JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

pubs.acs.org/JCTC

Rapid Calculation of Molecular Kinetics Using Compressed Sensing

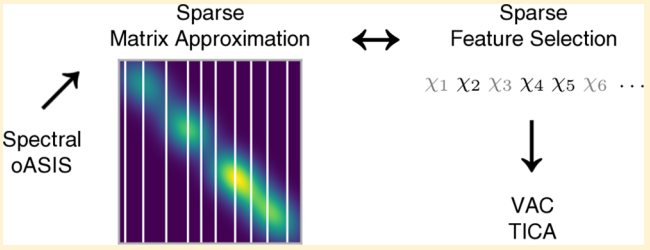
Florian Litzinger,[†] Lorenzo Boninsegna,[‡] Hao Wu,[†] Feliks Nüske,[‡] Raajen Patel,[§] Richard Baraniuk,[§] Frank Noé,[*,†,‡] and Cecilia Clementi[*,‡]

[†]Freie Universität Berlin, Department of Mathematics and Computer Science, Arnimallee 6, 14195 Berlin, Germany

[‡]Rice University, Center for Theoretical Biological Physics and Department of Chemistry, Houston, Texas 77005, United States

[§]Rice University, Department of Electrical and Computer Engineering, Houston, Texas 77005, United States

ABSTRACT: Recent methods for the analysis of molecular kinetics from massive molecular dynamics (MD) data rely on the solution of very large eigenvalue problems. Here we build upon recent results from the field of compressed sensing and develop the spectral oASIS method, a highly efficient approach to approximate the leading eigenvalues and eigenvectors of large generalized eigenvalue problems without ever having to evaluate the full matrices. The approach is demonstrated to reduce the dimensionality of the problem by 1 or 2 orders of magnitude, directly leading to corresponding savings in the computation and storage of the necessary matrices and a speedup of 2 to 4 orders of magnitude in solving the eigenvalue problem. We demonstrate the method on extensive data sets of protein conformational changes and protein−ligand binding using the variational approach to conformation dynamics (VAC) and time-lagged independent component analysis (TICA). Our approach can also be applied to kernel formulations of VAC, TICA, and extended dynamic mode decomposition (EDMD).



1. INTRODUCTION

Molecular dynamics (MD) simulation has become an essential instrument in the study of macromolecular systems. The use of ensemble simulation methods supported by novel hardware and middleware solutions in addition to distributed computational resources has been essential for achieving the sampling of processes spanning time scales of biological relevance.[1−6] However, the ability to generate large amounts of molecular dynamics data has created a need for tools that help analyze and interpret these data quickly and reliably. Given a large MD data set, one would like to extract molecular mechanisms as well as quantities that can be related to experimental measurements, such as the structures of long-lived states, their equilibrium probabilities and lifetimes, and the transition rates between them.

Recent years have seen a surge of interest in kinetic models to compute such structural and quantitative information, including Markov state models (MSMs) or Master-equation models,[7−14] multiensemble Markov models,[15−18] diffusion maps,[19,20] and VAMPnets.[21] The variational approach of conformation dynamics (VAC)[22,23] has shown that the approximation of the molecular kinetics, including transition rates and structural mechanisms, can be cast as a problem of combining basis sets in molecular state space. The recently developed variational approach of Markov processes (VAMP)[24] has generalized these results to dynamical processes that are out of equilibrium.

The key step in the VAC is the computation of matrices of time-correlations between basis functions and the solution of a generalized eigenvalue problem with these matrices. This approach is so general that many kinetic models can be cast as special cases of it, including MSMs, Master-equation models, VAMPnets,[21] core MSMs,[11,25] Markov transition models,[26] the time-lagged independent component analysis (TICA),[27,28] kernel TICA,[29] and kinetic maps.[30,31] Moreover, methods to extract dynamical components from high-dimensional time series have been developed in other fields as well, leading to equivalent mathematical problems. These include blind source separation,[32,33] dynamic mode decomposition,[34,35] extended dynamic mode decomposition (EDMD),[36] and their kernel formulations[37,38]—see refs 39 and 40 for recent reviews.

These developments suggest that the step of data clustering used in MSMs could be overcome if the state space were instead covered by many basis or kernel functions and the corresponding variational problem could be solved. Unfortunately, large MD data sets comprise millions of sampled configurations, and at present variational problems of only a few thousand samples can be solved efficiently. The main bottleneck is the calculation of the covariance matrices themselves that is linear in time and quadratic in the number of dimensions. Since these large matrices are dense, it is intractable to even compute all matrix entries and sparse eigenvalue solvers such as Krylov subspace methods are not efficient. The problem is even more dramatic in kernel approaches, such as diffusion maps, kernel TICA, or kernel EDMD, where the matrix sizes are quadratic in the number of timesteps. Unless most data points are discarded, this would

Received: January 29, 2018
Published: April 16, 2018

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

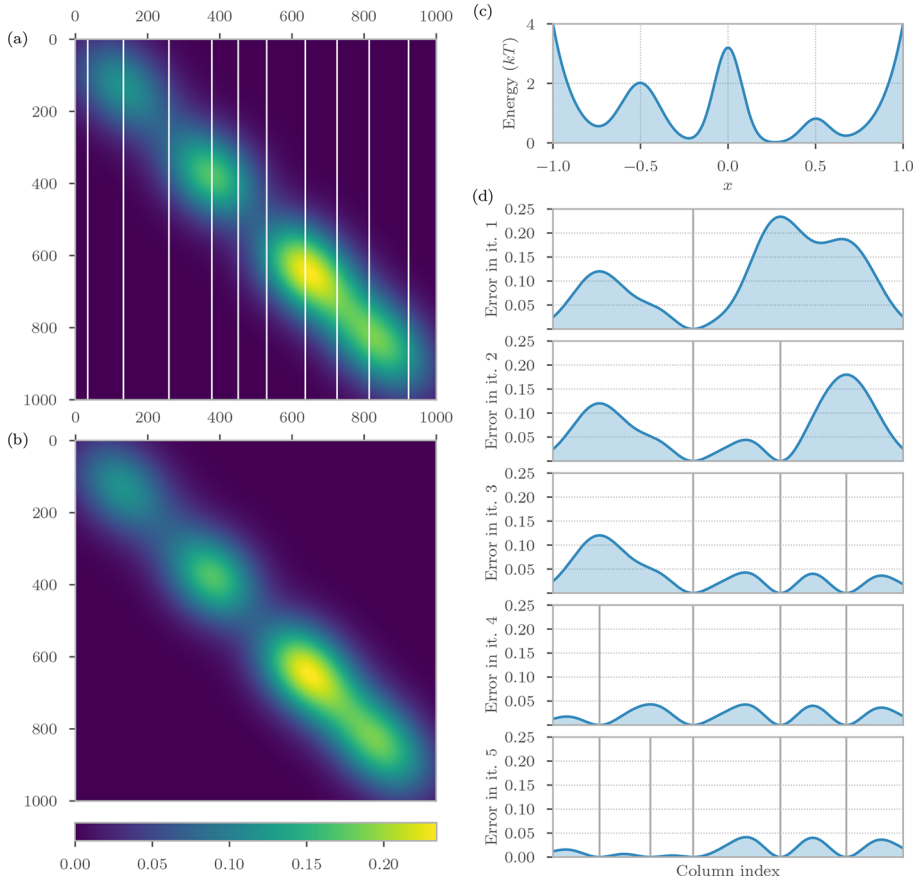


Figure 1. Sparse matrix approximation using the Nyström method. (a) 1000 × 1000 correlation matrix and a subset of 10 columns (white) chosen by an adaptive method. The correlation matrix was computed as described in the text (see section 3.1). (b) Reconstruction of the full correlation matrix using only the 10 chosen columns. The reconstruction is almost perfect, using only 1% of the data. Using this principle, eigenvalues and eigenvectors of large matrices can be approximated with very little computational effort. (c) Energy landscape of the Prinz potential[13] that was used to generate the covariance matrix in a and b. (d) Illustration of the oASIS algorithm. The gray area shows the oASIS reconstruction error of the diagonal elements of $\mathbf{C}(0)$. In the first iteration, a random column of $\mathbf{C}(0)$ is chosen and the Nyström approxim is made. The oASIS error at this column drops to zero. In each subsequent iteration, the column with the largest oASIS error is chosen. The first four iterations choose columns in different metastable states; then transition states are selected.

lead to matrices of several millions of rows and columns, which cannot be efficiently computed or stored.

At the same time, compressed sensing and sparse sampling methods[41,42] have emerged in mathematical research. This class of methods aims at accurately reconstructing a signal from sparse (i.e., compressed) samples. Sparse sampling has been applied successfully in different fields, such as image segmentation and matrix factorization.[43] Here we employ sparse sampling in the column space of the covariance matrices with the Nyström method in order to obtain a low-rank approximation of the matrices.[44,45]

The Nyström approximation has previously been used in related application areas. It has first been used in conjunction with Markov state models in order to define a coarse-graining (lumping) of states.[46] More recently, it has been employed to make the kernel TICA approach tractable.[47] The main difficulty, however, lies in making the choice of matrix columns such that the computed matrix eigenvalues and eigenvectors are sufficiently accurate. Recently, it has been shown that adaptive sampling strategies, which iteratively choose matrix columns one-by-one, enable highly accurate eigenfactorization approximations without having to compute the entire matrix.[48] For example, Figure 1 demonstrates the accurate reconstruction of a 1000 × 1000 matrix from only 10 of its columns (i.e., 1% of the data).

While adaptive sampling is far more efficient than random column selection, applying this idea to MD data analysis is hampered by two problems: (i) Each adaptive iteration requires a pass over the data and one would thus lose the gained efficiency by spending more I/O time; (ii) It is unknown how to efficiently approximate the eigenvalues and eigenvectors of a generalized eigenvalue problem in a sparse sampling framework, as this involves the simultaneous approximation of two matrices. Here we address (i) by developing a spectral adaptive method that samples columns suitable for approximating specific eigenvalue/eigenvector pairs. This is generalized to a spectral sampling method that selects m columns simultaneously to approximate the first m eigenvalue/eigenvector pairs, thus reducing the number of passes over the data. We address (ii) by arguing that a selection of columns for the Nyström approximation of the instantaneous correlation matrix can be understood as a selection of a subset of basis functions in the variational approach, whence approximate generalized eigenvalues and eigenvectors may be computed from a small-scale problem derived from a sparse approximation of the overlap matrix.

