

Integrated Variational Approach to Conformational Dynamics: A Robust Strategy for Identifying Eigenfunctions of Dynamical Operators

Published as part of *The Journal of Physical Chemistry* virtual special issue "Machine Learning in Physical Chemistry".

Chatipat Lorpaiboon,^{*,#} Erik Henning Thiede,^{*,#} Robert J. Webber,^{*,#} Jonathan Weare,^{*} and Aaron R. Dinner^{*}



Cite This: *J. Phys. Chem. B* 2020, 124, 9354–9364



Read Online

ACCESS |



Metrics & More

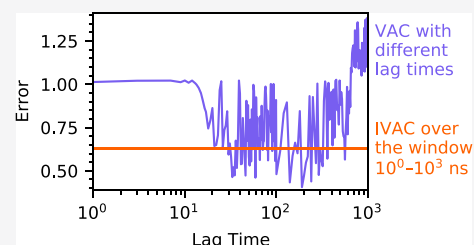


Article Recommendations



Supporting Information

ABSTRACT: One approach to analyzing the dynamics of a physical system is to search for long-lived patterns in its motions. This approach has been particularly successful for molecular dynamics data, where slowly decorrelating patterns can indicate large-scale conformational changes. Detecting such patterns is the central objective of the variational approach to conformational dynamics (VAC), as well as the related methods of time-lagged independent component analysis and Markov state modeling. In VAC, the search for slowly decorrelating patterns is formalized as a variational problem solved by the eigenfunctions of the system's transition operator. VAC computes solutions to this variational problem by optimizing a linear or nonlinear model of the eigenfunctions using time series data. Here, we build on VAC's success by addressing two practical limitations. First, VAC can give poor eigenfunction estimates when the lag time parameter is chosen poorly. Second, VAC can overfit when using flexible parametrizations such as artificial neural networks with insufficient regularization. To address these issues, we propose an extension that we call integrated VAC (IVAC). IVAC integrates over multiple lag times before solving the variational problem, making its results more robust and reproducible than VAC's.



INTRODUCTION

Many physical systems exhibit motion across fast and slow time scales. Whereas individual subcomponents may relax rapidly to a quasi-equilibrium, large collective motions occur over time scales that are orders of magnitude longer. These slow motions are often the most scientifically significant. For instance, observing the large-scale conformational changes that govern protein function requires microseconds to days, even though individual atomic vibrations have periods of femtoseconds. However, when exploring new systems, such slow collective processes may not be fully understood from the outset. Rather, they must be detected from time series data.

One approach for automating this process is the "variational approach to conformational dynamics" (VAC).^{1–3} In the VAC framework, slow dynamical processes are identified with functions that decorrelate slowly. These functions are the eigenfunctions of a self-adjoint operator associated with the system's dynamics known as the *transition operator*. The transition operator evolves expectations of functions over the system's state forward in time and completely defines the dynamics on a distributional level. VAC estimates the transition operator's eigenfunctions by constructing a linear or nonlinear model and using data to optimize parameters in the model. VAC encompasses commonly used approaches such as time-lagged

independent component analysis^{3–6} and eigenfunction estimates constructed using Markov state models.^{6–9} In addition, recent VAC approaches use artificial neural networks to learn approximations to the eigenfunctions.^{10,11}

While VAC has been successful in some applications, the approach has limitations. The accuracy of the estimated eigenfunctions depends strongly on the function space in which the eigenfunctions are approximated, the amount of data available, and a hyperparameter known as the lag time. In our previous work¹² we gave a comprehensive error analysis for the linear VAC algorithm. This error analysis showed that the choice of lag time can be critical to achieving an accurate VAC scheme. Choosing a lag time that is too short can cause substantial systematic bias in estimated eigenfunctions, while choosing a lag time that is too long can make VAC exponentially sensitive to sampling error.

Received: July 15, 2020

Revised: September 19, 2020

Published: September 21, 2020



In this paper, we present an extension of the VAC procedure in which we integrate the correlation functions in VAC over a time window. We term this approach integrated VAC (IVAC). Because IVAC is less sensitive to the choice of lag time, it reduces error compared to VAC. Additionally, when IVAC is applied using an approximation space parametrized by a neural network, the approach leads to stable training and mitigates the overfitting problems associated with VAC.

We organize the rest of the paper as follows. In the [Methods](#) section, we review the role of the transition operator and its eigenfunctions, and we introduce the VAC approach for estimating eigenfunctions. We then present the procedure for IVAC. In the [results](#), we evaluate the performance of IVAC on two model systems. We [conclude](#) with a summary and a discussion of further ways IVAC can be extended. Software implementing IVAC is available at <https://github.com/chatipat/ivac>.

METHODS

Background. In this section, we review the VAC theoretical framework^{1,13} that shows how the slowly decorrelating functions in a physical system can be identified using a linear operator known as the transition operator.

We assume that the system of interest is a continuous-time Markov process $X_t \in \mathbb{R}^n$ with a stationary, ergodic distribution μ (specifically a Feller process¹⁴). We use \mathbf{E} to denote expectations of the process X_t started from μ . For example, if μ is the Boltzmann distribution associated with the Hamiltonian H and temperature T , then expectations of the process satisfy

$$\mathbf{E}[f(X_t)] = \frac{\int f(x) e^{-H(x)/k_B T} dx}{\int e^{-H(x)/k_B T} dx} \quad (1)$$

for all $t \geq 0$. However, our results are valid for systems with other, more general, stationary distributions.

Transition Operator. To begin, we consider the space of real-valued functions with a finite second moment ($\mathbf{E}[f(X_0)^2] < \infty$). Equipped with the inner product

$$\langle f, g \rangle = \mathbf{E}[f(X_0) g(X_0)] \quad (2)$$

this forms a Hilbert space, which we denote L_μ^2 . We define the transition operator¹⁴ at a lag time τ to be the operator

$$\mathcal{T}_\tau f(x) = \mathbf{E}[f(X_\tau) | X_0 = x] \quad (3)$$

applied to a function $f \in L_\mu^2$. Here, we are interpreting the conditional expectation as a function of the initial point x .

The transition operator is also called the Markov or (stochastic) Koopman operator.^{6,15} We use the term transition operator as it is well-established in the literature on stochastic processes, and the terminology emphasizes the connection with finite-state Markov chains. For a finite-state Markov chain, f is a column vector and \mathcal{T}_τ is a row-stochastic transition matrix.

The transition operator lets us rewrite correlation functions in terms of inner products in L_μ^2 :

$$\mathbf{E}[f(X_0) g(X_\tau)] = \langle f, \mathcal{T}_\tau g \rangle \quad (4)$$

Moreover, we can express the slow motions of a system's dynamics in terms of the transition operator. The slow motions are identified by functions f for which the normalized correlation function

$$\frac{\mathbf{E}[f(X_0) f(X_\tau)]}{\mathbf{E}[f(X_0) f(X_0)]} = \frac{\langle f, \mathcal{T}_\tau f \rangle}{\langle f, f \rangle} \quad (5)$$

is large. We will show in the next subsection that these slowly decorrelating functions lie in the linear span of the top eigenfunctions of the transition operator.

Eigenfunctions of the Transition Operator. We can immediately see that \mathcal{T}_τ has the constant function as an eigenfunction, because

$$\mathcal{T}_\tau 1 = \mathbf{E}[1 | X_0 = x] = 1 \quad (6)$$

However, there is no guarantee that any other eigenfunctions exist. We must therefore impose additional assumptions.

We first assume that X_t obeys detailed balance. For any functions $f, g \in L_\mu^2$, we have

$$\mathbf{E}[f(X_0) g(X_\tau)] = \mathbf{E}[f(X_\tau) g(X_0)] \quad (7)$$

or equivalently

$$\langle f, \mathcal{T}_\tau g \rangle = \langle \mathcal{T}_\tau f, g \rangle \quad (8)$$

This detailed balance condition ensures that \mathcal{T}_τ is a self-adjoint operator on L_μ^2 .

