_{1\le i\le k}\gamma_i^\tau, \operatorname*{span}_{1\le i\le k} q_i\right)\frac{\lambda_k^\tau}{\lambda_{k+1}^\tau} = \frac{\langle q_{k+1}, T_\tau q_k\rangle}{\lambda_{k+1}^\tau}\left(1+o\left(1\right)\right) \tag{5.46}$$

$$= \frac{e^{-\sigma_{k+1}\tau}\left\langle\eta_{k+1},q_{k+1}\right\rangle\left\langle\eta_{k+1},q_k\right\rangle}{e^{-\sigma_{k+1}\tau}\left\langle\eta_{k+1},q_{k+1}\right\rangle^2}\left(1+o(1)\right) \tag{5.47}$$

$$= \frac{\langle\eta_{k+1},q_k\rangle}{\langle\eta_{k+1},q_{k+1}\rangle}\left(1+o(1)\right). \tag{5.48}$$

■

5.5. Formulas for the estimation error. In this section, we verify the formulas for the estimation error that are given in Theorem 3.6.

Proof of Theorem 3.6. First, define matrices

$$\Lambda\left(\tau\right) = \operatorname{diag}\left\{\begin{pmatrix}\lambda_1^\tau & \cdots & \lambda_n^\tau\end{pmatrix}\right\}, \qquad V\left(\tau\right) = \begin{pmatrix}v_1\left(\tau\right) & \cdots & v_n\left(\tau\right)\end{pmatrix}. \tag{5.49}$$

Due to the normalization $\delta_{ij} = \langle\gamma_i^\tau,\gamma_j^\tau\rangle$, we must have

$$V\left(\tau\right)^T C\left(0\right)V\left(\tau\right) = I, \qquad V\left(\tau\right)^T C\left(\tau\right)V\left(\tau\right) = \Lambda\left(\tau\right). \tag{5.50}$$

Therefore, idealized VAC eigenfunctions γ_i^τ and eigenvalues λ_i^τ are the eigenfunctions and eigenvalues of the multiplication operator

$$\sum\nolimits_{i,j=1}^{n} \gamma_i^\tau \Lambda_{ij}(\tau) \left\langle \gamma_j^\tau, \cdot \right\rangle. \tag{5.51}$$

In contrast, VAC eigenfunctions $\hat{\gamma}_i^\tau$ and eigenvalues $\hat{\lambda}_i^\tau$ are eigenfunctions and eigenvalues of

$$\sum\nolimits_{i,j=1}^{n} \gamma_i^\tau \left(V(\tau)^{-1} \hat{C}(0)^{-1} \hat{C}(\tau) V(\tau) \right)_{ij} \left\langle \gamma_j^\tau, \cdot \right\rangle. \tag{5.52}$$

As $\hat{C}(0) \to C(0)$ and $\hat{C}(\tau) \to C(\tau)$, we can calculate

$$C(0) \hat{C}(0)^{-1} \hat{C}(\tau) \tag{5.53}$$

$$= \left[I + \left[\hat{C}(0) C(0)^{-1} - I \right] \right]^{-1} \hat{C}(\tau) \tag{5.54}$$

$$= \hat{C}(\tau) - \left[\hat{C}(0) C(0)^{-1} - I \right] C(\tau) + \mathcal{O}\left(\left\| \hat{C}(0) - C(0) \right\|_{\mathrm{F}}^2 + \left\| \hat{C}(\tau) - C(\tau) \right\|_{\mathrm{F}}^2 \right) \tag{5.55}$$

$$= V(\tau)^{-T} \left[\Lambda(\tau) + \hat{L}(\tau) \right] V(\tau)^{-1} + \mathcal{O}\left(\left\| \hat{C}(0) - C(0) \right\|_{\mathrm{F}}^2 + \left\| \hat{C}(\tau) - C(\tau) \right\|_{\mathrm{F}}^2 \right), \tag{5.56}$$

where we have made repeated use of the identities (5.50). Multiplying on the left by $V(\tau)^T$ and on the right by $V(\tau)$, we find that VAC eigenspaces are unitarily equivalent to the eigenspaces of the matrix operator

$$V(\tau)^{-1} \hat{C}(0)^{-1} \hat{C}(\tau) V(\tau) = \Lambda(\tau) + \hat{L}(\tau) + \mathcal{O}\left(\left\| \hat{C}(0) - C(0) \right\|_{\mathrm{F}}^2 + \left\| \hat{C}(\tau) - C(\tau) \right\|_{\mathrm{F}}^2 \right), \tag{5.57}$$

and the two operators share the same eigenvalues. Theorem 3.6 then follows by applying first-order perturbation bounds [15] for eigenvalues and invariant subspaces of a diagonal matrix $\Lambda(\tau)$ that is perturbed by a matrix $\hat{L}(\tau)$. ∎