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

We demonstrate the validity of our approach and show that the otherwise prohibitively expensive analysis of complex macromolecular systems becomes feasible. The versatility of our approach is illustrated by applying it to two popular analysis approaches for MD data: the variational principle of conformation dynamics (VAC) and the time-lagged independent component analysis (TICA). We demonstrate the efficiency on extensive simulation data of protein conformational changes in bovine pancreatic trypsin inhibitor (BPTI)[49] and protein−ligand binding in the Trypsin−Benzamidine complex,[50,51] achieving a dimension reduction down to between 20% and 1% and thus a gain in efficiency of 2 to 4 orders of magnitude without significant accuracy loss.

Our method is implemented in PyEMMA (version 2.5.2 or later).

2. THEORY AND METHODS

2.1. Molecular Dynamics (MD). MD simulation can be described as a Markov process in a state space Ω sampling from an equilibrium distribution $\pi(\mathbf{x})$, assumed here to be the Boltzmann distribution. An MD trajectory $\{\mathbf{x}_t\}$ is a stochastic realization of this process. There is a probability density $p_\tau(\mathbf{y}|\mathbf{x})$ of making a transition, that is, to find the system in state $\mathbf{y}$ at a later time $t + \tau$ during the MD trajectory, given that it is in state $\mathbf{x}$ at time t. Under mild conditions, the time-evolution of probability distributions of molecular structures can be described with a finite number of eigenvalues and eigenfunctions of the so-called backward propagator (see the literature[13] for details):

$$u_{t+\tau} \approx \sum_{i=1}^{n} \mathbf{e}^{-\tau\kappa_i} \langle \psi_i | u_t \rangle_\pi \psi_i \tag{1}$$

where $\mathrm{e}_i^{-\tau\kappa} = \lambda_i(\tau)$ is the i-th eigenvalue, which decays with relaxation rate κ_i or time scale $t_i = \kappa_i^{-1}$, ψ_i is the corresponding eigenfunction, and $\langle \psi_i | u_t \rangle_\pi = \int \psi_i(\mathbf{x}) u_t(\mathbf{x}) \pi(x)\, \mathrm{d}\mathbf{x}$ is a weighted scalar product. The function u_t can be thought of as an indicator function identifying which conformations the system resides in at time t. By multiplying it with the stationary distribution, $\rho_t = \pi u_t$, it becomes the instantaneous probability density at time t; that is, ρ_t measures what fraction of the population is in which conformation for an ensemble of copies of a molecular system.

The first process is stationary, with rate $\kappa_1 = 0$, and has the associated constant eigenfunction $\psi_1 = 1$, while all other rates are positive: $0 = \kappa_1 < \kappa_2 \leq \ldots$. Thus, for $\tau \to \infty$, eq 1 becomes $u_\infty = 1$ and $\rho_\infty = \pi$; that is, the dynamics relax toward the Boltzmann density.

If we knew the significant eigenvalues λ_i and corresponding eigenfunctions ψ_i, we would have a complete description of the kinetic and equilibrium properties of the molecular system and could compute the long-lived structures, their equilibrium probabilities or free energies, and the transition rates between them.[13] Unfortunately, the eigenvalue/eigenfunction pairs (λ_i, ψ_i) are not directly available but must be estimated using trajectory samples $\{\mathbf{x}_t\}$.

2.2. Approximation Methods for Eigenvalues and Eigenfunctions. Different approaches can be used to approximate the eigenvalue/eigenfunction pairs (λ_i, ψ_i). Here we discuss two of the most popular methods.

2.2.1. Variational Approach to Conformation Dynamics (VAC). A very general variational approach has been proposed[22] to approximate the eigenfunctions and eigenvalues of the MD backward propagator. The idea is similar to the well-known variational principle used in quantum mechanics to approximate the ground state eigenfunction and energy of a given Hamiltonian operator. The variational principle states the following: Suppose we construct a set of n orthogonal trial functions $\hat{\psi}_1, \ldots, \hat{\psi}_n$, $\langle \hat{\psi}_i | \hat{\psi}_j \rangle_\pi = \delta_{ij}$. Then their normalized autocorrelation functions are Rayleigh coefficients and will systematically underestimate the true eigenvalues:

$$\hat{\lambda}_i = \frac{\langle \hat{\psi}_i(\mathbf{x}_t) \hat{\psi}_i(\mathbf{x}_{t+\tau}) \rangle_t}{\langle \hat{\psi}_i^2(\mathbf{x}_t) \rangle_t} = \frac{\langle \hat{\psi}_i, \mathcal{T}(\tau) \hat{\psi}_i \rangle_\pi}{\langle \hat{\psi}_i, \hat{\psi}_i \rangle_\pi} \leq \lambda_i \tag{2}$$

where the equality holds only if the trial functions are the true eigenfunctions $\hat{\psi}_i(\mathbf{x}) = \psi_i(\mathbf{x})$, and the approximation is better when the eigenvalue estimates $\hat{\lambda}_i$ are larger. This leads to the following variational principle: Given a set of basis functions $(\chi_1(\mathbf{x}), \ldots, \chi_M(\mathbf{x}))$, an approximation of the eigenfunctions can be constructed as a linear combination of such functions

$$\hat{\psi}_i(\mathbf{x}) = \sum_{j=1}^{M} a_{ij} \chi_j(\mathbf{x})$$

The optimal linear combination coefficients $\mathbf{a}_i = \{a_{ij}\}_{j=1}^{M}$ that simultaneously maximize each $\hat{\lambda}_i$ in the orthogonal set of functions (and at the same time maximize the partial eigensum $\sum_{i=1}^{n} \hat{\lambda}_i$)[22,52] is given by the solution of the generalized eigenvalue problem

$$\mathbf{C}(\tau)\mathbf{a}_i = \mathbf{C}(0)\mathbf{a}_i \lambda_i \tag{3}$$

where $\mathbf{C}(\tau)$ and $\mathbf{C}(0)$ are correlation matrices. When a realization of the trajectory $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T\}$ is given, these correlation matrices can be approximated by time averages:

$$\begin{aligned} c_{ij}(0) &\approx \langle \chi_i(\mathbf{x}_t) \cdot \chi_j(\mathbf{x}_t) \rangle_t \\ c_{ij}(\tau) &\approx \langle \chi_i(\mathbf{x}_t) \cdot \chi_j(\mathbf{x}_{t+\tau}) \rangle_t . \end{aligned} \tag{4}$$

In many cases one would like to use a very large basis set, as clearly the quality of the approximation depends on the quality of the basis set. Therefore, one is presented with a generalized eigenvalue problem involving huge matrices and this forms the bottleneck for the practical application of this technique. We show below that the Nyström approach can be used to tackle this problem.

2.2.2. Time-Lagged Independent Component Analysis (TICA).[27,28,32] TICA is a dimensionality reduction method that can be considered as a special case of the application of the variational principle presented above. In practice, given MD data, TICA performs a variational optimization using a set of input features $y_i(\mathbf{x}_t)$ (distances, angles, ...) and defining mean-free basis functions as

$$\chi_i(\mathbf{x}_t) = y_i(\mathbf{x}_t) - \langle y_i \rangle_t$$

These functions are then used to solve the generalized eigenvalue problem (eq 3) and find the optimal coefficients to obtain an approximation of the eigenfunctions as a linear combination of the input features. The low dimensional space spanned by the first few TICA coordinates provides an approximation for the subspace where the system's slowest processes live. Although TICA only provides a rather rough approximation to individual eigenvalues and eigenfunctions, the dominant space of eigenfunctions is generally well represented.

For this reason the first few TICA coordinates can be used to define a distance metric in the clustering stage in the

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

construction of Markov state models,[30,31] or in the diffusion map approach.[53] It has been shown recently[54] that heavy-atom distances are an excellent set of features to conduct a TICA analysis. Unfortunately, even for a small to medium protein (with about 1000 heavy atoms) there are nearly n = 500,000 distinct distances, making the $\sim Nn^2$ effort of computing the elements of the correlation matrices on a trajectory of N time frames in eq 4 completely intractable. Similarly, kernel TICA[29] requires the calculation of correlation matrices with $n = 2N$ dimensions, which also leads to intractable problems unless the MD data are massively downsampled.

In both of the above approaches, huge eigenvalue problems need to be solved in order to achieve a small discretization error. Note that the main computational problem is actually not the solution of eq 3 (which is $O(n^3)$ for the full solution and $O(n^2)$ for single eigenvalue/eigenvector pairs) but rather the calculation of the matrix entries (which is $O(Nn^2) \geq O(n^3)$). The key to solving the computational problem therefore does not lie in the eigenvalue solver but rather in ways to avoid computing the full correlation matrices.