Next we assume that \mathcal{T}_τ is a compact operator. In our context, assuming compactness is the same as assuming that the action of \mathcal{T}_τ can be decomposed as an infinite sum involving eigenfunctions and eigenvalues:

$$\mathcal{T}_\tau f(x) = \sum_{i=1}^{\infty} e^{-\sigma_i \tau} \langle \eta_i, f(x) \rangle \eta_i(x) \quad (9)$$

Our assumption of compactness is made for the sake of simplicity; in fact, a weaker assumption of quasi-compactness is sufficient. We refer the reader to Webber et al.¹² for a more general treatment.

At all lag times $\tau > 0$, the function η_i is an eigenfunction of the transition operator \mathcal{T}_τ with eigenvalue

$$\lambda_i^\tau = e^{-\sigma_i \tau} \quad (10)$$

The eigenvalues are indexed so that

$$0 = \sigma_1 < \sigma_2 \leq \sigma_3 \leq \dots \quad (11)$$

and $\lim_{i \rightarrow \infty} \sigma_i = \infty$. Because the process is ergodic, it is known that the largest eigenvalue $\lambda_1^\tau = 1$ is a simple eigenvalue and all other eigenvalues are bounded away from 1. The particular dependence of the eigenvalues on τ occurs because the transition operator can be written as

$$\mathcal{T}_\tau = e^{\tau \mathcal{L}} \quad \forall \tau \geq 0 \quad (12)$$

where \mathcal{L} is an operator known as the infinitesimal generator.¹⁴ We note that it is also common to consider the *implied time scale* (ITS) associated with eigenfunction i , defined as

$$\text{ITS}_i = \sigma_i^{-1} \quad (13)$$

We can use the eigenvalues and eigenvectors of the transition operator to rewrite the normalized correlation function (5). Observing that $\mathcal{T}_0 f(x) = f(x)$ and substituting (9) into the numerator and denominator of (5) gives

$$\frac{\mathbf{E}[f(X_0) f(X_\tau)]}{\mathbf{E}[f(X_0) f(X_0)]} = \frac{\sum_{i=1}^{\infty} e^{-\sigma_i \tau} \langle \eta_i, f \rangle^2}{\sum_{i=1}^{\infty} \langle \eta_i, f \rangle^2} \quad (14)$$

We now consider which functions maximize the normalized correlation function. Applying (11), we find that the normalized correlation function is maximized when we set f to be the constant function $f(x) = \eta_1(x) = 1$, because

$$\frac{\sum_{i=1}^{\infty} e^{-\sigma_i \tau} \langle \eta_i, f \rangle^2}{\sum_{i=1}^{\infty} \langle \eta_i, f \rangle^2} \leq \frac{\sum_{i=1}^{\infty} e^{-\sigma_i \tau} \langle \eta_i, f \rangle^2}{\sum_{i=1}^{\infty} \langle \eta_i, f \rangle^2} \quad (15)$$

$$= e^{-\sigma_1 \tau} \quad (16)$$

$$(15)$$

for all functions $f \in L^2_\mu$. If we constrain the search to functions that are orthogonal to η_1 , i.e., functions where

$$\langle \eta_1, f \rangle = \mathbf{E}[f(x)] = 0 \quad (17)$$

and assume $\sigma_2 > \sigma_3$, the normalized correlation function is maximized when $f = \eta_2$. If we constrain f to be orthogonal to both η_1 and η_2 , then the next slowest decorrelating function would be η_3 , and so forth. Maximizing the normalized correlation function at any lag time τ is therefore equivalent to identifying the eigenfunctions of the transition operator.

Because of the connection to slowly decorrelating functions, the eigenfunctions provide a natural coordinate system for dimensionality reduction. The first few eigenfunctions provide a compact representation of all the slowest motions of the system. Additionally, clustering data based on the eigenfunction coordinates makes it possible to identify metastable states.

Variational Approach to Conformational Dynamics. The “variational approach to conformational dynamics” (VAC) is a procedure for identifying eigenfunctions by maximizing the normalized correlation function. The first eigenfunction is known exactly and is set to the constant function $\eta_1(x) = 1$. To identify subsequent eigenfunctions, we parametrize a candidate solution f using a vector of parameters θ . We then construct an estimate γ_i for the i th eigenfunction by tuning the parameters to maximize (5). We set $\gamma_i = f_\theta$, where

$$\theta' = \arg \max_{\theta} \frac{\mathbf{E}[f_\theta(X_0) f_\theta(X_\tau)]}{\mathbf{E}[f_\theta(X_0) f_\theta(X_0)]} \quad (18)$$

subject to $\langle f_\theta, \gamma_j \rangle = 0$ for all $j < i$. In practice, we use empirical estimates of the correlations constructed from sampled data. For instance, if our data set consists of a single equilibrium trajectory $x_0, x_\Delta, \dots, x_{T-\Delta}$, we would then construct the estimate

$$\hat{\mathbf{E}}[f(X_0) g(X_\tau)] = \frac{\Delta}{T - \tau} \sum_{s=0}^{(T-\Delta-\tau)/\Delta} \frac{f(x_{s\Delta}) g(x_{s\Delta+\tau}) + f(x_{s\Delta+\tau}) g(x_{s\Delta})}{2} \quad (19)$$

Here and in the rest of the paper, we use the $\hat{}$ symbol to indicate quantities constructed using sampled data.

Once we have obtained an estimated eigenfunction $\hat{\gamma}_i$ using data, we can estimate the associated eigenvalue and implied time scale using

$$\hat{\lambda}_i^\tau = \frac{\hat{\mathbf{E}}[\hat{\gamma}_i(X_0) \hat{\gamma}_i(X_\tau)]}{\hat{\mathbf{E}}[\hat{\gamma}_i(X_0) \hat{\gamma}_i(X_0)]} \quad (20)$$

$$\hat{\sigma}_i = -\frac{1}{\tau} \log \hat{\lambda}_i^\tau \quad (21)$$

If the sampling is perfect, the variational principle ensures that VAC eigenvalues and VAC implied time scales are bounded

from above by the true eigenvalues $e^{-\sigma_i \tau}$ and implied time scales σ_i^{-1} , and the upper bound is achieved when the VAC eigenfunction is the true eigenfunction η_i . However, since the empirical estimate (20) is used in practice, it is possible to obtain estimates that exceed the variational upper bound.

The earliest VAC approaches estimated the eigenfunctions of the transition operator by using linear combinations of basis functions $\{\phi_i\}$, a procedure now known as linear VAC. In linear VAC, the optimization parameters are the unknown linear coefficients v_i , which solve the generalized eigenvalue problem

$$\hat{C}(\tau) v_i = \hat{\lambda}_i^\tau \hat{C}(0) v_i \quad (22)$$

where

$$\hat{C}_{jk}(t) = \hat{\mathbf{E}}[\phi_j(X_0) \phi_k(X_t)] \quad (23)$$

In approaches known as time-lagged independent component analysis⁴ and relaxation mode analysis,^{13,16} the basis functions $\{\phi_i\}$ were chosen to be the system’s coordinate axes. This choice of approximation space is still commonly used to construct collective variable spaces either for analyzing dynamics or for streamlining further sampling. Markov state models (MSMs) provide an alternative approach for estimating eigenfunctions using linear combinations of basis functions.^{7–9,17,18} MSMs can serve as general dynamical models for the estimation of metastable structures and chemical rates.^{18–21} When MSMs are applied to estimate eigenfunctions and eigenvalues, the approach is equivalent to performing linear VAC using a basis of indicator functions on disjoint sets.⁶

Noé and Nuske¹ unified the linear VAC approaches and exploited a general variational principle for identifying eigenvalues and eigenfunctions of the transition operator. Subsequent work further developed the methodology and introduced more general linear basis functions.^{2,22–24} Moreover, it was observed that the general variational principle allows one to model the eigenfunctions using nonlinear approximation spaces such as the output of a neural network.^{10,11} This can lead to very flexible and powerful approximation spaces. However, in our experience, the greater flexibility can also lead to overfitting problems that need to be addressed through regularization.