5.6. Distributional formulas for the estimation error. In this section, we verify the distributional formulas for the estimation error that are given in Theorem 3.7.

Proof of Theorem 3.7. We fix the lag time $t \geq 0$ and the indices $1 \leq i, j \leq n$, but allow the total trajectory length T to vary. Then, we write

$$\hat{C}_{ij}(t) = \frac{\Delta}{T-t} \sum_{s=0}^{\frac{T-t}{\Delta}-1} \phi_{ij}(X_{s\Delta}), \qquad \phi_{ij}(x,y) = \frac{\phi_i(x)\phi_j(y) + \phi_i(y)\phi_j(x)}{2}. \tag{5.58}$$

As $T \to \infty$, we will proceed to show that $\sqrt{T}(\hat{C}_{ij}(t) - C_{ij}(t))$ converges to an asymptotic normal distribution.

By assumption, X_t is started from the stationary distribution $X_0 \sim \mu$, so the random variables $(\phi_{ij}(X_{s\Delta}, X_{s\Delta+\tau}))_{s=0,1,\ldots}$ are strictly stationary [24, pp. 230–231] with mean $C_{ij}(t)$. Moreover, the conditional expectations $\mathrm{E}[\phi_{ij}(X_{s\Delta}, X_{s\Delta+\tau}) | X_0 = x]$ satisfy

$$\| \mathrm{E}[\phi_{ij}(X_{s\Delta}, X_{s\Delta+\tau}) | X_0 = x] - C_{ij}(t) \| \leq C e^{-\sigma_2 s\Delta}, \quad s \geq 0, \tag{5.59}$$

for a constant $C < \infty$ that is independent of s. Condition (5.59) is an asymptotic negligibility condition that guarantees the validity of the central limit theorem for $\phi_{ij}\left(X_{s\Delta}, X_{s\Delta+\tau}\right)$. Using the central limit theorem in [10, Chap. 5], we prove that

$$\sqrt{T}\left(\hat{C}_{ij}\left(t\right) - C_{ij}\left(t\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Delta \sum_{s=-\infty}^{\infty} \mathrm{Cov}_{\mu}\left[\phi_{ij}^{\tau}\left(X_0, X_\tau\right), \phi_{ij}^{\tau}\left(X_{s\Delta}, X_{s\Delta+\tau}\right)\right]\right). \tag{5.60}$$

For simplicity, we have considered the asymptotic distribution of $\sqrt{T}(\hat{C}_{ij}(t) - C_{ij}(t))$. However, by the same approach we can prove the asymptotic normality of any linear combination of random variables $\sqrt{T}(\hat{C}_{ij}(t) - C_{ij}(t))$ involving different values of i, j, and t. By the Cramér–Wold theorem [14], therefore, the array

$$\left[\sqrt{T}\left(\hat{C}\left(0\right) - C\left(0\right)\right), \sqrt{T}\left(\hat{C}\left(\tau\right) - C\left(\tau\right)\right)\right] \tag{5.61}$$

converges to a mean-zero multivariate normal distribution.