2.3. Nyström Approximation and Incomplete Cholesky Methods. A common approach to avoid the computation and storage of full symmetric matrices is to use the Nyström approximation.[45] Suppose that S is a set of column indices of the symmetric matrix $\mathbf{C} \in \mathbb{R}^{n\times n}$ with $|S|=k \ll n$ and forms the column-submatrix $\mathbf{C}_k = \mathbf{C}[:, S] \in \mathbb{R}^{n\times k}$. The Nyström approximation $\tilde{\mathbf{C}} \approx \mathbf{C}$ is then given by

$$\tilde{\mathrm{C}} = \mathbf{C}_k \mathbf{W}_k^{-1} \mathbf{C}_k^{\mathrm{T}}$$

where S is assumed to be such that the quadratic matrix $\mathbf{W}_k = \mathbf{C}_k[S, :] \in \mathbb{R}^{k\times k}$ is invertible. It can be shown that the leading eigenvalues λ_i and eigenvectors $\mathbf{u}_i$ of $\mathbf{C}$ can be approximated by a scaled version of the corresponding eigenvalues $\lambda_i^{(k)}$ and eigenvectors $\mathbf{u}_i^{(k)}$ of $\mathbf{W}_k$ as follows:

$$\lambda_i \approx \tilde{\lambda}_i = \frac{n}{k}\lambda_i^{(k)},$$

$$\mathbf{u}_i \approx \tilde{\mathrm{u}}_i = \sqrt{\frac{k}{n}}\frac{1}{\lambda_i^{(k)}}\mathbf{C}_k\mathbf{u}_i^{(k)}$$

Since $\mathbf{W}_k \in \mathbb{R}^{k\times k}$, this computation is much faster than solving the eigenvalue problem directly for the full matrix $\mathbf{C} \in \mathbb{R}^{n\times n}$, reducing the complexity from $O(n^3)$ to $O(k^3)$ with $k \ll n$.

The quality of the Nyström approximation evidently depends on which k columns are selected. Adaptive column sampling methods work by iteratively selecting columns in a greedy way so as to maximize the accuracy of the Nyström approximation with each selection. An optimal selection can only be made if the full matrix $\mathbf{C}$ is known, rendering the method useless for our purposes. Alternatively, it is possible to use adaptive column sampling strategies that intelligently choose which columns to sample even *before* these columns have been formed. This is possible by exploiting the symmetry of the matrix $\mathbf{C}$, as in this case sampling a number of columns is equivalent to sampling a subset of rows. These rows give us partial information about the unseen columns that can be used to make informed selections for subsequent columns.

The recently introduced oASIS method[48] provides an excellent choice of the k columns, and only the n diagonal elements of $\mathbf{C}$ need to be precomputed. The Nyström approximation produced by this adaptive method is equivalent to the more classical incomplete Cholesky decomposition (ICD).[55,56] Starting with a small initial number k of randomly selected columns, oASIS computes the error Δ_i of predicting each diagonal element i by virtue of the current Nyström approximation:

$$\Delta_i = c_{ii} - \mathrm{b}_i^{\mathrm{T}}\mathbf{W}_k^{-1}\mathbf{b}_i \tag{5}$$

where $\mathbf{b}_i = \mathbf{C}[S, i] \in \mathbb{R}^k$ indicates the k entries of the i-th column of $\mathbf{C}$ that are already available (remember that $\mathbf{C}$ is symmetric). The column with the maximal error is then selected as the next column to add, and the procedure is repeated until the diagonal error becomes smaller than a given cutoff or until some convergence criterion (for instance on the resulting eigenvalues) is satisfied.

oASIS can provide high-quality approximations while using only a very few columns. Figure 1 demonstrates this by reconstructing the covariance matrix $\mathbf{C}(0)$ generated by the Prinz potential[13] with only 10 out of 1000 columns. Whenever the oASIS method selects a column, the diagonal prediction error (eq 5) drops to zero at this column. Interestingly, the columns within the same metastable set (energy minimum) are highly correlated—when a column from one metastable set is selected, the errors of other columns in the same metastable state also drop. Hence, the first four columns selected by the oASIS procedure are selected to be close to the minima of the four metastable sets (Figure 1d).

2.4. Spectral oASIS. For our purpose of the analysis of MD simulation data, we propose a modified version of oASIS that we call *spectral oASIS*. The motivation for spectral oASIS is that when working with massive data, it is often not possible to store the entire data set in memory. In this case, one typically streams through data, reading one chunk at a time, and then uses this chunk to update the calculation of columns of $\mathbf{C}$. This procedure requires significant time to read the data from disk. It is thus undesirable or prohibitive to only add one column per pass through the data, as the resulting I/O cost will most likely dominate the computation and the computational advantage of only selecting a subset of columns to obtain a good approximation to the eigenvalue problem is lost.

Hence we aim at selecting multiple columns at a time. We start by m randomly selected columns and then add m new columns in each round—this approach is called batch selection. Unfortunately the oASIS method is not suitable for batch selection, as illustrated by Figure 1d. If m columns would be selected in iterations 2 to 5 based on the diagonal prediction error shown in Figure 1d, all columns in a given iteration would be in the same metastable set. Thus, only one column from the batch is informative, while the remaining $m-1$ columns are ineffective as they are redundant with the first selected column. This intuition is confirmed by Figure 2. While oASIS is effective when choosing one column at a time (Figure 2a), its efficiency is lost and may become no better than random when using oASIS with batch selection (Figure 2b).

Based on the intuition from Figure 1d that informative columns not only have a high diagonal reconstruction error (eq 5) but also are in distinct metastable states, we define the following spectral oASIS error:

$$\Delta_i^{(j)} = \Delta_i \cdot (\psi_j)_i \tag{6}$$

where the diagonal error for each column i is scaled by the corresponding component of the j-th eigenvector. When

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

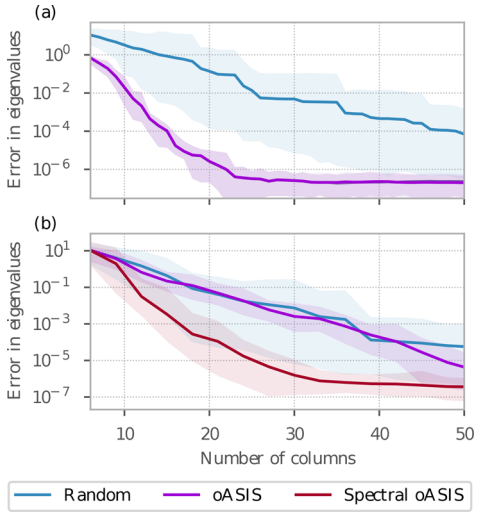


Figure 2. Comparison of different column selection methods evaluated on the covariance matrix shown in Figure 1: random selection, oASIS, and spectral oASIS. Errors (2-norm) of the four largest eigenvalues of $\mathbf{C}(0)$ are shown. (a) Single column selection (oASIS and spectral oASIS are equivalent in this case). (b) Batch selection of three columns at a time.

selecting m columns in a batch, we will simultaneously consider the first m eigenvectors. In spectral oASIS, we first select the row vector with maximum error in the first m eigenvectors and its index i defines the column which will be first selected and computed. As a second choice we select another vector with large spectral error, but as different as possible from the already selected vector, etc. With this approach we avoid making redundant choices within one selection round.

More formally, spectral oASIS is implemented as follows. We first define the weight matrix

$$\Delta = [\Delta^{(1)}, \Delta^{(2)}, \ldots, \Delta^{(m)}]$$

where

$$\Delta^{(j)} = (\Delta_1^{(j)}, \ldots, \Delta_n^{(j)})^{\mathrm{T}}$$

and $\Delta_i^{(j)}$ is defined in eq 6 above.

We then use the following column selection algorithm:

1 Select m initial columns at random without replacement. Perform Nyström approximation and calculate $\mathbf{\Delta}$
2 Define the set of selected row vectors V, and initialize it with a zero vector $V = \{0^{\mathrm{T}}\}$.
3 Repeat until convergence:
  (a) For $j = 1, \ldots, m$:
    i. Select column $c = \mathbf{argmax}_{c'} \sum_i \|V_i - \Delta_{c'}\|^2$
    ii. Add column to $V \leftarrow V + c$

Figure 2b compares the results obtained using batch selection with oASIS and spectral oASIS and random selection, for the same matrix as shown in Figure 1. While the efficiency of oASIS is lost when performing batch selection, spectral oASIS is still highly efficient and outperforms both regular oASIS and random selection in terms of the eigenvalue approximation error.

2.5. Variational Interpretation of the Nyström Approximation. The key insight for the derivation of a small-scale generalized eigenvalue problem whose eigenpairs approximate those of the full problem lies in the fact that the Nyström approximation of the overlap matrix $\mathbf{C}(0)$ in eq 4 allows an interpretation in terms of the variational approach. In fact, it has been pointed out[57] that the Nyström approximation $\tilde{\mathbf{C}}$ of a Gram matrix $\mathbf{C}=\mathbf{X}^{\mathrm{T}}\mathbf{X}$ using k columns $\mathbf{C}_k = \mathbf{C}[:,S]$, $|S| = k$, can be expressed using the orthogonal projection $\mathbf{P}_k = \mathbf{X}_k\mathbf{X}_k^{+}$ onto the selected data points $\mathbf{X}_k = \mathbf{X}[:,S]$, that is,

$$\tilde{\mathbf{C}} = \mathbf{X}^{\mathrm{T}}\mathbf{P}_k\mathbf{X} = (\mathbf{P}_k\mathbf{X})^{\mathrm{T}}(\mathbf{P}_k\mathbf{X})$$

In the same spirit, we may consider the projection $\mathcal{P}_k$ in function space onto the basis functions $\{\chi_i : i \in S\}$ used in the definition of the correlation matrices (eq 4). In Appendix A we argue that the overlap matrix constructed from the projected basis functions $\mathcal{P}_k\chi_i$ corresponds to the Nyström approximation $\tilde{\mathbf{C}}$ of $\mathbf{C}(0)$, whereby we conclude that the choice of columns for the Nyström approximation of the overlap matrix is directly linked to a choice of a subset of basis functions for the variational approach, to which again the variational principle applies. In essence, this means that a good approximation of the overlap matrix should provide us with a selection of basis functions in whose span the sought-after eigenfunctions can be well approximated.