In a common nonlinear VAC approach, a neural network outputs a set of functions $\phi_1, \phi_2, \dots, \phi_S$ that serves as a basis set for linear VAC calculations. The network parameters are then optimized to maximize the VAMP score,²⁵ which under our assumption of detailed balance can be calculated using

$$\text{VAMP-}k = \sum_{i=1}^S |\hat{\lambda}_i^\tau|^k \quad (24)$$

The hyperparameter k is typically set to 1 or 2. In this paper, we use the VAMP-1 score, since we find that it leads to more robust training. We note that the score function we use is also called the generalized matrix Rayleigh quotient.²⁶

Challenges in VAC Calculations. A major challenge in VAC calculations is selecting the lag time τ . Since the early days of VAC, it was noted that lag times that are too short or too long can lead to inaccurate eigenfunction estimates.^{27,28} Our recent work¹² revealed that the sensitivity to lag time is caused by a combination of approximation error at short lag times and estimation error at long lag times. In this section, we describe the impact of approximation error and estimation error and provide a schematic (Figure 1) that illustrates the trade-off between approximation error and estimation error at different lag times.

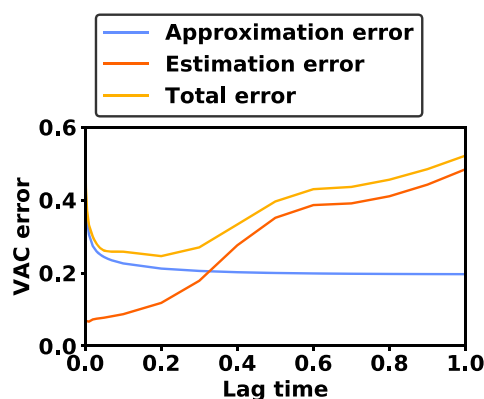


Figure 1. Schematic illustrating the sources of VAC error at different lag times. Even without sampling, VAC solutions have approximation error. Random variation due to sampling contributes additional estimation error.

Approximation error is the systematic error of VAC that exists even when VAC is performed with an infinite data set. We expect approximation error to dominate the calculation when the basis set is of poor quality and our approximation space cannot faithfully represent the eigenfunctions of the transition operator. The approximation error is greatest at short lag times, and it decreases and eventually stabilizes as the lag time is increased. Therefore, VAC users can typically reduce approximation error by avoiding the very shortest lag times.

Estimation error is the random error of VAC that comes from statistical sampling. As shown in our previous work,¹² with increasing lag time the results of VAC become exponentially sensitive to small variations in the data set, leading to high estimation error. At large enough lag times, all the eigenfunction estimates $\hat{\gamma}_1^\tau, \hat{\gamma}_2^\tau, \dots$ are essentially random noise.

In Webber et al.,¹² we proposed measuring VAC's sensitivity to estimation error using the condition number κ^τ . The condition number measures the largest possible changes that can occur in the subspace of VAC eigenfunctions $\{\gamma_j^\tau, \gamma_{j+1}^\tau, \dots, \gamma_k^\tau\}$ when there are small errors in the entries of $C(0)$ and $C(\tau)$. The condition number is calculated using the expression

$$\kappa^\tau = \frac{1}{\min\{\hat{\lambda}_{j-1}^\tau - \hat{\lambda}_j^\tau, \hat{\lambda}_k^\tau - \hat{\lambda}_{k+1}^\tau\}} \quad (25)$$

For a given problem and a given lag time, we can use the condition number to determine which subspaces of VAC eigenfunctions are highly sensitive to estimation error and which subspaces are comparatively less sensitive to estimation error.

Although we rigorously derived the condition number only in the case of linear VAC, we find that the condition number is also helpful for measuring estimation error in nonlinear VAC. If $\kappa^\tau \gtrsim 5$ at all lag times τ , then identifying eigenfunctions is very difficult and requires a large data set. We recommend that authors report the condition number along with their VAC results, helping readers to assess whether the results are potentially sensitive to estimation error.

Integrated VAC. To address the difficulty inherent in choosing a good lag time, we propose an extension of VAC called “integrated VAC” (IVAC) where we integrate over a range of different lag times before solving a variational problem. We find that the new approach is more robust to lag time selection and it often gives better results overall.

Just as VAC maximizes the correlation function in (5), IVAC solves a variational problem by identifying a subspace of functions f that maximize the integrated correlation function

$$\int_{\tau_{\min}}^{\tau_{\max}} \frac{\mathbb{E}[f(X_0) f(X_s)]}{\mathbb{E}[f(X_0) f(X_0)]} ds \quad (26)$$

As in VAC, the functions solving the variational problem are the eigenfunctions of the transition operator. When the eigenfunction η_i is substituted into the integrated correlation function (26), the resulting expression is related to the implied time scales by

$$\int_{\tau_{\min}}^{\tau_{\max}} \frac{\mathbb{E}[\eta_i(X_0) \eta_i(X_s)]}{\mathbb{E}[\eta_i(X_0) \eta_i(X_0)]} ds = \frac{e^{-\sigma_i \tau_{\min}} - e^{-\sigma_i \tau_{\max}}}{\sigma_i} \quad (27)$$

Therefore, like VAC, IVAC is a variational approach for identifying both eigenfunctions and implied time scales.

IVAC is a natural extension of VAC; in the limit as τ_{\max} approaches τ_{\min} , IVAC gives the same eigenfunction and implied time scale estimates as regular VAC. However, when τ_{\max} and τ_{\min} are separated from each other, the results of IVAC and VAC start to diverge. We find that IVAC with minimal tuning performs comparably to VAC with optimal tuning. IVAC has the desirable feature that it is not very sensitive to the values of τ_{\min} and τ_{\max} .

Previous approaches for estimating eigenfunctions using multiple time lags have attempted to reduce approximation error by accounting for unobserved degrees of freedom.^{29–32} In contrast, IVAC uses multiple time lags to reduce estimation error and improve robustness to parameter choice.

Linear IVAC. Linear IVAC uses linear combinations of basis functions to maximize the integrated autocorrelation function (26). However, as simulation data are sampled at discrete time points, we cannot directly calculate the integral. We therefore replace (26) with a discrete sum taken over uniformly spaced lag times. We seek to maximize

$$\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \frac{\mathbb{E}[f(X_0) f(X_\tau)]}{\mathbb{E}[f(X_0) f(X_0)]} \quad (28)$$

where $\tau = \tau_{\min}, \tau_{\min} + \Delta, \tau_{\min} + 2\Delta, \dots, \tau_{\max}$ and Δ is the sampling interval. The discrete sum (28) approximates (26) up to a constant multiple, and its value is maximized when f lies within the span of the top eigenfunctions of the transition operator. Setting f to be the eigenfunction η_i , we can sum the resulting finite geometric series:

$$\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \frac{\mathbb{E}[\eta_i(X_0) \eta_i(X_\tau)]}{\mathbb{E}[\eta_i(X_0) \eta_i(X_0)]} = \frac{e^{-\sigma_i \tau_{\min}} - e^{-\sigma_i (\tau_{\max} + \Delta)}}{1 - e^{-\sigma_i \Delta}} \quad (29)$$

In linear IVAC, we optimize linear combinations of basis functions $\{\phi_i\}$ to maximize the functional (28). The optimization parameters are the unknown linear coefficients v_i , which solve the generalized eigenvalue problem

$$\hat{I}(\tau_{\min}, \tau_{\max}) v_i = \hat{\lambda}_i \hat{C}(0) v_i \quad (30)$$

where we have defined

$$\hat{C}_{jk}(t) = \hat{\mathbb{E}}[\phi_j(X_0) \phi_k(X_t)] \quad (31)$$

$$\hat{I}(\tau_{\min}, \tau_{\max}) = \sum_{\tau=\tau_{\min}}^{\tau_{\max}} \hat{C}(\tau) \quad (32)$$

We solve the generalized eigenvalue problem to obtain estimates $\hat{\gamma}_i$ for the transition operator's eigenfunctions. Then, we form the sum

$$\sum_{\tau=\tau_{\min}}^{\tau_{\max}} \frac{\hat{\mathbf{E}}[\hat{\gamma}_i(X_0) \hat{\gamma}_i(X_\tau)]}{\hat{\mathbf{E}}[\hat{\gamma}_i(X_0) \hat{\gamma}_i(X_0)]} \quad (33)$$

and we estimate implied time scales by solving (29) for $\hat{\sigma}_i$ using a root-finding algorithm.