To complete the proof of Theorem 3.7, we then apply the "delta method" [52, p. 26] using the formulas (3.19) and (3.20). Since the $\sqrt{T}\|\hat{C}(0) - C(0)\|_{\mathrm{F}}^2$ and $\sqrt{T}\|\hat{C}(\tau) - C(\tau)\|_{\mathrm{F}}^2$ terms are $\mathcal{O}_p(\frac{1}{\sqrt{T}})$ as $T \to \infty$, these terms are asymptotically negligible. The matrix

$$\hat{L}\left(\tau\right) = V\left(\tau\right)^T \left(\hat{C}\left(\tau\right) - \hat{C}\left(0\right)\right) V\left(\tau\right) - V\left(\tau\right)^T \left(\hat{C}\left(0\right) - C\left(0\right)\right) V\left(\tau\right) \Lambda\left(\tau\right) \tag{5.62}$$

is a linear combination of matrices $\hat{C}\left(\tau\right) - C\left(\tau\right)$ and $\hat{C}\left(0\right) - C\left(0\right)$, with each matrix entry $\hat{L}_{ij}\left(\tau\right)$ satisfying

$$\sqrt{T}\hat{L}_{ij}\left(\tau\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Delta \sum_{s=-\infty}^{\infty} \mathrm{Cov}_{\mu}\left[F_{ij}^{\tau}\left(X_0, X_\tau\right), F_{ij}^{\tau}\left(X_{s\Delta}, X_{s\Delta+\tau}\right)\right]\right). \tag{5.63}$$

■

Therefore, the formulas (3.19) and (3.20) guarantee the results in Theorem 3.7.

6. Conclusions.

In this paper, we have identified and bounded the major error sources of "the variational approach to conformational dynamics" (VAC) [28, 5, 27, 13]. VAC is frequently applied in biomolecular simulation studies to estimate the largest eigenvalues $e^{-\sigma_1\tau} \geq e^{-\sigma_2\tau} \geq \cdots \geq e^{-\sigma_k\tau}$ for the Markov transition operator T_τ, along with the corresponding eigenfunctions $\eta_1, \eta_2, \ldots, \eta_k$.

We have proved that VAC accurately identifies subspaces of eigenfunctions $\mathrm{span}_{j\leq i\leq k}\, \eta_i$ when three conditions are satisfied:

1. The values $\{\sigma_j, \ldots, \sigma_k\}$ are separated from all other σ_i values by a spectral gap.
2. The library of basis functions $(\phi_i)_{1\leq i\leq n}$ becomes very rich so that linear combinations of basis functions can fully represent $\eta_1, \ldots, \eta_k$.
3. The data set becomes very large so that expectations $C_{ij}\left(0\right) = \mathrm{E}_\mu\left[\phi_i\left(X_0\right)\phi_j\left(X_0\right)\right]$ and $C_{ij}\left(\tau\right) = \mathrm{E}_\mu\left[\phi_i\left(X_0\right)\phi_j\left(X_\tau\right)\right]$ are evaluated with vanishing error.

VAC converges for any value of the lag time parameter $\tau > 0$, yet the choice of lag time can dramatically alter the speed of convergence. Hence, our main contribution is to prove error bounds that explicitly show how error depends on the lag time. These bounds provide

a full theoretical justification for why limitations in the basis set contribute to the error at short lag times and limitations in the data set contribute to the error at long lag times.

Our numerical analysis approach is flexible, and it could be extended to algorithms besides VAC that estimate dynamical quantities of interest using trajectory data. A broadly useful approach involves decomposing the total error into approximation error and estimation error. Another useful approach involves identifying asymptotic formulas for the estimation error. In future research, it is our goal to rigorously analyze the approximation and estimation error for other powerful algorithms used in biochemical simulation (e.g., [49]).

Lastly, while the main purpose of our work is to deepen theoretical understanding, we also provide diagnostic tools to assess VAC's estimation error and tune VAC's parameters to reduce this error source. We present the *VAC condition number* as a tool for identifying subspaces of VAC eigenfunctions that are prone to experiencing high estimation error. We also present the *mean squared estimation error* as a tool for calculating the estimation error at different lag times. Motivated by the present study, we have also developed an approach for reducing the lag time sensitivity and increasing VAC's robustness by integrating over a window of lag times [20]. These tools have direct relevance to the researchers using VAC, pointing the way toward a more streamlined lag time selection process and a more critical assessment of VAC's error for the future.

Acknowledgments. The authors would like to acknowledge helpful conversations with Gary Froyland, Chatipat Lorpaiboon, and John Strahan. The authors would also like to thank the anonymous reviewers for their helpful critiques and suggestions.

REFERENCES

[1] Y. F. Atchadé and F. Perron, *On the geometric ergodicity of Metropolis-Hastings algorithms*, Statistics, 41 (2007), pp. 77–84.

[2] L. Boninsegna, G. Gobbo, F. Noé, and C. Clementi, *Investigating molecular kinetics by variationally optimized diffusion maps*, J. Chem. Theory Comput., 11 (2015), pp. 5947–5960.

[3] N.-V. Buchete and G. Hummer, *Peptide folding kinetics from replica exchange molecular dynamics*, Phys. Rev. E, 77 (2008), 030902.

[4] W. Chen, H. Sidky, and A. L. Ferguson, *Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets*, J. Chem. Phys., 150 (2019), 214114.

[5] J. D. Chodera and F. Noé, *Markov state models of biomolecular conformational dynamics*, Current Opinion Struct. Biol., 25 (2014), pp. 135–144.

[6] C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation.* III, SIAM J. Numer. Anal., 7 (1970), pp. 1–46, https://doi.org/10.1137/0707001.

[7] N. Djurdjevac, M. Sarich, and C. Schütte, *Estimating the eigenvalue error of Markov state models*, Multiscale Model. Simul., 10 (2012), pp. 61–81, https://doi.org/10.1137/100798910.

[8] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353, https://doi.org/10.1137/S0895479895290954.

[9] G. Froyland, K. Padberg, M. H. England, and A. M. Treguier, *Detection of coherent oceanic structures via transfer operators*, Phys. Rev. Lett., 98 (2007), 224503.

[10] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*, Academic Press, New York, 2014.

[11] H. Hirao, S. Koseki, and H. Takano, *Molecular dynamics study of relaxation modes of a single polymer chain*, J. Phys. Soc. Japan, 66 (1997), pp. 3399–3405.

[12] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.

[13] B. E. Husic and V. S. Pande, *Markov state models: From an art to a science*, J. Amer. Chem. Soc.,

140 (2018), pp. 2386–2396.
[14] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York, 2006.
[15] M. Karow and D. Kressner, *On a perturbation bound for invariant subspaces of matrices*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 599–618, https://doi.org/10.1137/130912372.
[16] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé, *Data-driven model reduction and transfer operator approximation*, J. Nonlinear Sci., 28 (2018), pp. 985–1010.
[17] A. Knyazev, *Sharp a priori error estimates of the Rayleigh-Ritz method without assumptions of fixed sign or compactness*, Math. Notes, 38 (1985), pp. 998–1002.
[18] A. Knyazev, *New estimates for Ritz vectors*, Math. Comput., 66 (1997), pp. 985–995.
[19] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2008.
[20] C. Lorpaiboon, E. H. Thiede, R. J. Webber, J. Weare, and A. R. Dinner, *Integrated variational approach to conformational dynamics: A robust strategy for identifying eigenfunctions of dynamical operators*, J. Phys. Chem. B, 124 (2020), pp. 9354–9364, https://doi.org/10.1021/acs.jpcb.0c06477.
[21] A. Mardt, L. Pasquali, H. Wu, and F. Noé, *VAMPnets for deep learning of molecular kinetics*, Nature Commun., 9 (2018), 5.
[22] R. Mathias, *Quadratic residual bounds for the Hermitian eigenvalue problem*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 541–550, https://doi.org/10.1137/S0895479896310536.
[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.
[24] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, New York, 2012.
[25] L. Molgedey and H. G. Schuster, *Separation of a mixture of independent signals using time delayed correlations*, Phys. Rev. Lett., 72 (1994), 3634.
[26] Y. Naritomi and S. Fuchigami, *Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions*, J. Chem. Phys., 134 (2011), 065101.
[27] F. Noé and C. Clementi, *Collective variables for the study of long-time kinetics from molecular trajectories: Theory and methods*, Current Opinion Struct. Biol., 43 (2017), pp. 141–147.
[28] F. Noé and F. Nuske, *A variational approach to modeling slow processes in stochastic dynamical systems*, Multiscale Model. Simul., 11 (2013), pp. 635–655, https://doi.org/10.1137/110858616.
[29] F. Nüske, B. G. Keller, G. Pérez-Hernández, A. S. Mey, and F. Noé, *Variational approach to molecular kinetics*, J. Chem. Theory Comput., 10 (2014), pp. 1739–1752.
[30] F. Nüske, R. Schneider, F. Vitalini, and F. Noé, *Variational tensor approach for approximating the rare-event kinetics of macromolecular systems*, J. Chem. Phys., 144 (2016), 054105.
[31] F. Nüske, H. Wu, J.-H. Prinz, C. Wehmeyer, C. Clementi, and F. Noé, *Markov state models from short non-equilibrium simulations—analysis and correction of estimation bias*, J. Chem. Phys., 146 (2017), 094104.
[32] 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.
[33] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, *Identification of slow molecular order parameters for Markov model construction*, J. Chem. Phys., 139 (2013), 015102.
[34] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, *Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling*, Nature Chem., 9 (2017), 1005.
[35] M. Reed and B. Simon, *Methods of Modern Mathematical Physics.* 1: *Functional Analysis*, Academic Press, New York, 1975.
[36] M. Reed and B. Simon, *Methods of Modern Mathematical Physics.* 4: *Analysis of Operators*, Academic Press, New York, 1978.
[37] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, *Spectral analysis of nonlinear flows*, J. Fluid Mech., 641 (2009), pp. 115–127.
[38] Y. Saad, *Numerical Methods for Large Eigenvalue Problems: Revised Edition*, Classics Appl. Math. 66, SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611970739.
[39] M. Sarich, F. Noé, and C. Schütte, *On the approximation quality of Markov state models*, Multiscale Model. Simul., 8 (2010), pp. 1154–1177, https://doi.org/10.1137/090764049.
[40] P. J. Schmid, *Dynamic mode decomposition of numerical and experimental data*, J. Fluid Mech., 656 (2010), pp. 5–28.