2.6. Sparse Sampling of the Generalized Eigenvalue Problem. As a consequence of the previous section we propose the following new method to approximate the eigenpairs of the MD propagator: Instead of applying the variational approach to all basis functions $\chi_1, \ldots, \chi_M$, we directly employ the subset $\{\chi_i : i \in S\}$ corresponding to the selected indices S. Note that the space spanned by these k functions is identical to the space spanned by the projected basis functions $\mathcal{P}_k\chi_1, \ldots, \mathcal{P}_k\chi_M$ by construction. The result is the k-by-k generalized eigenvalue problem

$$\mathbf{C}^k(\tau)\mathbf{a}_i = \mathbf{C}^k(0)\mathbf{a}_i\lambda_i$$

where $\mathbf{C}^k(0), \mathbf{C}^k(\tau) \in \mathbb{R}^{k\times k}$ denote the instantaneous and time-lagged correlation matrices of the k selected basis functions $\{\chi_i : i \in S\}$. Since this method is a direct application of the variational approach, albeit to a smaller set of basis functions, the generalized eigenpairs retain their well-known meaning. The eigenfunctions of the propagator, for example, are approximated by the linear combination

$$\hat{\psi}_i(\mathbf{x}) = \sum_{j=1}^{k} a_{ij}\chi_{S(j)}(\mathbf{x})$$

Algorithmically, we proceed as follows. First, by means of the spectral oASIS method, we construct a Nyström approximation of the overlap matrix $\mathbf{C}(0)$,

$$\mathbf{C}(0) \approx \mathbf{C}_k(0)\mathbf{W}_k^{-1}(0)\mathbf{C}_k^{\mathrm{T}}(0)$$

and store the chosen indices in the set S. Second, we note that, in our notation above,

$$\mathbf{C}^k(0) = \mathbf{W}_k(0)$$

so that the required entries of the overlap matrix have already been computed over the course of the oASIS algorithm. Moreover, we compute the small-scale time-lagged correlation matrix

$$\mathbf{C}^k(\tau) = \mathbf{C}(\tau)[S, S]$$

Finally, we solve the above generalized eigenvalue problem to yield the eigenpairs $(\lambda_i, \mathbf{a}_i)$ which can then be used in further analysis as usual.

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

2.7. Alternative Methods for the Sparse Sampling of a Generalized Eigenvalue Problem. Besides the direct application of the VAC to the small basis set as determined by the selected columns of the overlap matrix $\mathbf{C}(0)$ presented above, we have considered three other possibilities to obtain a small-scale generalized eigenvalue problem whose solutions approximate those of the full problem (3). In short, they are based on the following: (i) a Nyström approximation of the time-lagged correlation matrix $\mathbf{C}(\tau)$, (ii) the CUR decomposition, a matrix decomposition more general than the Nyström approximation (that is restricted to symmetric positive semidefinite matrices), and (iii) the definition of a projected generalized eigenvalue problem in a lower dimensional space. A detailed description and comparison of these alternative methods is given in Appendix B. We assess their accuracy and compare them to the approach introduced above by means of the numerical examples of the following section. Although they still achieve significant dimensionality reduction in the practical examples, overall they appear less robust than the procedure presented above. Therefore, only the latter is used in the applications discussed below.

3. RESULTS

3.1. Illustration of a Low-Dimensional Dynamical Model. We first demonstrate the accuracy of our approach on a one-dimensional dynamical system for which the exact results can be computed without trajectory sampling. Figure 3a shows the potential of the model with four metastable states introduced in Prinz et al.[13] We approximate a diffusion process on this potential as described in Appendix C and define a basis set using 1000 Gaussian basis functions along the x coordinate. The resulting correlation matrix $\mathbf{C}(0)$ has been used in Figures 1 and 2. Using the VAC, we express the eigenfunctions of the system in this basis set, resulting in a nearly exact agreement between the true eigenfunctions and the variational approximation solving the full 1000-dimensional generalized eigenvalue problem (Figures 3c–f, compare light blue and violet).

The oASIS method was used to select columns of the overlap matrix $\mathbf{C}(0)$. As already illustrated in Figure 1, the matrix $\mathbf{C}(0)$ itself appears to be excellently approximated by using only 10 columns, but we are interested in the approximation of the generalized eigenfunctions and eigenvalues in eq 3. In order to present a critical convergence test, we compute the relaxation time scales $t_i = \kappa_i^{-1}$, which exhibit an exponential magnification of approximation errors in the eigenvalues and are therefore very sensitive observables. In Figure 4 we report the error in the top four generalized eigenvalues and time scales, respectively, while Figure 3b shows the convergence of the estimated time scales as a function of the number of columns selected in the approximation. In Figures 3c–f we compare the first four eigenfunctions approximated with spectral oASIS using only 20 out of 1000 columns (red) and the full solution (violet). Both time scales and eigenfunctions are indistinguishable from the full solution when only 2% of the columns are used, indicating the possibility for massive computational savings without significant accuracy loss.

3.2. Variational Approach for BPTI. The small protein bovine pancreatic trypsin inhibitor (BPTI) is a natural inhibitor of the serine protease Trypsin. A 1 ms trajectory generated on the Anton supercomputer[49] has previously been subject to a number of Markov model analyses.[54,58] As another proof of concept, we apply our sparse sampling methods to a set of basis functions

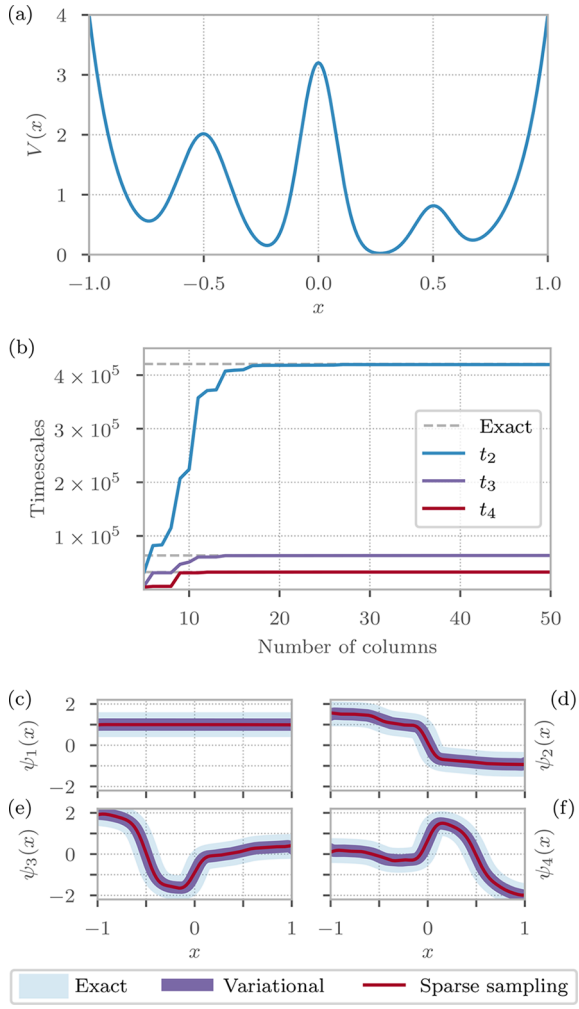


Figure 3. Sparse approximation of the dominant eigenvalues and eigenfunctions of 1-dimensional diffusion dynamics in a four-well potential. (a) Potential energy. (b) Approximated relaxation time scales as a function of the number of selected columns of the overlap matrix $\mathbf{C}(0)$ sampled with spectral oASIS. (c–f) Exact (light blue) and approximated eigenfunctions using the full variational approach (violet) and the sparse approximation (red) from 20 columns.

$$\chi_i(x_1, \ldots, x_N) = \frac{1 - (x_i/r)^{64}}{1 - (x_i/r)^{96}}, \qquad i = 1, \ldots, N$$

evaluated on $N = 1540$ coordinates x_1, ..., x_N given by the distances between pairs of Cα atoms that are at most four slots apart in the molecular chain. The values of the basis functions change smoothly from one to zero around $x_i = r$, where $r = 0.7$ nm is a cutoff parameter.

We used sparse sampling with three different column selection strategies and evaluated the quality of the column sampling procedures by computing the relaxation time scales, as described above, and comparing to the relaxation time scales of the full solution shown in Figure 5a. We compare the following strategies: oASIS (i.e., selection of a single column at a time, requiring many passes over the data), spectral oASIS (selecting $m = 20$ columns at a time, greatly reducing the required number of passes over the data), and random selection (Figure 5b–e). Each experiment was performed as a function of the number of columns selected and was repeated 50 times in order to compute mean and average errors. Random column

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

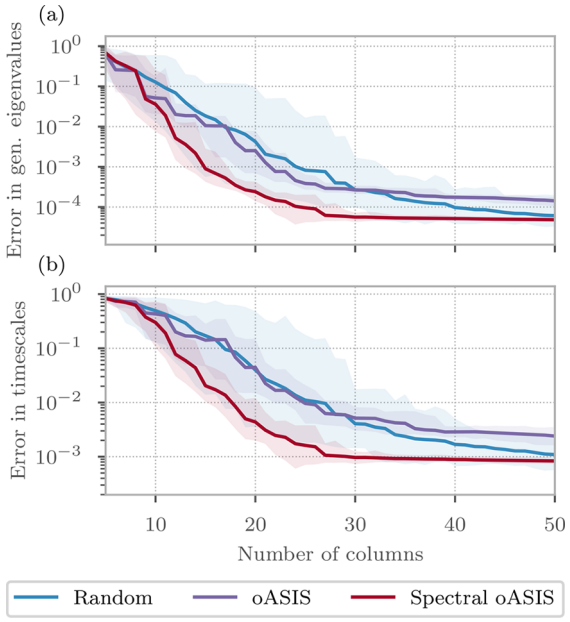


Figure 4. Comparison of the random, oASIS, and spectral oASIS selection methods when selecting three columns at a time for the system shown in Figure 3. (a) Errors (2-norm) of the four largest generalized eigenvalues. (b) Errors (2-norm) of the four largest implied time scales.

selection generally exhibits very large relative errors even when a comparatively large number of columns is used. The adaptive oASIS methods outperform random sampling by several orders of magnitude. Both oASIS and spectral oASIS reach an error level that can be considered as numerical noise using much less than 20% of all available columns (even when considering the exponential magnification of errors from eigenvalues to time scales). This means that oASIS and spectral oASIS achieve accurate results while using only a fraction of the data, resulting in a substantial speedup.