Nonlinear IVAC. Nonlinear IVAC maximizes the integrated correlation function (26) by constructing approximations in a nonlinear space of functions, for example, those represented by a neural network. Specifically, the nonlinear model provides a set of functions $\phi_1, \phi_2, \dots, \phi_S$ that serves as a basis set for linear IVAC. The parameters are trained to maximize the VAMP- k score

$$\sum_{i=1}^S |\hat{\lambda}_i|^k \quad (34)$$

where the eigenvalues $\hat{\lambda}_i$ are defined using eq 30. In a linear approximation space, all values of VAMP- k scores lead to identical eigenfunction estimates. In a nonlinear approximation space, it is theoretically possible that minimizing with different values of k would lead to different estimates. However, in practice we find there is little difference between estimates at the minima. We present our results using $k = 1$ because it leads to the most stable convergence; we found that higher values of k are prone to large gradients and, in turn, unstable training. When $k = 1$, the score function can be computed using

$$\text{tr}(\hat{\mathbf{C}}(0)^{-1} \hat{\mathbf{I}}(\tau_{\min}, \tau_{\max})) \quad (35)$$

The main practical challenge in an application of nonlinear IVAC is that the basis functions $\phi_1, \phi_2, \dots, \phi_S$ change at every iteration, requiring costly re-evaluation of $\hat{\mathbf{C}}(0), \hat{\mathbf{I}}(\tau_{\min}, \tau_{\max})$, and the gradient of (35) with respect to the parameters. To reduce this cost, we have developed the batch subsampling approach described in Algorithm 1, which we apply at the start of each optimization iteration.

Algorithm 1: subsampling routine

input : data $x_0, \dots, x_{T-\Delta}, \tau_{\min}, \tau_{\max}$,
 number of samples N
for $n \in 1, 2, \dots, N$ **do**
 Sample τ_n from $\{\tau_{\min}, \dots, \tau_{\max}\}$;
 Sample s_n from $\{0, \dots, T - \tau_n - \Delta\}$;
end
output: sampled pairs $(x_{s_n}, x_{s_n+\tau_n})$

In the subsampling approach, we draw a randomly chosen set of data points, which allow us to estimate the matrix entries $\hat{C}_{ij}(0)$ using

$$\sum_{n=1}^N \frac{\phi_i(x_{s_n}) \phi_j(x_{s_n}) + \phi_i(x_{s_n+\tau_n}) \phi_j(x_{s_n+\tau_n})}{2N} \quad (36)$$

and the matrix entries $\hat{I}_{ij}(\tau_{\min}, \tau_{\max})$ using

$$\sum_{n=1}^N \frac{\phi_i(x_{s_n}) \phi_j(x_{s_n+\tau_n}) + \phi_i(x_{s_n+\tau_n}) \phi_j(x_{s_n})}{2N\Delta/(\tau_{\max} - \tau_{\min} + \Delta)} \quad (37)$$

After constructing these random matrices, we calculate the score function (35). We then use automatic differentiation to obtain the gradient of the score function with respect to the parameters, and we perform an optimization step. By randomly drawing new data points at each optimization step, we ensure a thorough sampling of the data set and we are able to train the nonlinear representation at reduced cost. Typically, we find that 10^3 – 10^4 data points per batch is enough for the score function (35) to be estimated with low bias.

RESULTS AND DISCUSSION

In this section, we provide evidence that IVAC is more robust than VAC and can give more accurate eigenfunction estimates. First, we show results from applying IVAC and VAC to the alanine dipeptide. VAC can provide accurate eigenfunction estimates for this test problem owing to the large spectral gap and the approximation space that overlaps closely with the eigenfunctions of the transition operator. However, VAC requires a careful tuning of the lag time. In contrast, IVAC is much less sensitive to lag time choice. IVAC gives solutions that are comparable to VAC with the optimal lag time parameter and substantially better than VAC with a poorly chosen lag time.

Second, we show results for the villin headpiece protein. Because the data set has a small number of independent samples and the neural network approximation space is flexible and prone to overfitting, VAC and IVAC suffer from estimation error at long lag times. Despite these challenges, we present a robust protocol for choosing parameters in IVAC to limit the estimation error, and we show that IVAC is less sensitive to overfitting for this problem compared with VAC.

Application to the Alanine Dipeptide. In this section we compare linear IVAC and VAC applied to Langevin dynamics simulations of the alanine dipeptide (i.e., *N*-acetylalanine-*N'*-methylamide) in aqueous solvent; further simulation details are given in the [Supporting Information](#).

The alanine dipeptide is a well-studied model for conformational changes in proteins. Like many protein systems, the alanine dipeptide has dynamics that are dominated by transitions between metastable states. The top eigenfunctions are useful for locating barriers between states, as these eigenfunctions change sharply when passing from one well to another. We focus on estimating η_2 and η_3 , as large changes in these eigenfunctions correspond to transitions over the alanine dipeptide's two largest barriers. We refer to the span of η_1, η_2 , and η_3 as the 3D subspace.

In our experiments, we consider trajectories of length 10 and 20 ns. The trajectories are long enough to observe approximately 15 or 30 transitions respectively along the dipeptide's slowest degree of freedom. Folding simulations of proteins, such as the villin headpiece considered below, often have a similar number of transitions between the folded and unfolded states.

There are several features that make it possible for VAC to perform well on this example. First, the linear approximation space, which consists of all the dihedral angles in the molecular backbone, is small (just 9 basis functions), and it is known to overlap heavily with the top eigenfunctions of the dynamics. Second, we are estimating a well-conditioned subspace with a minimum condition number of just

$$\min_{\tau} \kappa^{\tau} = \min_{\tau} (\hat{\lambda}_3^{\tau} - \hat{\lambda}_4^{\tau})^{-1} = 1.4 \quad (38)$$

and therefore we do not expect a heavy amplification of sampling error that degrades eigenfunction estimates.

To evaluate the error in our eigenfunction estimates, we compare to “ground truth” eigenfunctions computed using a Markov state model built with a very long time series (1.5 μ s) and a fine discretization of the dihedral angles. We measure error using the projection distance,³³ which evaluates the overlap between one subspace and the orthogonal complement of another subspace. For subspaces \mathcal{U} and \mathcal{V} with orthonormal basis functions $\{u_i\}$ and $\{v_j\}$, the projection distance is given by

$$d(\mathcal{U}, \mathcal{V}) = \sqrt{\sum_{i,j} (\delta_{ij} - \langle u_i, v_j \rangle)^2} \quad (39)$$

This measure, which combines the error in the different eigenfunctions into a single number, is useful because VAC is typically used to identify subspaces of eigenfunctions rather than individual eigenfunctions. The maximum possible error when estimating k eigenfunctions is \sqrt{k} .

Our main result from the alanine dipeptide application is that IVAC is more robust to the selection of lag time parameters than VAC. In Figure 2, we report the accuracy of IVAC and VAC for

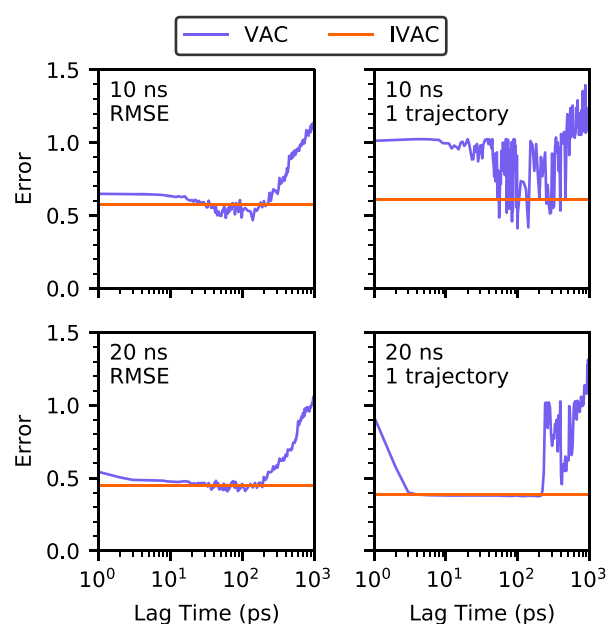


Figure 2. Linear IVAC and VAC errors for alanine dipeptide trajectories. IVAC was applied with $\tau_{\min} = 1$ ps and $\tau_{\max} = 1$ ns. VAC was applied with variable lag time τ (horizontal axis). Errors are computed using the projection distance to the MSM reference for the span of η_2 and η_3 . (Left) Root mean square errors (RMSE) over 30 independent trajectories. (Right) Errors for a single trajectory.

different lag times and trajectory lengths. In the left column, we show the root-mean-square errors (RMSE) for IVAC (orange) and VAC (purple), aggregated over 30 independent trajectories. From the aggregated results, IVAC performs nearly as well as VAC with the best possible τ and consistently gives results much better than VAC with a poorly chosen τ . The RMSE of IVAC is just 0.58 with 10 ns trajectories and 0.45 with 20 ns trajectories. These low error levels are not far from the minimum error of 0.37 that is possible using our linear approximation space.

In the right column of Figure 2, we show results for a 10 ns trajectory and a 20 ns trajectory. The trajectories were selected to help illustrate differences in the error profiles for VAC and IVAC; similar plots for all other trajectories can be found in the Supporting Information. We observe two key differences. First,

VAC error can exhibit high-frequency stochastic variability as a function of lag time, a source of variability that does not affect integrated VAC results. Second, VAC can have high error levels at very short and long lag times. The projection distance against our reference often reaches 1.0, which might indicate that a true eigenfunction is completely orthogonal to our estimated subspace. The error of IVAC is unlikely to reach such high extremes.

We note that the parameter values $\tau_{\min} = 1$ ps and $\tau_{\max} = 1$ ns used in IVAC are not hard to tune. The range 1 ps to 1 ns is a broad window of lag times over which VAC eigenvalues λ_2^* and λ_3^* decrease from values near one to values near zero. In contrast, it is much harder to tune the VAC lag time τ . VAC results are very sensitive to high or low lag times as seen in Figure 2.

When eigenfunction estimates are accurate, we expect that the eigenfunction coordinates will help identify the system's metastable states. In Figure 3, we compare the results of clustering configurations in the 20 ns alanine dipeptide trajectory in Figure 2 using the associated IVAC and VAC estimates. We plot the predicted metastable states against the dipeptide's ϕ and ψ dihedral angles. In the figure, we present VAC results taken at a short lag time, an intermediate lag time, and a long lag time. We also present results for the MSM reference. Comparing against the reference, we find that IVAC identifies clusters as accurately as VAC at a well-chosen lag time, and IVAC performs far better than VAC at a poorly chosen lag time.

Next, we present additional analyses applied to a single 20 ns alanine dipeptide trajectory that provide insight into why IVAC is more robust to lag time selection than VAC. To start, we examine the discrepancy in VAC results at different lag times. In Figure 4, left, we performed VAC with a range of different lag times, and we measured the projection distance between the VAC results obtained at one lag time τ_1 (horizontal axis) and the VAC results obtained at a different lag time τ_2 (vertical axis). The square with low projection distance between 3 and 200 ps indicates that VAC results with lag times chosen within this range are similar to one another, but not to those with lag times taken from outside this range.

The discrepancy between VAC results at both low and high lag times can be explained by a plot of VAC eigenvalues (Figure 4, center). At 3 ps, there is an eigenvalue crossing between the eigenvalues λ_3^* and λ_4^* (shown in purple and magenta). The eigenvalue crossing causes VAC to misidentify the third VAC eigenfunction (which is inside the 3D subspace) and the fourth VAC eigenfunction (which is outside the 3D subspace). At 200 ps, there is a different problem related to insufficient sampling. The third eigenvalue descends into noise, causing VAC to fit the first two eigenfunctions at the expense of the 3D subspace.

With integrated VAC, the problem of finding a single good lag time is replaced with the problem of finding two end points for a range of lag times. This proves to be an easier task as IVAC is more tolerant of lag times outside the region where VAC gives good results. In Figure 4, right, we show the error of IVAC as a function of τ_{\min} and τ_{\max} (horizontal and vertical axes, respectively). This figure, which shows the error of IVAC estimates computed from comparison with the reference, is different from the figure on the left which shows only the discrepancy between VAC results at different lag times. Figure 4, right, also shows the error of VAC, which appears along the diagonal of the plot corresponding to the case $\tau_{\min} = \tau_{\max}$.

Figure 4, right, reveals that the range of lag time parameters for which IVAC exhibits low error levels is much broader than the

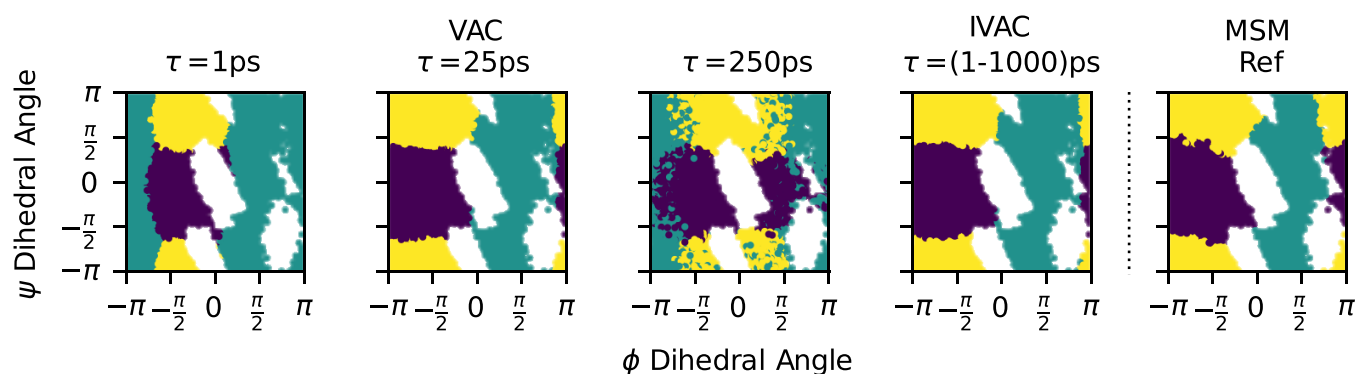


Figure 3. Clusters on the eigenfunctions estimated using VAC and IVAC compared with clusters on an accurate MSM. (Left of the dashed line) VAC and IVAC results for the 20 ns trajectory from Figure 2. (Right of the dashed line) Clustering on η_2 and η_3 evaluated using an accurate MSM reference.

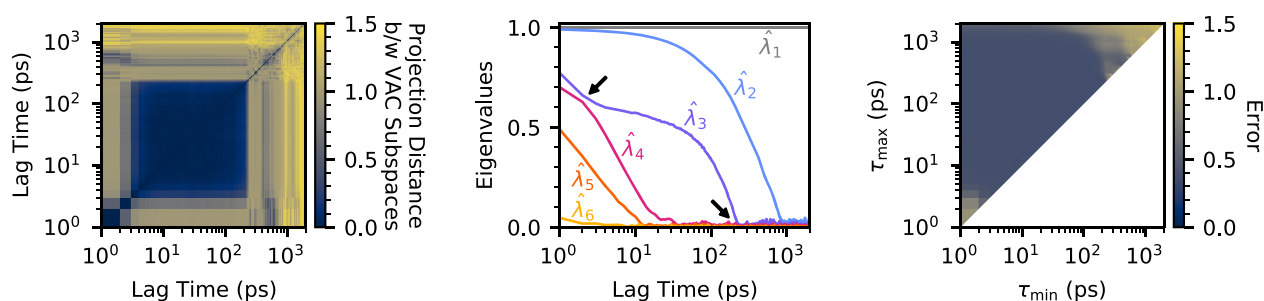


Figure 4. Lag time dependence of VAC and IVAC results. All results shown are for the single 20 ns alanine dipeptide trajectory in Figure 2. (Left) Projection distance between VAC results at the horizontal axis lag time and VAC results at the vertical axis lag time. (Center) First six estimated eigenvalues of the transition operator. (Right) Error in IVAC results at different values of τ_{\min} and τ_{\max} , evaluated using the projection distance to the MSM reference for the span of η_2 and η_3 .

range of lag times for which VAC exhibits low error levels. This supports our basic argument that choosing good parameters in IVAC is easier than choosing good parameters in VAC. To achieve low errors, we do not need to identify the optimal VAC lag times but only integrate over a window that contains the optimal VAC lag times while ensuring that τ_{\max} is not excessively high.

Application to the Villin Headpiece. Next we apply IVAC to a difficult spectral estimation problem with limited data. We seek to estimate the slow dynamics for an engineered 35-residue subdomain of the villin headpiece protein. Our data consist of a 125 μ s molecular dynamics simulation performed by Lindorff-Larsen et al.³⁴ Villin is a common model system for protein folding for both experimental and computational studies,^{34–37} where the top eigenfunctions correlate with the folding and unfolding of the protein.

On the surface, the villin data set would seem to be much larger and more useful for spectral estimation compared to the 10–20 ns trajectories we examined for the alanine dipeptide. However, the villin headpiece relaxes to equilibrium orders of magnitude more slowly than the alanine dipeptide. The data set contains just 34 folding/unfolding events with a folding time of 2.8 μ s. The limited number of observed events is characteristic of simulations of larger and more complex biomolecules, since simulations require massive computational resources and conformational changes take place slowly over many molecular dynamics time steps. The fact that the dynamics of villin are not understood nearly as well as the dynamics of the alanine dipeptide presents an additional challenge. Compared to the alanine dipeptide, villin has a more complex free energy surface and a larger number of degrees of freedom. Since the true

eigenfunctions of the system are unknown, it is appropriate to apply spectral estimation using a large and diverse feature set. However, the large size and diversity of the feature set increases the risk of estimation error.

In contrast to the alanine dipeptide results, where we applied IVAC using linear combinations of basis functions, here we apply IVAC using a neural network. The increased flexibility of the neural network approximation reduces approximation error. However, the procedure for optimizing the neural network is more complicated than the procedure for applying linear VAC. Moreover, the complexity of the neural network representation (around 5×10^4 parameters) makes overfitting a concern for this example.

We use a slight modification of the neural network architecture published in Sidky et al.,³⁸ with 2 hidden layers of 50 neurons, tanh nonlinearities, and batch normalization between layers. The network is built on top of a rich set of features, consisting of all the C_α pairwise distances as well as sines and cosines of all dihedral angles. At each optimization step, we subsample 10^4 data points using Algorithm 1. We optimize the neural network parameters using AdamW³⁹ with a learning rate of 10^{-4} and a weight decay coefficient of 10^{-2} . Following standard practice, we use the first half of the data set for training and the second half for validation. We validate the neural network against the testing data set every 100 optimization steps, and perform early stopping with a patience of 10.

We present our results for villin in two parts. First we describe our procedure for selecting parameters in nonlinear IVAC. Next we highlight evidence that nonlinear IVAC shows greater robustness to overfitting compared to nonlinear VAC.

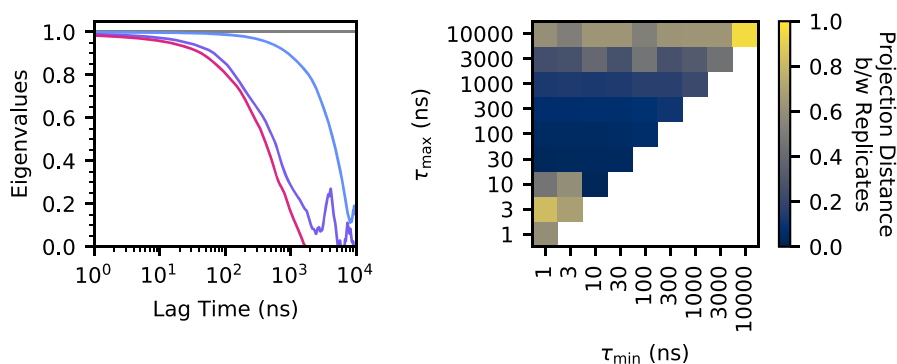


Figure 5. Nonlinear IVAC results for the 125 μ s villin headpiece trajectory. (Left) Estimated eigenvalues of the transition operator. (Right) Root mean square projection distance between 10 replicates of nonlinear IVAC at the specified values of τ_{\min} and τ_{\max} .

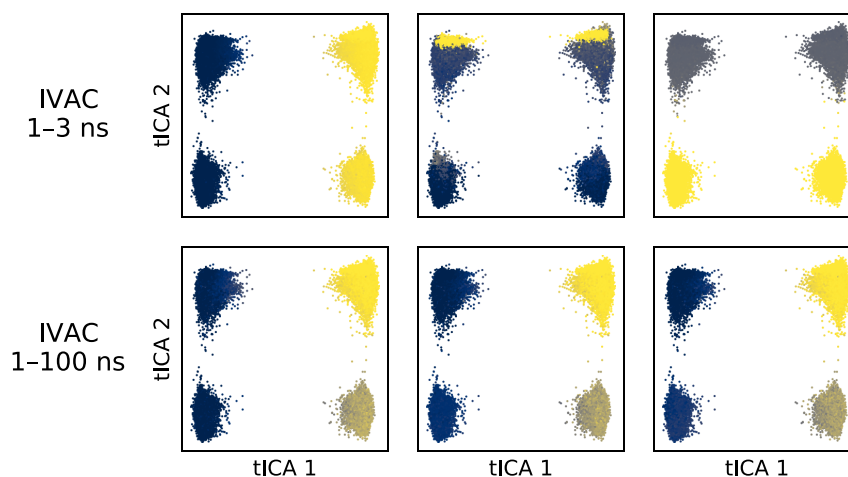


Figure 6. Nonlinear IVAC results plotted on the first two time-lagged independent component analysis (tICA) coordinates. (top) IVAC with a 1–3 ns integration window and three different random seeds. (bottom) IVAC with a 1–100 ns integration window and three different random seeds.

Selection of Parameters. Here, we describe the protocols we use for selecting IVAC parameters. By establishing clear protocols, we help ensure that IVAC performs to the best of its ability, providing robust eigenfunction estimates even in a high-dimensional setting with limited data.

Our first protocol is to evaluate the condition number for the subspace of eigenfunctions that we are estimating. This protocol is motivated by the theoretical error analysis in Webber et al.,¹² where we showed that spectral estimates are less sensitive to estimation error for a well-conditioned subspace. To ensure that we are estimating a well-conditioned subspace, we first use IVAC to estimate eigenvalues for the transition operator. We then identify a subspace of eigenfunctions $\eta_1, \eta_2, \dots, \eta_k$ that is separated from all other eigenfunctions by a large spectral gap $\hat{\lambda}_k^\tau - \hat{\lambda}_{k+1}^\tau$.

For the villin data, we choose the subspace consisting only of the constant eigenfunction $\eta_1 = 1$ and the first nontrivial eigenfunction η_2 . This is a well-conditioned subspace with a minimum condition number

$$\min_{\tau} \kappa^{\tau} = \min_{\tau} (\hat{\lambda}_2^{\tau} - \hat{\lambda}_3^{\tau})^{-1} = 1.6 \quad (40)$$

Our second protocol for ensuring robustness is to check that eigenfunction estimates remain consistent when the random seeds used in initialization and subsampling are changed. We train ten nonlinear IVAC neural networks and quantify the inconsistency in the results using the root-mean-square projection distance between eigenspace estimates from different

runs. The results of this calculation are plotted in Figure 5 across a range of τ_{\min} and τ_{\max} values. The results for VAC appear along the diagonal of the plot in Figure 5, corresponding to the case $\tau_{\min} = \tau_{\max}$.

Figure 5 reveals problems with consistency for both IVAC and VAC. IVAC is robust to the choice of τ_{\min} . However, setting $\tau_{\max} < 30$ ns or $\tau_{\max} > 300$ ns leads to poor consistency. If we train the neural network with these problematic τ_{\max} values, then solutions can look very different depending on the random seeds that are used for optimizing. With VAC, setting $\tau < 10$ ns or $\tau > 300$ ns would lead to inconsistent results.

IVAC provides more flexibility to address the consistency issues compared to VAC, since we can integrate over a range of lag times. For the villin data, we choose to set $\tau_{\min} = 1$ ns and $\tau_{\max} = 100$ ns. For these parameter values, the consistency score is very good. The typical projection distance between subspaces with different random seeds is just 0.05. Moreover, 1–100 ns is a wide range of lag times, helping to ensure that optimal or near-optimal VAC lag times are included in the integration window.

To help explain why the consistency is so poor for small τ_{\max} values, we present in Figure 6 a set of IVAC solutions obtained with an integration window of 1–3 ns and three different random seeds. We see that all three solutions identify clusters in the data, but the clusters are completely different in the three cases. We conjecture that IVAC is randomly fitting three different eigenspaces. This is supported by the eigenvalue plot in Figure 5, which shows that three nontrivial eigenvalues of the

transition operator lie close together over the 1–3 ns time window, making it possible that eigenspaces are randomly misidentified by IVAC.

In contrast to the inconsistent results obtained with an integration window of 1–3 ns, we obtain more reasonable results with an integration window of 1–100 ns. As shown in Figure 6, the IVAC solutions are nearly identical regardless of the random seed.

In summary, we propose a robust procedure for approximating eigenfunctions of the villin headpiece system. We choose to approximate a well-conditioned eigenspace that is separated from other eigenspaces by a wide spectral gap. Moreover, we ensure that IVAC results are consistent regardless of the random initialization and randomly drawn data subsets used to train the neural net. Because of these protocols, the neural network estimates shown in Figure 6 reliably identify clusters in the trajectory data indicative of folded/unfolded states.

Robustness to Overfitting. In this section, we present results suggesting that nonlinear IVAC is more robust to overfitting than nonlinear VAC. This is crucial if the data set is too small for cross-validation.

To identify the overfitting issue with small data sets, we eliminate the early stopping and we train IVAC and VAC until the training loss stabilizes. We calculate implied time scales by performing linear VAC on the outputs of the networks trained using IVAC and VAC, which we present in Figure 7.

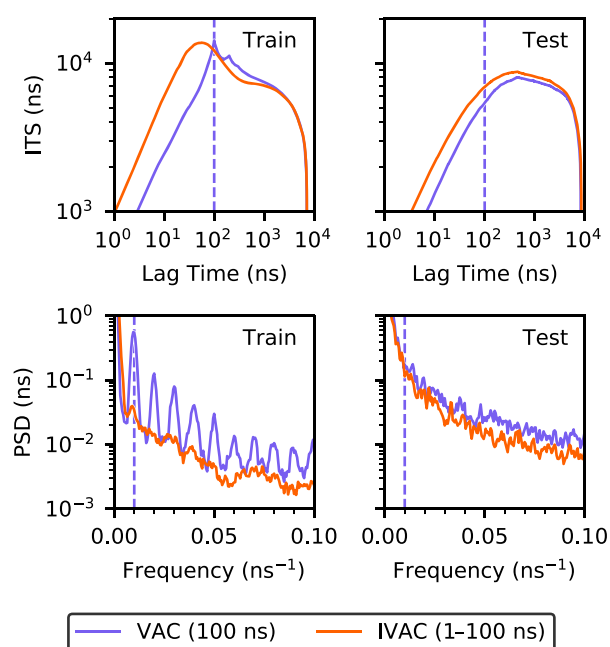


Figure 7. Implied time scales (ITS) and power spectral densities (PSD) obtained with nonlinear IVAC and VAC with neural network basis functions applied to the villin headpiece data set. The VAC training lag time is marked by the dotted line in each panel.

We first compare the estimated implied time scales between the training and validation data sets. For both algorithms, the implied time scales calculated on the training data are larger than those calculated on the validation data. This is clear evidence of overfitting. However, we see that IVAC gives larger implied time scales on the validation data compared to VAC. In combination with the variational principle associated with the implied time

scales, this suggests that IVAC is giving an improved estimate for the slow eigenfunctions.

Examining the implied time scales estimated on training data shows further signs of overfitting. The VAC implied time scale estimates for the training data exhibit sharp peaks at the training lag time that are absent in the implied time scale estimates of the validation data. This suggests a hypothesis for the mechanism of overfitting: with a sufficiently flexible approximation space, VAC is able to find spurious correlations between features that happen to be separated by τ . This explains the smaller peaks at integral multiples of the lag time, as features artificially correlated at τ will be correlated at 2τ as well.

To confirm our hypothesis, we plot the power spectral density (PSD)⁴⁰ of the time trace of eigenfunction estimates in Figure 7. The PSD confirms the existence of a periodic component in VAC results with a frequency at the inverse training lag time. In contrast, IVAC does not exhibit such a periodic component. In Figure 7, we see that the 1–100 ns integration window leads to implied time scale estimates that depend smoothly on the data both for the training and the test data set. The PSD shows no periodic components in the spectra for IVAC, providing further evidence that IVAC is comparatively robust while VAC results can be very sensitive to the particular lag time that is used.

CONCLUSION

In this paper we have presented integrated VAC (IVAC), a new extension to the popular variational approach to conformational dynamics (VAC). By integrating correlation functions over a window of lag times, IVAC provides robust estimates of the eigenfunctions of a system's transition operator.

To test the efficacy of the new approach, we compared IVAC and VAC results on two molecular systems. First, we applied the spectral estimation methods to simulation data from the alanine dipeptide. This is a relatively simple system that permits generation of extensive reference data for validating our calculations. As we varied the lag time parameters and the amount of data available, we observed the improved robustness of IVAC compared to VAC. IVAC gives low-error eigenfunction estimates even when the lag times range over multiple orders of magnitude. In contrast, VAC requires more precise lag-time tuning to give reasonable results.

Next we applied IVAC to analyze a folding/unfolding trajectory for the villin headpiece. These data contain relatively few folding/unfolding events despite pushing the limits of present computing technology. For this application, we used a flexible neural network representation built on top of a rich feature set. We presented a procedure for selecting parameters in IVAC that helps lead to robust performance in the face of uncertainty. For the application to villin data, we found that VAC exhibited pronounced artifacts from overfitting when precautions were not taken to specifically prevent it, while IVAC did not.

Our work highlights the sensitivity of VAC calculations to error from insufficient sampling. Examining our results on the villin headpiece, we see that regularization (here, by early stopping) and validation are crucial when running VAC with neural networks or other flexible approximation spaces. With insufficient regularization or poor validation these schemes easily overfit. Even for the alanine dipeptide example, where we employ a simple basis on a statistically well-conditioned problem, we see that VAC has a high probability of giving spurious results with insufficient data.

Integrated VAC addresses this problem by considering information across multiple time lags. Future extensions of the work could further leverage this information. For instance, employing a well-chosen weighting function within the integral in (5) could further decrease hyperparameter sensitivity. Additionally, future numerical experiments could point to improved procedures for selecting τ_{\min} and τ_{\max} values. Finally, we could integrate over multiple lag times in other formalisms using the transition operator, such as schemes that estimate committors and mean-first-passage times.³² These extensions would further strengthen the basic message of our work: combining information from multiple lag times leads to improved estimates of the transition operator and its properties.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c06477>.

Alanine dipeptide simulation details and error plots for additional individual alanine dipeptide trajectories and loss functions for villin nonlinear VAC and IVAC training (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Chatipat Lorpaiaboon – Department of Chemistry and James Franck Institute, University of Chicago, Chicago, Illinois 60637, United States; Email: chatipat@uchicago.edu

Erik Henning Thiede – Flatiron Institute, New York, New York 10037, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60637, United States; Email: thiede@uchicago.edu

Robert J. Webber – Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States; Email: rw2515@nyu.edu

Jonathan Weare – Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States; Email: weare@cims.nyu.edu

Aaron R. Dinner – Department of Chemistry and James Franck Institute, University of Chicago, Chicago, Illinois 60637, United States; orcid.org/0000-0001-8328-6427; Email: dinner@uchicago.edu

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c06477>

Author Contributions

#C.L., E.H.T., and R.J.W. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.H.T. was supported by DARPA grant HR00111890038. R.J.W. was supported by the National Science Foundation through award DMS-1646339. C.L., A.R.D., and J.W. were supported by the National Institutes of Health award R35 GM136381. J.W. was supported by the Advanced Scientific Computing Research Program within the DOE Office of Science through award DE-SC0020427. The villin headpiece data set was provided by D. E. Shaw Research. Computing resources where provided by the University of Chicago Research Computing Center.

■ REFERENCES

- (1) Noé, F.; Nüske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **2013**, *11*, 635–655.
- (2) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S.; Noé, F. Variational approach to molecular kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739–1752.
- (3) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102.
- (4) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **1994**, *72*, 3634.
- (5) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000–2009.
- (6) Klus, S.; Nüske, F.; Koltai, P.; Wu, H.; Kevrekidis, I.; Schütte, C.; Noé, F. Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.* **2018**, *28*, 985–1010.
- (7) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146–168.
- (8) Swope, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- (9) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. C.; Zheshkov, Y.; et al. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and a β -hairpin peptide. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.
- (10) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
- (11) Chen, W.; Sidky, H.; Ferguson, A. L. Nonlinear discovery of slow molecular modes using state-free reversible VAMPnets. *J. Chem. Phys.* **2019**, *150*, 214114.
- (12) Webber, R. J.; Thiede, E. H.; Dow, D.; Dinner, A. R.; Weare, J. Error bounds for dynamical spectral estimation. *arXiv:2005.02248* **2020**.
- (13) Takano, H.; Miyashita, S. Relaxation modes in random spin systems. *J. Phys. Soc. Jpn.* **1995**, *64*, 3688–3698.
- (14) Kallenberg, O. *Foundations of Modern Probability*; Springer Science & Business Media, 2006.
- (15) Eisner, T.; Farkas, B.; Haase, M.; Nagel, R. *Operator Theoretic Aspects of Ergodic Theory*; Springer, 2015; Vol. 272.
- (16) Hirao, H.; Koseki, S.; Takano, H. Molecular dynamics study of relaxation modes of a single polymer chain. *J. Phys. Soc. Jpn.* **1997**, *66*, 3399–3405.
- (17) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.
- (18) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99–105.
- (19) Vanden-Eijnden, E. Transition path theory. In *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R.; Pande, V. S.; Noé, F., Eds.; Springer, 2014; pp 91–100.
- (20) Noé, F.; Prinz, J.-H. Analysis of Markov models. In *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Bowman, G. R.; Pande, V. S., Noé, F., Eds.; Springer, 2014; pp 75–90.
- (21) Keller, B. G.; Aleksic, S.; Donati, L. In *Biomolecular Simulations in Drug Discovery*; Gervasio, F. L., Spiwok, V., Eds.; Wiley-VCH, 2019; Chapter 4.
- (22) Vitalini, F.; Noé, F.; Keller, B. A basis set for peptides for the variational approach to conformational kinetics. *J. Chem. Theory Comput.* **2015**, *11*, 3992–4004.

- (23) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating molecular kinetics by variationally optimized diffusion maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947–5960.
- (24) Schwantes, C. R.; McGibbon, R. T.; Pande, V. S. Perspective: Markov models for long-timescale biomolecular dynamics. *J. Chem. Phys.* **2014**, *141*, 090901.
- (25) Wu, H.; Nüske, F.; Paul, F.; Klus, S.; Koltai, P.; Noé, F. Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations. *J. Chem. Phys.* **2017**, *146*, 154104.
- (26) McGibbon, R. T.; Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **2015**, *142*, 124105.
- (27) Naritomi, Y.; Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J. Chem. Phys.* **2011**, *134*, 06S101.
- (28) Husic, B. E.; Pande, V. S. Note: MSM lag time cannot be used for variational model selection. *J. Chem. Phys.* **2017**, *147*, 176101.
- (29) Wu, H.; Prinz, J.-H.; Noé, F. Projected metastable Markov processes and their estimation with observable operator models. *J. Chem. Phys.* **2015**, *143*, 144101.
- (30) Suárez, E.; Adelman, J. L.; Zuckerman, D. M. Accurate estimation of protein folding and unfolding times: beyond Markov state models. *J. Chem. Theory Comput.* **2016**, *12*, 3473–3481.
- (31) Cao, S.; Montoya-Castillo, A.; Wang, W.; Markland, T. E.; Huang, X. On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J. Chem. Phys.* **2020**, *153*, 014105.
- (32) Thiede, E. H.; Giannakis, D.; Dinner, A. R.; Weare, J. Galerkin approximation of dynamical quantities using trajectory data. *J. Chem. Phys.* **2019**, *150*, 244111.
- (33) Edelman, A.; Arias, T. A.; Smith, S. T. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications* **1998**, *20*, 303–353.
- (34) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.
- (35) McKnight, J. C.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. A Thermostable 35-residue subdomain within villin headpiece. *J. Mol. Biol.* **1996**, *260*, 126.
- (36) Kubelka, J.; Eaton, W. A.; Hofrichter, J. Experimental tests of villin subdomain folding simulations. *J. Mol. Biol.* **2003**, *329*, 625–630.
- (37) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740–744.
- (38) Sidky, H.; Chen, W.; Ferguson, A. L. High-resolution Markov state models for the dynamics of Trp-cage miniprotein constructed over slow folding modes identified by state-free reversible VAMPnets. *J. Phys. Chem. B* **2019**, *123*, 7999–8009.
- (39) Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv:1711.05101* **2017**.
- (40) Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on. IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73.