[41] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *A direct approach to conformational dynamics based on hybrid Monte Carlo*, J. Comput. Phys., 151 (1999), pp. 146–168.

[42] C. R. Schwantes and V. S. Pande, *Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9*, J. Chem. Theory Comput., 9 (2013), pp. 2000–2009.

[43] A. Sokal, *Monte Carlo methods in statistical mechanics: Foundations and new algorithms*, in Functional Integration, Plenum, New York, 1997, pp. 131–192.

[44] G. W. Stewart, *Matrix Algorithms, Volume* II: *Eigensystems*, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898718058.

[45] W. C. Swope, J. W. Pitera, and F. Suits, *Describing protein folding kinetics by molecular dynamics simulations.* 1. *Theory*, J. Phys. Chem. B, 108 (2004), pp. 6571–6581.

[46] W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. C. Ward, Y. Zhestkov, and R. Zhou, *Describing protein folding kinetics by molecular dynamics simulations.* 2. *Example applications to alanine dipeptide and a β-hairpin peptide*, J. Phys. Chem. B, 108 (2004), pp. 6582–6594.

[47] K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, *Modal analysis of fluid flows: An overview*, AIAA J., 55 (2017), pp. 4013–4041.

[48] H. Takano and S. Miyashita, *Relaxation modes in random spin systems*, J. Phys. Soc. Japan, 64 (1995), pp. 3688–3698.

[49] E. H. Thiede, D. Giannakis, A. R. Dinner, and J. Weare, *Galerkin approximation of dynamical quantities using trajectory data*, J. Chem. Phys., 150 (2019), 244111.

[50] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, *Indeterminacy and identifiability of blind identification*, IEEE Trans. Circuits Syst., 38 (1991), pp. 499–509.

[51] S. M. Ulam, *A Collection of Mathematical Problems*, Interscience Tracts Pure Appl. Math. 8, Interscience, New York, London, 1960.

[52] A. W. Van der Vaart, *Asymptotic Statistics*, Camb. Ser. Stat. Probab. Math. 3, Cambridge University Press, Cambridge, UK, 2000.

[53] F. Vitalini, F. Noé, and B. Keller, *A basis set for peptides for the variational approach to conformational kinetics*, J. Chem. Theory Comput., 11 (2015), pp. 3992–4004.

[54] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, *A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition*, J. Nonlinear Sci., 25 (2015), pp. 1307–1346.

[55] H. Wu, F. Nüske, F. Paul, S. Klus, P. Koltai, and F. Noé, *Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations*, J. Chem. Phys., 146 (2017), 154104.