3.3. Slow Coordinates of the Trypsin−Benzamidine Complex. Trypsin is a serine protease that can be reversibly inhibited by benzamidine, which competes with trypsin's natural substrates. Trypsin has been the subject of many computational studies in the past.[50,51,59−61] It was found that trypsin has multiple long-lived conformations that exchange on the time scale of microseconds. The Markov state model previously used in Plattner and Noé[51] was based on a relatively coarse TICA analysis which employed the pairwise distances between groups of two subsequent residues as input coordinates. The reason for this choice was that the 223 residues involved in the simulated system would lead to a full set of 24753 pairwise distances, and thus dense correlation matrices with 24753^2 elements. The computational effort of estimating such matrices over approximately 10^6 frames and the storage of the resulting matrices are unpractical. Equipped with the present sparse sampling methods we can now solve the full system.

Here we use 100 μs of MD simulations in trajectories of 1 or 2 μs length that were generated in Plattner and Noé[51] in order to explore the conformational dynamics of the trypsin−benzamidine complex. Heavy-atom distances were calculated for each pair of residues separated by two or more residues, and additionally the distance of benzamidine to each of the residues in the trypsin molecule was computed, for a total of 24533

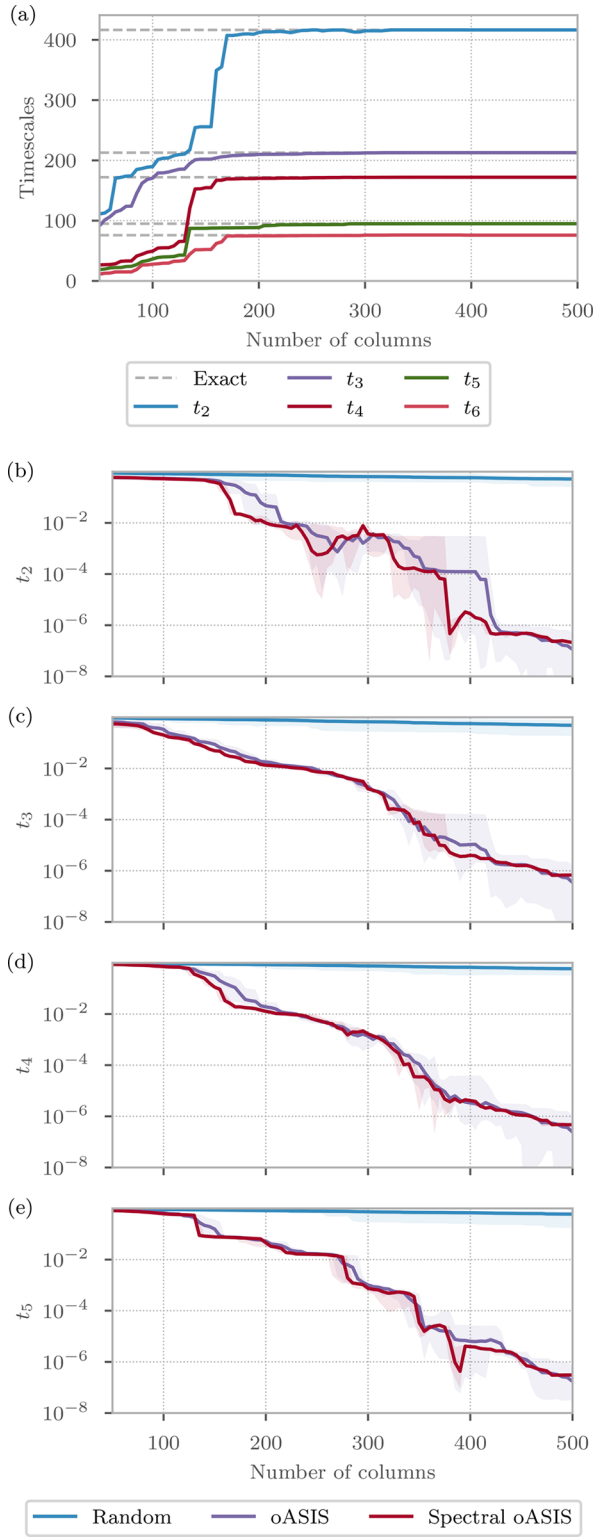


Figure 5. (a) Dominant implied time scales associated with the dynamics of BPTI as approximated by the sparse sampling method using spectral oASIS (selecting 20 columns at a time) as well as reference values computed from the full basis set (dashed). (b−e) Relative error of the slowest time scales as a function of the number of columns used in the Nyström approximation. Different methods (oASIS, violet lines, and spectral oASIS, red lines) are compared with the results obtained by choosing the columns at random (blue lines).

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

distances. The inverses of these distances were used as input features for the TICA analysis with a lagtime τ = 140 ns in the calculation of the time-lagged correlation matrix $\mathbf{C}(\tau)$.

Figures 6a–b show the sparse sampling results for TICA on the full 24533-dimensional basis set using spectral oASIS (with m = 50 columns selected at a time). The fully converged TICA time scales are around 1.29, 1.13, and 0.85 μs. Note that these time scales cannot be compared directly with those obtained previously,[51] as in that study a Markov state model analysis was performed where all the system configurations were presorted into bound, associated, and unbound. Since the purpose of the present study is the demonstration of the sparse computation and not the analysis of the binding kinetics of this complex, we used TICA without presorting the configurations. Using spectral oASIS, only about 20% of the basis functions were needed to estimate the TICA time scales and eigenvectors accurately. This results in massive savings in computer time and storage, making the use of the full TICA basis set feasible.

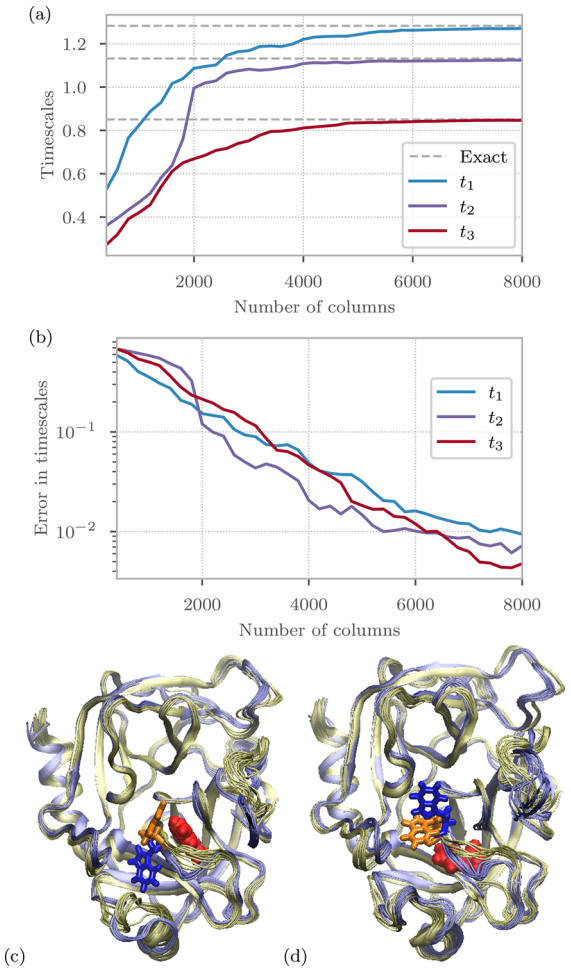


Figure 6. Application of spectral oASIS to TICA of trypsin–benzamidine. (a) Convergence of the estimation of the first three TICA time scales (in μs) as a function of the number of columns used. The full matrix has 24533 columns. About 20% of the total number of columns are needed to obtain a reliable estimate. (b) Relative error of the estimation of the three slowest time scales associated with the spectral oASIS method. (c-d) Comparison of structures at opposite extremes of the first (c) and second (d) TIC projections. The first TIC corresponds to the flipping of Trp215 and fluctuations of the loop 215–221 and its flanking loops. The second TIC corresponds to fluctuations in the calcium binding loop.[51]

In agreement with the result in the literature,[51] we find benzamidine bound in either of the two binding pockets 1 and 1*.[50,59−61] In both pockets benzamidin forms a salt bridge to Asp170 but in each pocket benzamidine can bind either above or below the binding loop. Figures 6c–d show that the first two TICA eigenvectors clearly indicate an open/close conformational change of the Trp215 side-chain that acts as a lid in front of the binding pocket. This finding is in agreement with the previous Markov state model analysis[51] where the opening and closing of the binding pockets were associated with the slowest transitions.

4. CONCLUSIONS

We used recent results from the field of compressed sensing to significantly reduce the computational effort in the analysis of molecular kinetics from molecular dynamics simulations. In particular, starting from the adaptive oASIS approach for eigenvalue approximation we propose a spectral adaptive method that samples m columns of a dense matrix simultaneously to approximate the first few eigenvalue/eigenvector pairs, additionally reducing the number of passes over the molecular dynamics trajectory in the data analysis. In addition, we propose the use of the Nyström approximation as a selection method for a subset of basis functions in the variational approach to conformation dynamics, hence extending the oASIS approach to approximate a generalized eigenvalue problem.

The power of this approach is illustrated on a number of examples, including the analysis of massive molecular dynamics data such as the binding of the ligand benzamidine to the protein trypsin. We show that in all cases considered the spectral oASIS method for the generalized eigenvalue problem achieves a dimension reduction between 20% and 1%. As the problem scales cubically with the number of basis functions used, such a reduction corresponds to a gain in efficiency of 2 to 4 orders of magnitude without significant accuracy loss. We believe that the proposed approach can facilitate the analysis of complex molecular dynamics data for large systems over long time scales, a problem that is becoming increasingly prominent as the data become easier to generate.

APPENDIX A: NYSTRÖM APPROXIMATION OF THE OVERLAP MATRIX

Let $\tilde{\mathbf{C}}$ denote the Nyström approximation of the correlation matrix $\mathbf{C}(0)$ using k columns, and let S be the set containing their indices. Without loss of generality, we may assume that $S = \{1, ..., k\}$. Moreover, let $\mathcal{P}_k$ be the orthogonal projection onto the space spanned by the k functions $\{\chi_i : i \in S\}$. Suppose that $\mathbf{C}(0) = \mathbf{X}^T\mathbf{X}$, so that $\tilde{\mathbf{C}} = \mathbf{X}^T\mathbf{P}_k\mathbf{X}$, where $\mathbf{P}_k = \mathbf{X}_k\mathbf{X}_k^+$ is the orthogonal projection onto the space spanned by the selected data points $\mathbf{X}_k = \mathbf{X}[:,S]$.[57] Similarly, the projection $\mathcal{P}_k$ in function space is given by

$$\mathcal{P}_k f = \sum_{l=1}^{k} \alpha_l(f)\chi_l$$

where $\alpha \in \mathbb{R}^k$ solves the linear system

$$\mathbf{C}^k\alpha(f) = \beta(f)$$

with $\beta \in \mathbb{R}^k$ defined by

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

$$\beta_j(f) = \langle f, \chi_j \rangle_\pi, \qquad j = 1, \ldots, k$$

By the linear independence of the basis functions χ_i, the exact correlation matrix $\mathbf{C}^k$ is an invertible matrix, so that the unique solution to the linear system is given by

$$\alpha(f) = (\mathbf{C}^k)^{-1}\beta(f)$$

Denoting the s-th row of $\mathbf{X}_k$ by $(\mathbf{X}_k)_{(s)}$, i.e., $(\mathbf{X}_k)_{(s)} = \mathbf{X}[s,S] = (\chi_1(\mathbf{x}_s), \ldots, \chi_k(\mathbf{x}_s))$, where $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s, \ldots, \mathbf{x}_T\}$ is a realization of the trajectory, we thus obtain

$$(\mathcal{P}_k\chi_i)(\mathbf{x}_s) = \langle (\mathbf{X}_k)_{(s)}, (\mathbf{C}^k)^{-1}\beta(\chi_i) \rangle$$

On the other hand, writing the i-th column of $\mathbf{X}$ as $\mathbf{X}^{(i)}$=$\mathbf{X}[:,i]$, for the orthogonal projection $\mathbf{P}_k$ we get

$$(\mathbf{P}_k\mathbf{X}^{(i)})_s = \langle (\mathbf{X}_k)_{(s)}, (\overline{\mathbf{C}^k})^+\gamma(\chi_i) \rangle$$

where $\overline{\mathbf{C}^k}$ explicitly denotes the estimate of $\mathbf{C}^k$ from data as in eq 4, and $\gamma(\chi_i)$ is the estimate of $\beta(\chi_i)$ from data:

$$\beta_j(\chi_i) \approx \gamma_j(\chi_i) = \langle \chi_j(\mathbf{x}_t) \cdot \chi_i(\mathbf{x}_t) \rangle_t$$

Consequently, it can be shown that for long enough sampling trajectories the difference between the estimated correlation matrix of the smaller basis set and the Nyström approximation of the estimated correlation matrix of the full basis set can be made arbitrarily small. The remaining error between the exact correlation matrices of the full and the projected basis set, respectively, thus only stems from the error caused by the Nyström approximation, which is a quantity we can control using the column selection algorithm.

APPENDIX B: ALTERNATIVE SPARSE SAMPLING METHODS FOR THE GENERALIZED EIGENVALUE PROBLEM

Nyström Approximation of the Time-Lagged Correlation Matrix

While in principle the time-lagged correlation matrix $\mathbf{C}(\tau)$ is neither symmetric nor positive semidefinite, in practice one might assume that it is still possible to construct a Nyström approximation that is reasonably close to it. Indeed, this seems to be the case when the time lag employed is small compared to the length of the sample trajectory. Using the same set of columns as previously selected for the purpose of the Nyström approximation of the overlap matrix $\mathbf{C}(0)$, we write both approximations as

$$\tilde{\mathbf{C}}(0) = \mathbf{C}_k(0)\mathbf{W}_k(0)^{-1}\mathbf{C}_k(0)^{\mathrm{T}}$$

$$\tilde{\mathbf{C}}(\tau) = \mathbf{C}_k(\tau)\mathbf{W}_k(\tau)^{-1}\mathbf{C}_k(\tau)^{\mathrm{T}}$$

where $\mathbf{C}_k(\cdot)$ denotes the submatrix of selected columns and $\mathbf{W}_k(\cdot)$ is the intersection matrix of rows and columns with the same indices. We proceed by computing a square root $\mathbf{L}_0 \in \mathbb{R}^{k\times k}$ of $\mathbf{W}_k(0)$ that is numerically full-rank and form the product $\mathbf{L} = \mathbf{C}_k(0)\mathbf{L}_0 \in \mathbb{R}^{M\times k}$. Consequently,

$$\mathbf{C}(0) \approx \tilde{\mathbf{C}}(0) = \mathbf{L}\mathbf{L}^{\mathrm{T}}$$

Using the pseudoinverse $\mathbf{L}^+$ of $\mathbf{L}$ as a transformation of the data, that is,

$$\hat{\mathbf{X}} = \mathbf{X}(\mathbf{L}^+)^{\mathrm{T}}$$

we get

$$\hat{\mathbf{C}}(0) = \mathbf{L}^+\mathbf{C}(0)(\mathbf{L}^+)^{\mathrm{T}} = \mathbf{I}_k \in \mathbb{R}^{k\times k}$$

$$\hat{\mathbf{C}}(\tau) = \mathbf{L}^+\mathbf{C}(\tau)(\mathbf{L}^+)^{\mathrm{T}} \in \mathbb{R}^{k\times k}$$

As a result, the generalized eigenvalue problem (eq 3) reduces to the standard eigenvalue problem

$$\mathbf{L}^+\tilde{\mathbf{C}}(\tau)(\mathbf{L}^+)^{\mathrm{T}}\mathbf{b}_i = \lambda_i\mathbf{b}_i$$

so that the generalized eigenvectors $\mathbf{a}_i$ in (eq 3) satisfy

$$\mathbf{b}_i = \mathbf{L}^{\mathrm{T}}\mathbf{a}_i$$

Note that we can compute the product

$$\mathbf{L}^+\tilde{\mathbf{C}}(\tau)(\mathbf{L}^+)^{\mathrm{T}} = [\mathbf{L}^+\mathbf{C}_k(\tau)][(\mathbf{W}_k(\tau)^{-1}\mathbf{C}_k(\tau)^{\mathrm{T}})(\mathbf{L}^+)^{\mathrm{T}}]$$

without the need for storage of full-size matrices.

CUR Decomposition

In order to relax the assumption that the Nyström approximation of the time-lagged correlation matrix $\mathbf{C}(\tau)$ be accurate even in the absence of symmetry and positive definiteness, we can replace the Nyström approximation by a more general matrix decomposition. Specifically, the *CUR decomposition*[62−65] is applicable to any matrix. Using a selected set of rows $\mathbf{R}_k$ and columns $\mathbf{C}_k$ of a matrix $\mathbf{C}$, the CUR decomposition of $\mathbf{C}$ is given by

$$\mathbf{C} = \mathbf{C}_k\mathbf{U}_k\mathbf{R}_k$$

where $\mathbf{U}_k$ is usually defined as the pseudoinverse of the intersection matrix of columns and rows. Given a selection of columns, corresponding row indices that provide for maximum accuracy within the constraint of the initially selected column indices can be found using the *maximal volume* algorithm.[66−68] Proceeding as in the previous paragraph, we thus obtain

$$\mathbf{L}^+\tilde{\mathbf{C}}(\tau)(\mathbf{L}^+)^{\mathrm{T}} = [\mathbf{L}^+\mathbf{C}_k(\tau)][(\mathbf{U}_k(\tau)\mathbf{R}_k(\tau))(\mathbf{L}^+)^{\mathrm{T}}]$$

Projected Generalized Eigenvalue Problem

Suppose that the matrices $\mathbf{C}(0)$, $\mathbf{C}(\tau)$ are given by

$$\mathbf{C}(0) = \mathbf{X}^{\mathrm{T}}\mathbf{X}$$

$$\mathbf{C}(\tau) = \mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

respectively. We then define the associated matrices

$$\mathbf{G}(0) = \mathbf{X}\mathbf{X}^{\mathrm{T}}$$

$$\mathbf{G}(\tau) = \mathbf{Y}\mathbf{X}^{\mathrm{T}}$$

Consider the generalized eigenvalue problem

$$\mathbf{G}(\tau)\mathbf{v}_i = \lambda_i\mathbf{G}(0)\mathbf{v}_i \tag{7}$$

Denoting a submatrix of k columns of $\mathbf{X}$ by $\mathbf{X}_{:k}$, we may approximate the generalized eigenpairs $(\mathbf{v}_i,\lambda_i)$ by the solutions $(\tilde{\mathbf{u}}_i,\tilde{\lambda}_i)$ of the small-scale problem[69]

$$\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{G}(\tau)\mathbf{X}_{:k}\mathbf{y}_i = \tilde{\lambda}_i\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{G}(0)\mathbf{X}_{:k}\mathbf{y}_i \tag{8}$$

where $\tilde{\mathbf{u}}_i = \mathbf{X}_{:k}\mathbf{y}_i$.

There is an interesting relationship between the generalized eigenvalue problem (eq 7) and our original problem (eq 3): Every generalized eigenpair $(\lambda_i,\mathbf{v}_i)$ of eq 7 yields a generalized eigenpair $(\lambda_i,\mathbf{a}_i)$ of eq 3, as we now show.

Let $\mathbf{X} = \mathbf{Q}\mathbf{S}\mathbf{V}^{\mathrm{T}}$ be the thin singular value decomposition of $\mathbf{X}$. Then $\mathbf{G}(0) = \mathbf{Q}\mathbf{S}^2\mathbf{Q}^{\mathrm{T}}$, and multiplying eq 7 by $\mathbf{G}(0)^+$ yields

$$\mathbf{G}(0)^+\mathbf{G}(\tau)\mathbf{v}_i = \lambda_i\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{v}_i$$

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

However, it is also true that $\mathbf{G}(\tau)\mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{G}(\tau)$, so that $(\lambda_i\hat{\mathbf{v}}_i) = (\lambda_i\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{v}_i)$ is an eigenpair of the matrix $\mathbf{G}(0)^{+}\mathbf{G}(\tau)$:

$$\mathbf{G}(0)^{+}\mathbf{G}(\tau)\hat{\mathbf{v}}_i = \lambda_i\hat{\mathbf{v}}_i$$

We multiply this equation by $\mathbf{X}^{\mathrm{T}}$ to get

$$\mathbf{V}\mathbf{S}^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{Y}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{v}}_i = \lambda_i\mathbf{X}^{\mathrm{T}}\hat{\mathbf{v}}_i$$

or equivalently

$$\mathbf{C}(0)^{+}\mathbf{C}(\tau)\mathbf{X}^{\mathrm{T}}\hat{\mathbf{v}}_i = \lambda_i\mathbf{X}^{\mathrm{T}}\hat{\mathbf{v}}_i$$

Thus, $(\lambda_i,\mathbf{a}_i) = (\lambda_i\mathbf{X}^{\mathrm{T}}\hat{\mathbf{v}}_i) = (\lambda_i\mathbf{X}^{\mathrm{T}}\mathbf{v}_i)$ is an eigenpair of the matrix $\mathbf{C}(0)^{+}\mathbf{C}(\tau)$. Multiplication of the equation

$$\mathbf{C}(0)^{+}\mathbf{C}(\tau)\mathbf{a}_i = \lambda_i\mathbf{a}_i$$

by $\mathbf{C}(0)^{+}$ from the left finally results in the generalized eigenvalue problem (eq 3).

Interestingly, the matrices involved in the generalized eigenvalue problem (eq 8) can be computed without the need to store full-size matrices. Observe that

$$\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{G}(0)\mathbf{X}_{:k} = (\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{X})(\mathbf{X}^{\mathrm{T}}\mathbf{X}_{:k}) = \mathbf{C}_{k:}(0)\mathbf{C}_{k:}(0)^{\mathrm{T}}$$

$$\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{G}(\tau)\mathbf{X}_{:k} = (\mathbf{X}_{:k}^{\mathrm{T}}\mathbf{Y})(\mathbf{X}^{\mathrm{T}}\mathbf{X}_{:k}) = \mathbf{C}_{k:}(\tau)\mathbf{C}_{k:}(0)^{\mathrm{T}}$$

where $\mathbf{C}_{k:}(\cdot)$ denotes the submatrix containing the first k rows of the respective matrix. We thus have to solve the generalized eigenvalue problem

$$\mathbf{C}_{k:}(0)\mathbf{C}_{k:}(0)^{\mathrm{T}}\mathbf{y}_i = \lambda_i\mathbf{C}_{k:}(\tau)\mathbf{C}_{k:}(0)^{\mathrm{T}}\mathbf{y}_i$$

and set $\mathbf{a}_i = \mathbf{C}_{k:}(0)^{\mathrm{T}}\mathbf{y}_i$ to obtain an approximate generalized eigenpair of eq 3.

Note that this approach has two shortcomings: First, there might be eigenpairs of eq 3 that are not approximated by eigenpairs of the small-scale problem. Second, the approximation is not guaranteed to be exact even if the Nyström approximation of $\mathbf{C}(0)$ for which the column selection is performed is exact.

Comparison of Methods

In order to assess the approximation properties of these alternative methods for sparse sampling of our generalized eigenvalue problem, we compare them among each other and also to the direct method of applying the VAC to the small basis set (as presented in the manuscript), which is called the "reduced" method in what follows.

In Figure 7, we compare the relative approximation error in the slowest implied time scale in the numerical examples discussed in the manuscript. For the four-well example, only the solution obtained from the projected generalized eigenvalue problem is significantly less accurate than the other three. In the case of the BPTI data, the reduced problem provides the best result, while the methods using low-rank decompositions and the projected problem perform worse by several orders of magnitude. Since the data for the trypsin−benzamidine example consists of many short trajectories, we observe a very unstable behavior of the Nyström-based method. It is worth noting that the method using the CUR decomposition performs best in this example. However, as the condition number of this particular problem is rather high, this result should be taken with reservations.

All in all, we conclude that the direct application of the VAC to the small basis set provides a competitive performance whilst incurring the least computational effort of all methods presented. Still, we believe that further investigation and comparison of the proposed methods would prove to be interesting.

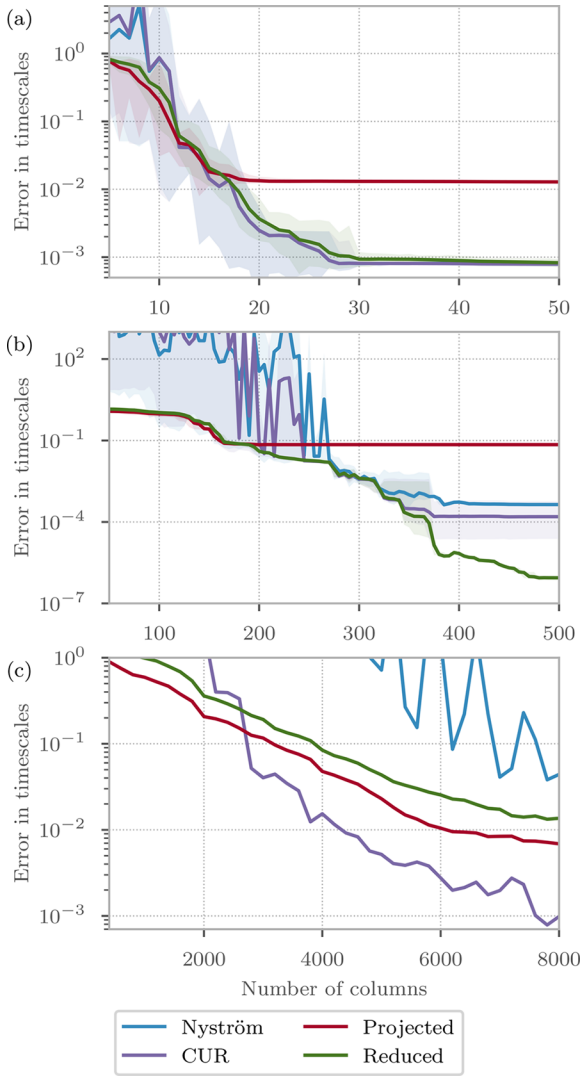


Figure 7. Comparison of the approximation behavior of the four sparse sampling methods. We show the norm of the relative error in the dominant implied time scales for (a) the four-well example, (b) BPTI, and (c) the Trypsin−Benzamidine system.

APPENDIX C: ONE-DIMENSIONAL MODEL SYSTEM

The one dimensional energy function $V(x)$ from ref 13 used in section 3.1 is given by

$$\begin{aligned} V(x) = 4(&x^8 + 0.8\exp(-80x^2) \\ &+ 0.2\exp(-80(x-0.5)^2) \\ &+ 0.5\exp(-40(x+0.5)^2)) \end{aligned}$$

A plot of the potential over the interval $[-1,1]$ is shown in Figure 3a.

We sample the energy landscape by using the discrete dynamics defined as follows. Let $(q_1, \ldots, q_N)$ be a regular grid finely approximating the x-axis ($N = 1000$ in the reported results). We define the dynamics by considering the probability of taking steps between neighbors given by a tridiagonal transition probability matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$ defined by

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$p_{i,i+1} = \nu M(V(q_i), V(q_{i+1}))$$

$$p_{i,i} = 1 - \nu M(V(q_i), V(q_{i+1})) - \nu M(V(q_i), V(q_{i-1}))$$

$$p_{i,i-1} = \nu M(V(q_i), V(q_{i-1}))$$

with the Metropolis function

$$M(a, b) = \mathbf{min}\{1, \mathbf{exp}(-\beta(b - a))\}$$

In the following, we use $\nu = 0.5$ and $\beta = 1$.

We employ the variational principle to estimate the eigenfunctions and eigenvalues associated with the Markov operator by defining a basis set of N Gaussian functions

$$\chi_i(j) = a \exp\left(-\frac{(q_i - q_j)^2}{2\sigma^2}\right)$$

each of them centered at a grid point q_i and evaluated at every grid point q_j. In the above expression, a and σ are parameters set to $a = 1$ and $\sigma = 0.15$.

The exact correlation matrix elements can be easily computed by

$$c_{ij}(\tau) = \sum_k \sum_l \chi_i(k)\pi_i(\mathbf{P}^\tau)_{ij}\chi_j(l)$$

The exact eigenfunctions ψ_k and eigenvalues $\lambda_k(\tau)$ can be calculated directly from $\mathbf{P}^\tau$.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: frank.noe@fu-berlin.de.

*E-mail: cecilia@rice.edu.

ORCID

Cecilia Clementi: 0000-0001-9221-2358

Present Address

(H.W.): School of Mathematical Sciences, Tongji University, Shanghai, 200092, P.R. China.

Funding

This work was partially supported by Deutsche Forschungsgemeinschaft grant SFB 1114/A04 (to H.W. and F.N.), European Commission ERC starting grant 307494 "pcCell" (to F.N.), National Science Foundation grants CHE-1265929, CHE-1738990, and PHY-1427654 (to C.C.), the Welch Foundation grant C-1570 (to C.C.), ARO grant W911NF-15-1-0316, AFOSR grant FA9550-14-1-0088, and ONR grant N00014-17-1-2551 (to R.B.). Calculations have been performed on the computer clusters of the Center for Research Computing at Rice University, supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award (NSF grant CNS-1338099), and on the clusters of the Department of Mathematics and Computer Science at Freie Universität Berlin.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

F.N. is grateful to the Center for Theoretical Biological Physics (CTBP, supported by NSF PHY-1427654) at Rice University for hosting his sabbatical visit, during which most of this work was performed. We are grateful to D. E. Shaw research for sharing the BPTI simulation data, and to Nuria Plattner for sharing the simulation data on the Trypsin−Benzamidine complex. We are indebted to Tom Goldstein, Eva Dyer, and Azalia Mirhoseini for their contribution to the development of oASIS.

REFERENCES

(1) Shirts, M.; Pande, V. S. Screen Savers of the World Unite! *Science* **2000**, *290*, 1903−1904.

(2) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. High-throughput all-atom molecular dynamics simulations using distributed computing. *J. Chem. Inf. Model.* **2010**, *50*, 397.

(3) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461−469.

(4) Pronk, S.; Pouya, I.; Lundborg, M.; Rotskoff, G.; Wesén, B.; Kasson, P. M.; Lindahl, E. Molecular Simulation Workflows as Parallel Algorithms: The Execution Engine of Copernicus, a Distributed High-Performance Computing Platform. *J. Chem. Theory Comput.* **2015**, *11*, 2600−2608.

(5) Doerr, S.; Harvey, M. J.; Noé, F.; Fabritiis, G. D. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, *12*, 1845−1852.

(6) Plattner, N.; Doerr, S.; Fabritiis, G. D.; Noé, F. Protein-protein association and binding mechanism resolved in atomic detail. *Nat. Chem.* **2017**, *9*, 1005−1011.

(7) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, 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. *J. Phys. Chem. B* **2004**, *108*, 6571−6581.

(9) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. *J. Chem. Phys.* **2007**, *126*, 155102.

(10) Chodera, J. D.; Dill, K. A.; Singhal, N.; Pande, V. S.; Swope, W. C.; Pitera, J. W. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.

(11) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.

(12) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* **2009**, *131*, 124101.

(13) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B. G.; 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.

(14) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. Estimation and uncertainty of reversible Markov models. *J. Chem. Phys.* **2015**, *143*, 174101.

(15) Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E3221−E3230.

(16) Wu, H.; Mey, A. S. J. S.; Rosta, E.; Noé, F. Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states. *J. Chem. Phys.* **2014**, *141*, 214106.

(17) Rosta, E.; Hummer, G. Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J. Chem. Theory Comput.* **2015**, *11*, 276−285.

(18) Chodera, J. D.; Swope, W. C.; Noé, F.; Prinz, J.-H.; Pande, V. S. Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures. *J. Chem. Phys.* **2011**, *134*, 244107.

(19) Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7426−7431.

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(20) Rohrdanz, M. A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* **2011**, *134*, 124116.
(21) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets: Deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
(22) Noé, F.; Nüske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **2013**, *11*, 635−655.
(23) Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739−1752.
(24) Wu, H.; Noé, F. Variational approach for learning Markov processes from time series data. **2017**, arXiv:1707.04659. arXiv.org ePrint archive. https://arxiv.org/abs/1707.04659.
(25) Schütte, C.; Noé, F.; Lu, J.; Sarich, M.; Vanden-Eijnden, E. Markov State Models Based on Milestoning. *J. Chem. Phys.* **2011**, *134*, 204105.
(26) Wu, H.; Noé, F. Gaussian Markov transition models of molecular kinetics. *J. Chem. Phys.* **2015**, *142*, 084104.
(27) Perez-Hernandez, 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.
(28) 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.
(29) Schwantes, C. R.; Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *J. Chem. Theory Comput.* **2015**, *11*, 600−608.
(30) Noé, F.; Clementi, C. Kinetic distance and kinetic maps from molecular dynamics simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002−5011.
(31) Noé, F.; Banisch, R.; Clementi, C. Commute maps: separating slowly-mixing molecular configurations for kinetic modeling. *J. Chem. Theory Comput.* **2016**, *12*, 5620−5630.
(32) Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **1994**, *72*, 3634−3637.
(33) Ziehe, A.; Müller, K.-R. TDSEP — an efficient algorithm for blind separation using time structure. *Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN)*, Skövde, Sweden, Sep 2−4, 1998. pp 675−680.
(34) Schmid, P. J.; Sesterhenn, J. Dynamic mode decomposition of numerical and experimental data. *Proceedings of the 61st Annual Meeting of the APS Division of Fluid Dynamics*, San Antonio, TX, Nov 23−25, 2008.
(35) Tu, J. H.; Rowley, C. W.; Luchtenburg, D. M.; Brunton, S. L.; Kutz, J. N. On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.* **2014**, *1*, 391−421.
(36) Williams, M. O.; Kevrekidis, I. G.; Rowley, C. W. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlinear Sci.* **2015**, *25*, 1307−1346.
(37) Harmeling, S.; Ziehe, A.; Kawanabe, M.; Müller, K.-R. Kernel-Based Nonlinear Blind Source Separation. *Neur. Comp.* **2003**, *15*, 1089−1124.
(38) Williams, M. O.; Rowley, C. W.; Kevrekidis, I. G. A kernel-based approach to data-driven Koopman spectral analysis. **2014**, arXiv:1411.2260. arXiv.org ePrint archive. https://arxiv.org/abs/1411.2260.
(39) Noé, F.; Clementi, C. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Curr. Opin. Struct. Biol.* **2017**, *43*, 141−147.
(40) 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*, 1−26.
(41) Candès, E. J.; Romberg, J. K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 1207−1223.
(42) Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289−1306.
(43) Baraniuk, R. Compressive Sensing. *IEEE Signal Process. Mag.* **2007**, *24*, 118−121.
(44) Williams, C.; Seeger, M. Using the Nyström method to speed up kernel machines. *Adv. Neur. Inf. Proc. Syst.* **2001**, *13*, 682−688.
(45) Drineas, P.; Mahoney, M. W. On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J. Mach. Learn. Res.* **2005**, *6*, 2153−2175.
(46) Yao, Y.; Cui, R. Z.; Bowman, G. R.; Silva, D.-A.; Sun, J.; Huang, X. Hierarchical Nyström methods for constructing Markov state models for conformational dynamics. *J. Chem. Phys.* **2013**, *138*, 174106.
(47) Harrigan, M. P.; Pande, V. S. Landmark Kernel tICA For Conformational Dynamics. 2017, bioRxiv:123752. bioRxiv.org ePrint archive. https://www.biorxiv.org/content/early/2017/04/04/123752.
(48) Patel, R. J.; Goldstein, T.; Dyer, E. L.; Mirhoseini, A.; Baraniuk, R. G. oASIS: Adaptive Column Sampling for Kernel Matrix Approximation; TREE1402, Rice University, Department of Electrical and Computer Engineering, 2015; arXiv:1505.05208. arXiv.org ePrint archive. https://arxiv.org/abs/1505.05208.
(49) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(50) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184−10189.
(51) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.
(52) Szabo, A.; Ostlund, S. S. *Modern Quantum Chemistry*; Dover Publications: Mineola, NY, 1982.
(53) Boninsegna, L.; Gobbo, G.; Noé, F.; Clementi, C. Investigating Molecular Kinetics by Variationally Optimized Diffusion Maps. *J. Chem. Theory Comput.* **2015**, *11*, 5947−5960.
(54) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Perez-Hernandez, G.; Hoffmann, M.; Plattner, N.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation and analysis of Markov models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(55) Bach, F. R.; Jordan, M. I. Predictive low-rank decomposition for kernel methods. *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, Bonn, Germany, Aug 7−11, 2005. pp 33−40.
(56) Fine, S.; Scheinberg, K. Efficient SVM training using low-rank kernel representations. *J. Mach. Learn. Res.* **2002**, *2*, 243−264.
(57) Arcolano, N. F.; Wolfe, P. J. Estimating principal components of large covariance matrices using the Nyström method. *2011 IEEE International Conference on Acoustics*; Speech and Signal Processing (ICASSP), Prague, Czech Republic, May 22−27, 2011. pp 3784−3787.
(58) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and Hidden Markov Models for calculating kinetics and metastable states of complex molecules. *J. Chem. Phys.* **2013**, *139*, 184114.
(59) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.* **2009**, *30*, 1701−1711.
(60) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein−ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386−E391.
(61) Teo, I.; Mayne, C. G.; Schulten, K.; Lelièvre, T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *J. Chem. Theory Comput.* **2016**, *12*, 2983−2989.
(62) Goreinov, S. A.; Zamarashkin, N. L.; Tyrtyshnikov, E. E. Pseudoskeleton approximations of matrices. *Dokl. Math.* **1995**, *52*, 18−19.

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(63) Goreinov, S. A.; Tyrtyshnikov, E. E.; Zamarashkin, N. L. A theory of pseudoskeleton approximations. *Linear Algebra Appl.* **1997**, *261*, 1−21.
(64) Drineas, P.; Kannan, R.; Mahoney, M. W. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. *SIAM J. Comput.* **2006**, *36*, 184−206.
(65) Mahoney, M. W.; Drineas, P. CUR matrix decompositions for improved data analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 697−702.
(66) Goreinov, S. A.; Tyrtyshnikov, E. E. The maximal-volume concept in approximation by low-rank matrices. *Contemp. Math.* **2001**, *280*, 47−51.
(67) Goreinov, S. A.; Zamarashkin, N. L.; Tyrtyshnikov, E. E. Pseudo-skeleton approximations by matrices of maximal volume. *Math. Notes* **1997**, *62*, 515−519.
(68) Goreinov, S. A.; Oseledets, I. V.; Savostyanov, D. V.; Tyrtyshnikov, E. E.; Zamarashkin, N. L. In *Matrix Methods: Theory, Algorithms, Applications*; Olshevsky, V., Tyrtyshnikov, E. E., Eds.; World Scientific: Hackensack, NJ, 2010; pp 247−256.
(69) Saad, Y. *Numerical Methods for Large Eigenvalue Problems*, revised ed.; Classics in Applied Mathematics 66; Society for Industrial and Applied Mathematics: Philadelphia, PA, 2011.

DOI: 10.1021/acs.jctc.8b00089
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX