Addressing the Embeddability Problem in Transition Rate Estimation

Curtis Goolsby,[1] James Losey,[1] Yuchen Xu,[2] Marie-Christine Düker,[2]
Mila Getmansky Sherman,[3] David S. Matteson,[2] and Mahmoud Moradi[1,*]

[1]*Department of Chemistry and Biochemistry,*
University of Arkansas, Fayetteville, AR 72701, U.S.A.
[2]*Department of Statistics and Data Science,*
Cornell University, Ithaca, NY 14850, U.S.A.
[3]*Department of Finance, Isenberg School of Management,*
University of Massachusetts at Amherst, Amherst, MA 01003, U.S.A.

(Dated: February 26, 2021)

Abstract

Markov State Models (MSM) and related techniques have gained significant traction as a tool for analyzing and guiding molecular dynamics (MD) simulations due to their ability to extract structural, thermodynamic, and kinetic information on proteins using computationally feasible MD simulations. The MSM analysis often relies on spectral decomposition of empirically generated transition matrices. Here, we discuss an alternative approach for extracting the thermodynamic and kinetic information from the so-called rate/generator matrix rather than the transition matrix. Although the rate matrix itself is built from the empirical transition matrix, it provides an alternative approach for estimating both thermodynamic and kinetic quantities, particularly in diffusive processes. We particularly discuss a fundamental issue with this approach, known as the embeddability problem and offer ways to address this issue. We describe eight different methods to overcome the embeddability problem, including a novel approach developed for this work. The algorithms were tested on data from a one-dimensional toy model to show the workings of these methods and discuss the robustness of each method in terms of its dependence in lag time and trajectory length.

* moradi@uark.edu

I. INTRODUCTION

Proteins and other biological macromolecules are associated with complex conformational spaces that are virtually impossible to be fully characterized at the atomic level using current experimental and computational tools [1]. With increasing computational capabilities due to hardware improvements, all-atom molecular dynamics simulations are used to explore the possibility of investigating the important regions of free energy landscapes of proteins and other biomolecules[2 − −4]. The tools provided by computational methods are limited by their computational costs. With this limitation, it is of significant interest to employ statistical techniques, which allow the inference of relevant thermodynamic and kinetic properties from shorter, less costly simulations. Markov State Models or MSMs[5 − −10] provide some of the most powerful tools for analyzing the ensembles of short molecular dynamics trajectories to extract information on both thermodynamics and kinetics of complex biomolecular systems and predict the behavior of such systems at much longer timescales than would be possible to simulate with current computing capabilities. These methods, however, are based on assumptions and simplifications that introduce limitations to the reliability and interpretability of these methods[11 − −13].

Building MSMs involves (1) discretization of conformational space, (2) extracting transition statistics from simulation trajectories, and (3) analyzing the transition statistics to estimate kinetic and thermodynamic properties[14]. Here, we only focus on the third component, which assumes an empirical transition matrix has been generated. Here we discuss a known[15] but less commonly used approach for analyzing empirical transition matrices that in some cases could provide an alternative approach to the more common eigendecomposition technique. The latter relies on the eigenvectors and eigenvalues of the empirical transition matrix to estimate thermodynamic and kinetic properties[16]. The approach we will discuss here, however, is based on building a rate matrix to estimate the kinetic and thermodynamic quantities using standard methods from chemical kinetics literature (Figure 1). In the finance literature, the rate matrix is known as the generator matrix[17]. Producing a rate matrix from the empirical transition matrix is known to be associated with the embeddability problem in the field of finance. The embeddability problem is found in taking the matrix logarithm of a time-dependent transition probability matrix in order to determine the generator matrix or the rate matrix, as known in chemical kinetics literature. In theory, the lag-time dependent transition probability matrix should only have positive eigenvalues; however, in practice, insufficient sampling can result in the presence of negative eigenvalues. Thus, its ma-

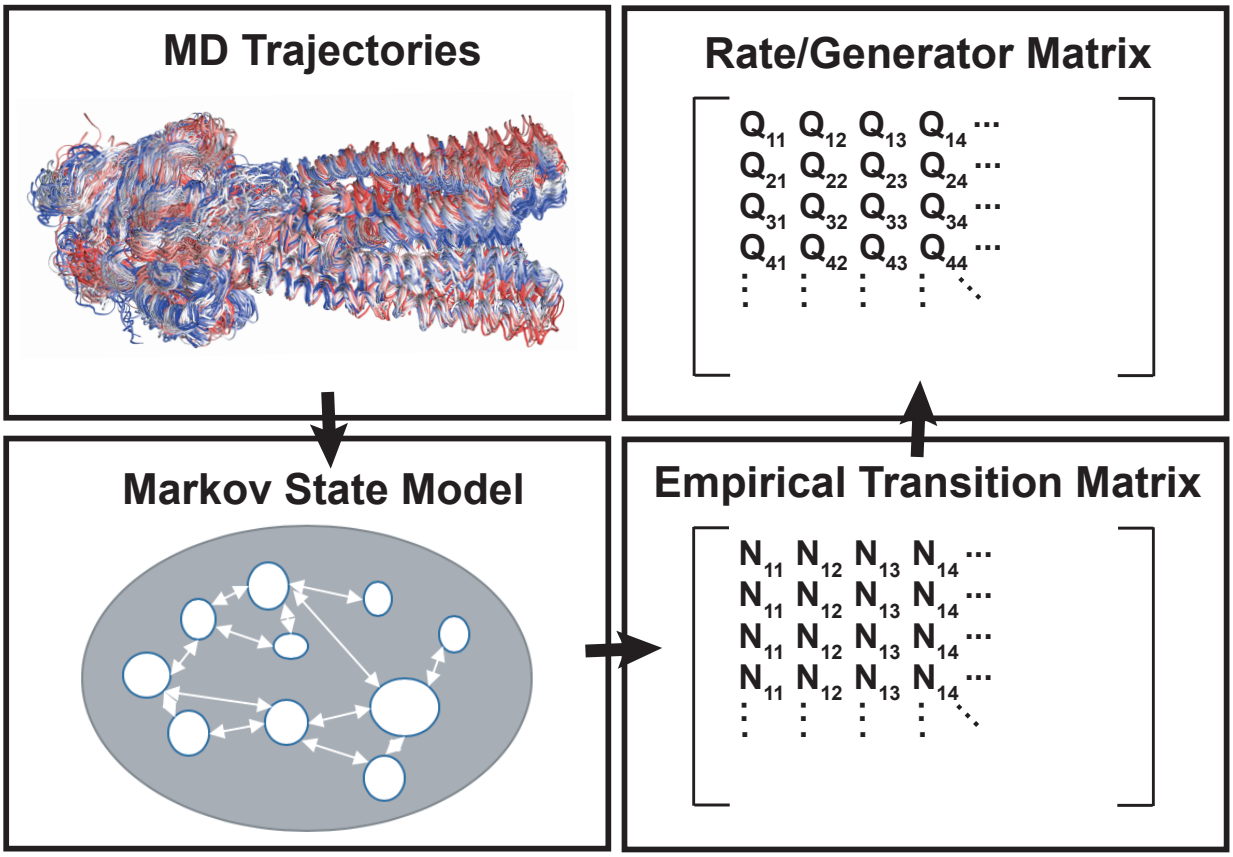

FIG. 1. The schematic representation of the proposed approach to analyze the MSM data. MD trajectories are first used within a clustering scheme to generate a Markov state model and build an empirical transition matrix using a given lag time (see MSM literature[7]). Instead of spectral decomposition of the transition matrix, however, we propose to estimate a lag-time independent rate/generator matrix from the empirical transition matrix obtained from the MD trajectories.

trix logarithm has non-real values and is an invalid generator matrix. Solving this embeddability problem involves accurately predicting the true generator matrix from an invalid generator matrix. We explored the validity of various algorithms in the chemical literature as well as from the field of finance. These methods were applied to data generated from a simple one-dimensional toy model, and compared to assess the relative performance of proposed algorithms empirically.

II. BACKGROUND

Enhanced sampling[18] has found success in the field of computational chemistry/biophysics as an effective remedy for the costs involved with simulating large molecular systems. Methods such as umbrella sampling (US)[19 – –22], metadynamics[23 – –26], and replica exchange[27 – –31]

have grown increasingly popular by aiming to enhance the sampling of configuration space by manipulating the energetics of the system. Enhancing the sampling by biasing the potential or kinetic energy, of course requires post-analysis reweighting techniques such as weighted histogram analysis method or multistate Bennett acceptance ratio method to determine the thermodynamic and kinetic quantities such as free energy and mean first passage time (MFPT). However, it is reasonable to assume that the reweighting schemes cannot fully remove the inherent bias in the MD data generated using enhanced sampling techniques. As a consequence, an alternative approach to characterize the thermodynamic and kinetic properties of biomolecular systems has been suggested that relies on Markovian analysis of transition probabilities between discrete states obtained from short but numerous unbiased MD trajectories. These methods are often known as Markov State Models or MSMs[5 − −10].

In this paper we explore the possibility of different algorithmic techniques for solving what is known as the "embeddability problem"[32, 33] in financial literature. We present a comparison between eight algorithms, used for predicting bond rating transitions, implemented in Inamura[34] and an extension by Marada[35] of Inamura's work with known and largely used methods of estimating free energies, as well as a maximum likelihood approach based on Hummer *et al.*[15]. In order to explore the efficacy of each of these methods we used a 1D bistable potential model. It is important to stress the comparison between algorithms found in finance for performing work on bond rating matrices, and our work here. The underlying mathematics is identical, and in fact there may be hidden gems in other seemingly unrelated fields for the application of this work.

Determination of kinetic properties is fundamental to the understanding of molecular systems. In order to use molecular dynamics as more than a virtual microscope, detailed calculations are necessary in order to search the conformational space for properties and conformations of interest. Various methods, such as those mentioned above, have been developed to improve the sampling in order to determine thermodynamic and kinetic properties of systems. These models are largely based upon defining a collective variable space[36, 37] in which to reduce the dimensionality of the sampling space. The methods then address how to sample more efficiently along a predefined collective variable.

MSMs[11, 38–40] allow for the reduction of the complexities of a dynamic molecular system into a lower dimensional model. The conformational space can be reduced by clustering and then brought into a square empirical transition matrix allowing for the determination of various parameters of interest such as the relative free energies and diffusion constants among others.

MSMs as with other computational methodologies still require a sufficient amount of sampling in order to obtain the parameters of interest. The present work adapts several algorithms from the field of finance in order to compare their efficacy with that of the state of the art in the chemical physics literature for addressing problems arising from insufficient sampling. In finance, MSMs are built in order to determine the probabilities of transition between rating grades for bonds[34]. Often there is insufficient data in the bond marketplace to observe transitions from every rating to every other rating in the same fashion that the conformational space of a protein is too complex in order to sample every transition with MD[34]. The discrete nature of the problem and the assumption of memoryless-ness cause the same models to work in both finance and chemical physics.

The algorithms, described below, are applied to a 1D bistable toy model. We test the efficacy of the various algorithms under different conditions, first in sufficient sampling to show that the algorithms do not distort the correct answer, and then in the case of increasing lag times as well as the case of insufficient data points. By doing this we are able to show the various strengths and weaknesses of each algorithm and hope to provide guidance for future researchers in their decision making about this matter.

Let us assume that we can assign any given sampled conformation from a set of MD simulations to one of the K well-defined states that are determined by clustering or other dimensionality reduction techniques. The empirical transition matrix, N, is defined as a $K \times K$ matrix, where $N_{i,j}$ is the number of observations of system being at state i at time t and being at state j at time $t+\tau$, in which τ is a given lag time. An estimator $\mathbf{P}$ for the transition matrix can be obtained by normalizing N.

$$P_{i,j} = \frac{N_{i,j}}{\sum_k N_{i,k}},$$

where $P_{i,j}$ is the empirical probability of observing the system at state j at time $t+\tau$, given the system is observed at state i at time t.

From the transition matrix $\mathbf{P}$, we can find the rate/generator matrix $\mathbf{Q}$. While the $\mathbf{P}$ matrix is a more common quantity to work with in Markov chain based models such as MSM, the rate matrix is more well-known in chemical kinetics literature and is the focus of our work here. $\mathbf{Q}$ is relevant for continuous time models. There are advantages in using $\mathbf{Q}$ over $\mathbf{P}$ for such models; e.g., $\mathbf{P}$ is by construct dependent on the lag time, while $\mathbf{Q}$ is not.

Transition matrix $\mathbf{P}$ is a function of τ and is related to rate/generator matrix $\mathbf{Q}$ by $\mathbf{P}(\tau) = \exp(\mathbf{Q}\tau)$. The matrix $\mathbf{Q}$ can thus be estimated from $\mathbf{P}$ within the Markovian approximation using

the relation

$$\mathbf{Q} = \frac{\log \mathbf{P}}{\tau}. \tag{2.1}$$

However, the elements of the matrix $\mathbf{Q}$ need to be real-valued and satisfy the criteria

$$\begin{cases} \sum_{j=1}^{K} Q_{i,j} = 0, & \text{for } 1 \leq i \leq K, \\ Q_{i,j} \geq 0, & \text{for } 1 \leq i, j \leq K \text{ with } i \neq j. \end{cases} \tag{2.2}$$

to be a valid generator matrix. The criteria in (2.2) lead to

$$Q_{i,i} \leq 0, \text{ for } 1 \leq i \leq K.$$

Additional criteria need to be imposed in order to guarantee the validity of detailed balance relation, which is a common feature of equilibrium processes. Put simply, at equilibrium each process should be equilibrated by its reverse process. The detailed balance or reversibility feature then allows for the definition of free energies based on the equilibrium probability of states

$$G_i = -k_B T \ \log(\pi_i),$$

where k_B is the Boltzmann constant and T is the temperature, G_i is the free energy of state i and π_i is the probability of observing the system at state i at equilibrium. The free energy between any two states i and j, $\Delta G_{i,j} = G(j) - G(i)$, can be calculated from $\mathbf{Q}$ using the detailed balance relation:

$$\Delta G_{i,j} = -k_B T \ \log\left(\frac{Q_{i,j}}{Q_{j,i}}\right). \tag{2.3}$$

Another interesting feature of relevance to molecular processes is the diffusivity, where no jump is allowed beyond the immediate neighbors of a state. The diffusivity condition is satisfied when the generator matrix $\mathbf{Q}$ is tridiagonal. For a diffusive process, the generator matrix can be used to determine the diffusion constant as[15]

$$D_{i \leftrightarrow i+1} = \sqrt{Q_{i+1,i} Q_{i,i+1}},$$

where $D_{i \leftrightarrow i+1}$ is related to the continuous diffusion constant of a diffusive process by the following approximate relationship:

$$D\left(\frac{x_i + x_{i+1}}{2}\right) \approx D_{i \leftrightarrow i+1}(x_{i+1} - x_i)^2,$$

where x_i represents the position along "reaction coordinate" associated with state i. If the states are equidistantly distributed along the reaction coordinate, we can estimate

$$D_i \approx \frac{D_{i-1\leftrightarrow i} + D_{i\leftrightarrow i+1}}{2}.$$

which leads to $D(x) = D_i \Delta x^2$, in which $\Delta x = x_i - x_{i-1}$ for all i.

The diffusion constant and free energy of states can be used to determine the mean first passage time (MFPT). This is the time of transition from the reactant (i.e, free energy minimum at x_R) to the product (i.e., free energy minimum at x_P, with an effective infinite free energy at x_0). Lifson and Jackson[41] and others[42] have shown the MFPT can be calculated using

$$MFPT_{R\to P} = \int_{x_R}^{x_P} \frac{1}{D(x)} \exp(G(x)/k_B T) \left(\int_{x_0}^{x} \exp(-G(y)/k_B T)\, dy \right) dx, \tag{2.4}$$

where the free energy and diffusion constant along x are described by $G(x)$ and $D(x)$, respectively. The MFPT can be estimated from G_i and D_i values as

$$MFPT_{R\to P} = \sum_{i=i_R}^{i_P} \frac{1}{D_i} \exp(G_i/k_B T) \sum_{j=1}^{i} \exp(-G_j/k_B T), \tag{2.5}$$

where i_P and i_R are the reactant and product states.

III. THE EMBEDDABILITY PROBLEM

For a given empirical transition matrix, it has been shown that with a sufficient but not necessary condition[43], an exact generator matrix exists such that $\mathbf{Q} = \frac{\log \mathbf{P}}{\tau}$ as in (2.1). However, the resulting $\mathbf{Q}$ may not be a valid generator matrix. For instance, consider states i and j such that transitions between the two are possible given infinite time but there is no recorded transition from i to j in the empirical transition matrix. Unfortunately, this is not an uncommon case for empirical transition matrices, due to under sampling certain transitions. Such behavior in empirical transition matrices would lead to empirical transition matrices that do not not satisfy the criteria in (2.2). One may even calculate negative eigenvalues for the empirical transition matrix, which would result in a non-real generator matrix.

In Section IV below, we present several methods to address this problem, namely either adjusting directly the generator matrix produced by the matrix logarithm of our insufficiently sampled transition matrix, performing a maximum likelihood estimation, or using a Markov chain Monte Carlo technique.

IV. ALGORITHMS

We present here a wide range of algorithms whose comparison can be found in Section VI below. These algorithms range from very simple methods such as the Diagonal Adjustment algorithm to more intricate methods such as Expectation Maximization. Besides using existing algorithms suggested in related literature for comparison, we also introduce a novel algorithm in order to estimate the generator matrix $\mathbf{Q}$. The procedure is called Polynomial Adjustment (PA) algorithm and is based on the common eigenvector structure of $\mathbf{Q}$ and the transition matrix $\mathbf{P}$.

A. Diagonal Adjustment (DA)

The Diagonal Adjustment (DA)[34] algorithm is a simple and sometimes effective solution to the embeddability problem. The algorithm adjusts the generator matrix in two steps:

$$Step\ 1.\ \widetilde{Q}_{i,j} = \begin{cases} 0, & \text{if } i \neq j \text{ and } Q_{i,j} < 0, \\ Q_{i,j}, & \text{otherwise.} \end{cases}$$

$$Step\ 2.\ \widetilde{Q}^{DA}_{i,i} = -\sum_{j=1, j\neq i}^{K} \widetilde{Q}_{i,j},$$

where $\widetilde{Q}^{DA}_{i,j}$ denotes the elements of an estimated generator matrix obtained by DA algorithm.

B. Weighted Adjustment (WA)

The Weighted Adjustment (WA)[34] is another simple algorithm very similar to the DA. The algorithm adjusts the generator matrix in two steps.

$$Step\ 1.\ \widetilde{Q}_{i,j} = \begin{cases} 0, & \text{if } i \neq j \text{ and } Q_{i,j} < 0, \\ Q_{i,j}, & \text{otherwise.} \end{cases}$$

$$Step\ 2.\ \widetilde{Q}^{WA}_{i,j} = Q_{i,j} - |\widetilde{Q}_{i,j}| \frac{\sum_{k=1}^{K} \widetilde{Q}_{i,k}}{\sum_{k=1}^{K} |\widetilde{Q}_{i,k}|},$$

where $\widetilde{Q}^{WA}_{i,j}$ denotes the elements of an estimated generator matrix obtained by WA algorithm.

C. Quasi-Optimization of the Generator (QOG)

DA and WA are very similar methodologies. Unfortunately, they are not based upon an optimization strategy and become increasingly hard to trust in sparse data situations due to their unphysicality. To this end, Krenin and Sidelnikova[44] have extended the above work by implementing a post-adjustment optimization method called quasi-optimization of the generator (QOG). QOG works by first noting that the generator has a restriction on each row which allows the problem to be split into K distinct minimization problems of the sum of the squared deviation between $\log \mathbf{P}$ and $\mathbf{Q}\tau$. Thus we can write the problem as

$$\widetilde{Q}^{QOG} = \arg\min_{\mathbf{Q}} \|\mathbf{Q}\tau - \log \mathbf{P}\|,$$

where $\|\cdot\|$ denotes the so-called Frobenius norm defined as $\|\mathbf{A}\|^2 = \sum_{i,j=1}^{K} A_{i,j}^2$ for a real-valued matrix $\mathbf{A} = (A_{i,j})_{i,j=1,\ldots,K}$. By reducing the problem to a distinct problem for each row, we can define

$$C_i = \left\{ \mathbf{z} \in \mathbb{R}^K \mid \sum_{j=1}^{K} z_j = 0,\ z_i \leq 0,\ z_j \geq 0 \text{ for } j \neq i \right\},$$

where $\mathbf{z}$ represents possible valid values for the i'th row of $\mathbf{Q}$. The optimum $\mathbf{z}$ can be determined from the i'th row of $\log(\mathbf{P})$ (denoted by $\mathbf{a}$) as

$$\arg\min_{\mathbf{z} \in C_i} \sum_{j=1}^{K} (a_j \tau - z_j)^2.$$

An easy implementation of this algorithm can be found in the Jupyter notebook contained in the supplementary information.

D. Component-Wise Optimization (CWO)

Component-Wise Optimization (CWO)[35] is a somewhat more complex version of DA or WA which bears resemblance to QOG. The general idea is to divide the problem into K-1 separate optimization problems. First, a DA or WA is performed followed by optimizing each individual value in the generator according to

$$\widetilde{Q}_{i,j}^{CWO} = \arg\min_{Q_{i,j} \in [0,c]} \|\exp(\mathbf{Q}(Q_{i,j})\tau) - \mathbf{P}\|.$$

This is followed by rebalancing the row to maintain a valid generator matrix. Constant c determines the desired convergence; the smaller the c, the faster the convergence defined as 0.0001 in

our work. Marada goes on to note that the CWO is not capable of distinguishing between local and global minima and as such should be used judiciously, possibly as a way to further optimize results from other algorithms[35].

E. Expectation Maximization (EM)

The Expectation Maximization (EM)[34] algorithm is based on iterating two steps, the Expectation-step and the Maximization-step. To be more precise, recall that $N_{i,j}$ denotes the number of transitions from state i to j. We further write R_i for the number of time steps the system stays in state i. Then, write

$$\mathbb{E}[R_i(\tau)] = \frac{1}{M}\sum_{\alpha} \frac{1}{d_\alpha}\, e_{i_\alpha}^{\top} \int_0^{\tau} \exp(\mathbf{Q}\, s)\, (e_i\, e_i^{\top})\, \exp(\mathbf{Q}\, (\tau - s))\, ds\, e_{j_\alpha}, \tag{4.1}$$

$$\mathbb{E}[N_{i,j}(\tau)] = \frac{1}{M}\sum_{\alpha} \frac{1}{d_\alpha}\, e_{i_\alpha}^{\top}\, Q_{i,j} \int_0^{\tau} \exp(\mathbf{Q}\, s)\, (e_i\, e_i^{\top})\, \exp(\mathbf{Q}\, (\tau - s))\, ds\, e_{j_\alpha}, \tag{4.2}$$

where α is an index pointing to a specific transition from state i_α (observed at time t_α) to state j_α (observed at time $t_\alpha + \tau$), and M is the total number of transition observations for a given lag time τ. Furthermore, e_i denotes a unit vector with the i'th element being one and the rest of elements being zero, we write $e^{\top}$ for the transpose of e. The quantity d_α is defined by

$$d_\alpha = e_{i_\alpha}^{\top}\, \exp(-\mathbf{Q}\, \tau)\, e_{j_\alpha}.$$

The relations (4.1) and (4.2) allow to estimate the expected values of $N_{i,j}$ and R_i as a function of the generator matrix $\mathbf{Q}$. The generator matrix is then estimated as

$$\widetilde{Q}_{i,j}^{EM} = \frac{\mathbb{E}[N_{i,j}(\tau)]}{\mathbb{E}[R_i(\tau)]}. \tag{4.3}$$

The procedure of the EM is done by calculating (4.1) and (4.2) for each element, and then using (4.3) to construct a new generator matrix. The iteration proceeds until convergence completes the algorithm.

F. Maximum Likelihood Estimator (MLE)

A common optimization approach is to use maximum likelihood estimation (MLE), in this case to determine a generator matrix to maximize the likelihood of observing all of the transitions, α

(summarized in the empirical transition matrix for a given lag time τ)[15]. The relation,

$$L = \prod_{\alpha} p(j_\alpha, t_\alpha + \tau | i_\alpha, t_\alpha) = \prod_{\alpha} \big(\exp(\mathbf{Q}\tau)\big)_{i_\alpha, j_\alpha},$$

which leads to

$$\log L = \sum_{i,j} N_{i,j} \log\left((\exp(\mathbf{Q}\tau))_{i,j}\right),$$

in which the off-diagonal elements of $\mathbf{Q}$ matrix are varied but stay positive by construct and the diagonal elements are determined by

$$Q_{i,i} = -\sum_{j \neq i} Q_{i,j}.$$

Other criteria can be easily imposed on $\mathbf{Q}$ such as reversibility (to satisfy the detailed balance) and diffusivity. To satisfy both reversibility and diffusivity, for instance, only $Q_{i,i\pm1}$ elements are varied and the other elements are determined by

$$\widetilde{Q}^{MLE}_{i,j} = \begin{cases} 0, & \text{if } |i-j| > 1, \\ Q_{i,j}, & \text{if } |i-j| = 1, \\ -Q_{i,i-1} - Q_{i,i+1}, & \text{if } i = j. \end{cases}$$

G. Quadratic Programming (QP)

A quadratic programming approach was proposed by D.T. Crommelin and E. Vanden-Eijnden[45] to take into account the eigen-structures of the transition probability matrix $\mathbf{P}$ and the generator matrix $\mathbf{Q}$. Indeed, if $\mathbf{P}$ has the eigendecomposition

$$\mathbf{P} = \sum_{i=1}^{K} \Lambda_i \phi_i \psi_i,$$

where $\mathbf{P}\phi_i = \Lambda_i \phi_i,\ \psi_i \mathbf{P} = \Lambda_i \psi_i,\ \Lambda_i \in \mathbb{C},\ \forall i = 1, \ldots, K$, then with $\lambda_i = \tau^{-1}\log(\Lambda_i)$,

$$\tau^{-1}\log(\mathbf{P}) = \sum_{i=1}^{K} \lambda_i \phi_i \psi_i. \tag{4.4}$$

Note that $(\phi_i, \psi_i, \Lambda_i, \lambda_i)$, for $1 \leq i \leq K$, can be complex-valued.

Since the generator matrix $\mathbf{Q}$ has a similar eigen-structure as $\tau^{-1}\log(\mathbf{P})$ given in (4.4), an estimator $\widetilde{\mathbf{Q}}$ can be obtained by solving the following optimization problem

$$\widetilde{\mathbf{Q}}^{QP} = \arg\min_{\mathbf{Q} \in \mathbb{D}} \sum_{i=1}^{K} \left(\alpha_i \|\mathbf{Q}\phi_i - \lambda_i \phi_i\|^2 + \beta_i \|\psi_i \mathbf{Q} - \lambda_i \psi_i\|^2 + \gamma_i |\psi_i \mathbf{Q}\phi_i - \lambda_i|^2\right), \tag{4.5}$$

where $\alpha_i, \beta_i, \gamma_i$ for $i = 1, \ldots, K$ are positive weights chosen to stabilize the algorithm numerically, and domain $\mathbb{D}$ is a subspace of $\mathbb{R}^{K\times K}$ defined by (2.2). The norm-operators and the absolute values in (4.5) are compatible with complex cases. Note that the objective function in (4.5) is a quadratic function of the entries $\mathbf{Q}_{i,j}$ for all $1 \le i, j \le K$, and (2.2) only imposes linear constraints to the domain space $\mathbb{D}$, the problem (4.5) can thus be well handled by quadratic programming[46].

H. Polynomial Adjustment (PA)

We propose here a novel routine of polynomial adjustment (PA) that not only meets the generator matrix constraints in (2.2) but also strictly maintains the eigenvectors structure.

The generator matrix $\mathbf{Q}$ can be expressed as a polynomial of the transition probability matrix $\mathbf{P}$ with order at most $K-1$, or equivalently, the vectorization of the generator matrix can be given by a linear combination of the column vectors of the matrix

$$S = (\mathrm{vec}(I_K), \mathrm{vec}(\mathbf{P}), \ldots, \mathrm{vec}(\mathbf{P}^{K-1})) \in \mathbb{R}^{K^2\times K}$$

with $\mathrm{vec}(\cdot)$ the operator that stacks all the columns of a matrix, and I_K the $K \times K$ identity matrix. It is then possible to parametrize the adjustment $\widetilde{\mathbf{Q}}$ as

$$\mathrm{vec}(\widetilde{\mathbf{Q}}^{PA}) = \widetilde{S}x,\ x \in \mathbb{R}^r \tag{4.6}$$

with $\widetilde{S} = (\lambda_1 u_1, \ldots, \lambda_r u_r)$ the column space basis of S. The singular value decomposition of S yields non-increasing singular values $\lambda_1, \ldots, \lambda_K$ and corresponding left singular eigenvectors $u_1, \ldots, u_K \in \mathbb{R}^{K^2}$, while r is the index that satisfies $\lambda_r > \varepsilon \ge \lambda_{r+1}$ (set $\lambda_{K+1} := 0$) for a given threshold $\varepsilon \ge 0$ for numeric stability. With $\mathbf{Q}$ based on the power series expression of matrix logarithm of $\mathbf{P}$, the minimizer $\hat{x}$ in the optimization problem

$$\hat{x} = \underset{x\in\mathbb{R}^r}{\arg\min} \|\widetilde{S}x - \mathrm{vec}(\mathbf{Q})\|^2 \quad \text{subject to} \quad \begin{cases} (\mathbf{1}_K \otimes e_1)\widetilde{S}x = 0, \\ \Big(I_K \otimes I_K - 2\,\mathrm{diag}\big(\mathrm{vec}(I_K)\big)\Big)\widetilde{S}x \ge 0 \end{cases} \tag{4.7}$$

is available from quadratic programming with all linear constraints[46]. Here $\otimes$ is the Kronecker product, $\mathbf{1}_K = (1, \ldots, 1) \in \mathbb{R}^K$, $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^K$, $\mathrm{diag}(\cdot)$ generates a diagonal matrix, and the last inequality holds entry-wise. The adjusted generator matrix $\widetilde{\mathbf{Q}}$ can be obtained by reshaping the vector $\widetilde{S}\hat{x}$ according to (4.6).

Though QP and PA both use the property of shared eigen-structures and incorporate the quadratic programming toolkit, the two methods still differ in many ways. First, the QP algorithm explicitly computes the possibly complex eigendecompositions of **P** and **Q**, while PA avoids doing it by using a different idea of parametrization. Second, the QP algorithm is supposed to find **Q** with eigenvectors that are sufficiently close to those of **P**, while PA forces the eigenvectors of **P** and **Q** to be strictly the same. In addition, due to the different parametrizations, PA also reduces the dimension of the search domain to $O(K)$, compared to $O(K^2)$ in QP.

V. TOY MODEL

In order to examine the workings of different generator matrix estimators, we use a 1D bistable toy model, whose dynamics is modeled by an overdamped Langevin equation using the Euler-Maruyama[47] method with parameters: temperature $T = 298K$, $mass = 1$, collision frequency $\gamma = 1$, and time step $\delta t = 10^{-6}$, where time and position are unitless. The potential was of the form

$$U(x) = \frac{k}{4}(x^2 - 1)^2, \tag{5.1}$$

where $k = 10\frac{kcal}{mol}$. For sampling, 48 sets of simulations were performed. The starting points were equidistantly distributed between $x = -2$ and 2 with increments of $\Delta x = 4/47$. For each of the 48 starting points we ran 100 repeats of simulations, each for 1,000,000,000 steps. After performing the simulations we created empirical transitions matrices for various lag times, which were then used as the seed information for our algorithms and subsequent detailed balance free energy calculations.

In order to test the efficacy of each algorithm we decided to vary lag time and trajectory lengths. Trajectory length was varied by cutting off our counting statistics for each copy at the appropriate point. For instance, a 10% trajectory length (denoted by $L = 10\%$) includes the first tenth of each trajectory. For lag time, we utilized a sliding window approach in creating our empirical transition matrix. In this approach, a transition from state i at any time t to state j at time $t + \tau$ would count towards $N_{i,j}$. The total number of transition observations based on a given trajectory would be $\tau/\delta t$ data points less than the total number of data points in the trajectory. The sliding window approach guarantees a minimal dependence on the lag time for the number of observations to avoid bias.

From the results of our algorithms we calculated the percent error from our expected values,

particularly we expected a total ΔG of 0 $\frac{kcal}{mol}$ between the two energy minima at $x = \pm 1$, and a ΔG of 2.5 $\frac{kcal}{mol}$ between the energy barrier at $x = 0$ and the average of the two energy minima at $x = \pm 1$.

VI. RESULTS AND DISCUSSION

A. Bistable Potential

We used the data generated for the toy model above to estimate the generator matrix based on various lengths of trajectories and with various lag times using eight algorithms discussed above including DA, WA, QOG, CWO, EM, MLE, QP, and PA. Before we begin it is worth noting the approximate relative computational costs of the algorithms. The DA, and WA algorithms were very low cost, with QOG and CWO only slightly more expensive. EM and MLE were very expensive, requiring specialized computing resources to run for long periods of time. Finally, the QP and PA algorithms were relatively low cost, similar to DA and WA but somewhat more expensive.

The robustness of these methods in terms of their performance was tested by varying lag time and trajectory length. First, we used a fixed trajectory length (in this case, $L = 10\%$) to examine the dependence on the lag time. Figure 2A-C shows the estimated free energies along with the analytical potential energy for the toy model. At low τ, all algorithms made reasonable predictions, at least in terms of their free energy estimates, with deviations from the model potential function occurring when $|x| > 1.5$. At longer lag times, CWO was the first algorithm to break down, which can be seen visibly in the free energy profiles based on $\tau = 50\delta t$ and greater. The longer lag times increased the CWO free energy estimate for the local minima at $x = 1$. At the longest lag time in panel C, the QP algorithm began to greatly underestimate the extremes of the free energy landscape. Longer lag times tended to make QP less reliable as an estimator for values of $|x| > 1.5$.

Panels D-F in Figure 2 more clearly demonstrate the observed τ-dependent behavior of barrier height (defined as $G(0) - \frac{G(1)+G(-1)}{2}$) and well symmetry (defined as $|G(1) - G(-1)|$) as important features of the free energy profile. All models gave reasonable barrier height estimates at low τ with slight over-estimations, WA being the highest. DA, WA, CWO and QOG estimates all decreased with increasing *tau*, where CWO had the fastest rate of decrease, culminating in the worst barrier height estimate of all the algorithms. The EM, MLE, QP, and PA models had

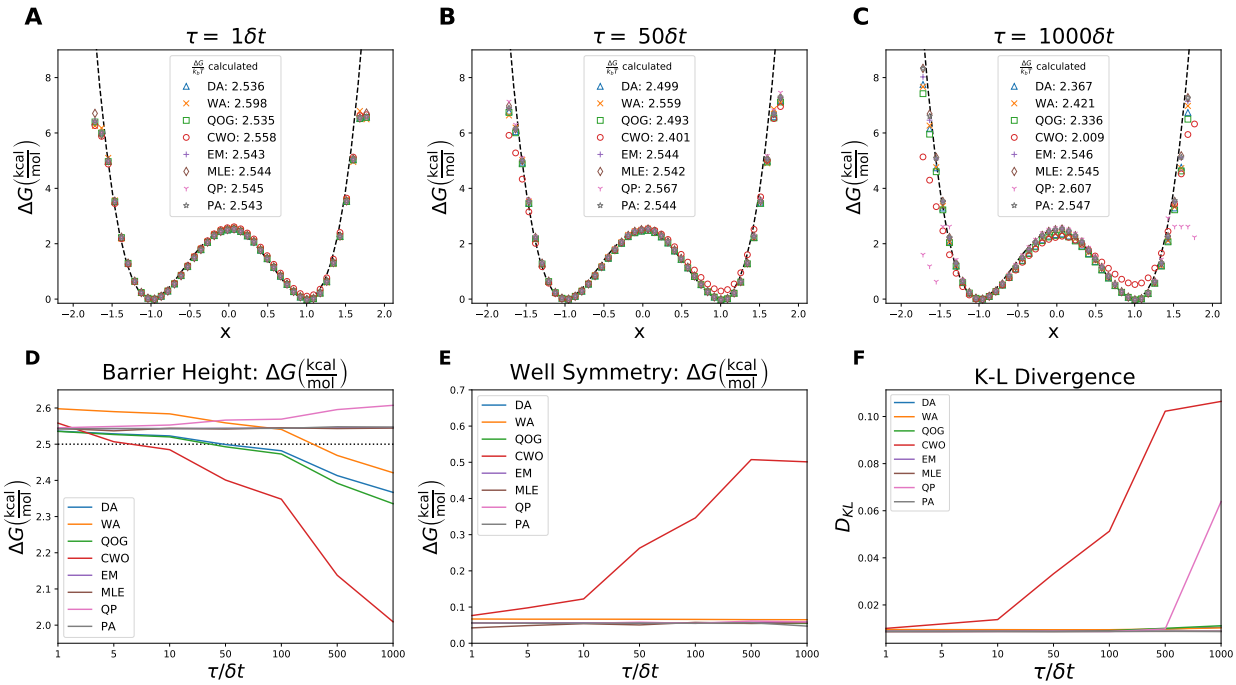

FIG. 2. (A-C) Estimated free energy along x from DA (blue triangle), WA (orange x), QOG (green square), CWO (red circle), EM (purple +), MLE (brown diamond), QP (pink Y), and PA (gray $\star$) algorithms for various lag times (expressed in terms of the number of time steps) based on 10% of the data ($L = 10\%$); (A) $\tau = \delta t$, (B) $\tau = 50\delta t$, (C) $\tau = 1000\delta t$. The reference black dashed line is the model potential: $2.5(x^2-1)^2$. (D) The estimated energy barrier height as a function of τ, with an analytical value of 2.5 shown as a black dotted line. (E) The absolute difference of the energy minima as a function of τ with an analytical value of 0. (F) The D_{KL} of the estimated positional probability from the analytic positional probability as a function of τ.

slight overestimates of barrier height that remained consistent for all τ values, though the QP estimates increased above these other algorithms at the longest lag times. Well symmetry results in Figure 2E, were consistently low for all models except CWO which increased to a plateau with increasing τ. The qualitative observation of poor performance in well symmetry for the CWO algorithm was confirmed with Kullback-Leibler divergence[48] (D_{KL}), shown in Figure 2F. D_{KL} is a measure of difference between probability distributions. The positional probability inside the bistable potential, $U(x)$, was calculated as

$$P(x) = \frac{e^{-\beta U(x)}}{\sum e^{-\beta U(x)}} \tag{6.1}$$

for the analytic potential energy and the estimated potential energy from the different optimization

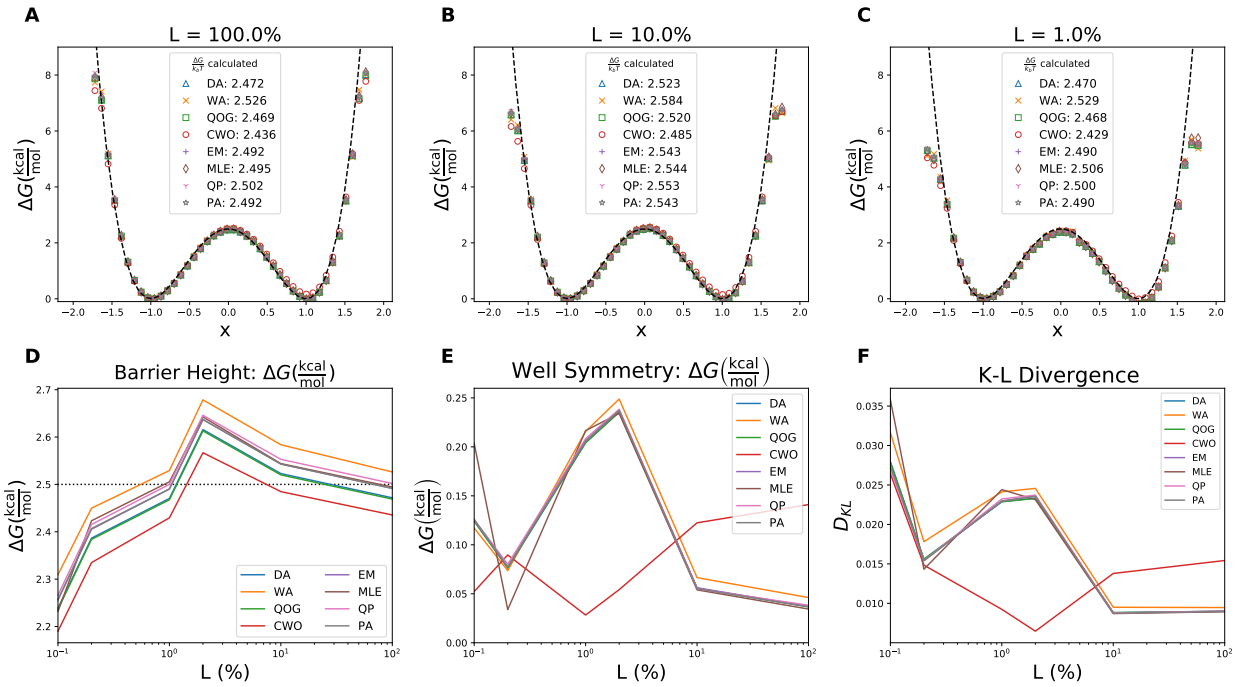

FIG. 3. (A-C) Estimated free energy along x from DA (blue triangle), WA (orange x), QOG (green square), CWO (red circle), EM (purple +), MLE (brown diamond), QP (pink Y), and PA (gray $\star$) for various trajectory lengths; (A) $L = 100\%$, (B) $L = 10\%$, (C) $L = 0.1\%$. (D) The energy barrier height ΔG (as a function of trajectory length) with an analytical value of 2.5 $\frac{kcal}{mol}$, shown as a black dotted line for reference. (E) The absolute difference of the energy minima as a function of $L\%$ with an analytical value of 0. (F) D_{KL} of the estimated positional probability from the analytic positional probability as a function of trajectory length.

algorithms. The divergence of the estimated D_{KL} was calculated by the relation

$$D_{KL} = \sum P_{Ana} \log \left(\frac{P_{Ana}}{P_{Est}} \right), \tag{6.2}$$

where P_{Ana} was the analytic probability and P_{Est} was the estimated probability. D_{KL} for other values of τ are shown in the other panels of Figure S5. The well symmetry and D_{KL} were qualitatively similar, even though the well symmetry was only using information on the basins. One exception was the deviations in free energy estimates by the QP algorithm at long lag times that are captured in the D_{KL} figure but not the well symmetry.

Figure 3A-C illustrates the free energy profiles estimated from various methods for a fixed τ ($\tau = 10\delta t$) and varying trajectory length. Using all of the trajectory, all the algorithms provided reasonable estimates for the free energy. Deviations in the basin of the well at $x = 1$ appeared as the length of the trajectory was reduced, as well as poor estimations at the extremes of the

potential, $|x| > 1.5$.

All algorithms had qualitative similarities in their estimates of barrier height in Figure 3D, with low estimates for the shortest trajectory increasing the a peak at $L = 5\%$, then approaching the true value at $L = 100$. The relative spacing of the barrier height remained mostly consistent for a given trajectory length. The most accurate algorithms for the full trajectory were EM, MLE, QP, and PA. WA overestimated the barrier height, though this behavior was dependent on the τ chosen, while DA, QOG, and CWO algorithms all underestimated the barrier height. At the lowest trajectory length, $L = 0.1\%$, the estimations of barrier peak was inaccurate, all significantly below the analytic value of 2.5. Differences between the related QP and PA algorithms were also observed at lower trajectory lengths, with QP giving slightly higher estimates than PA.

When we used shorter trajectories, the barrier estimates deviated from the correct values. From Figure 3D, the estimated barriers were flatter for the shortest trajectories, $L < 1\%$. This can be easily understood from observing that with shorter trajectory lengths, there was more dependence on the starting point of trajectories, which were distributed evenly over the space. As the trajectory length shortens, the free energy profiles resemble more the initial flat distribution. A similar overestimation of the barrier at $L = 2\%$ for all algorithms indicated further the dependence on the data for accurate results.

The well symmetry and D_{KL} panels in Figure 3E and F confirmed earlier observations that CWO behaved differently than the other algorithms. The D_{KL} values in Figure 3F confirmed that CWO performed poorly at the longer trajectory lengths, but at $L < 2\%$ the CWO deviations was comparable or superior to the other algorithms. D_{KL} for other trajectory lengths are shown in the other panels of Figure S5.

Tables S1 and S2 provide more extensive data on the dependence of these algorithms on lag time and trajectory length as they summarize various estimates based on varying both quantities. In the tables one sees at the first row for each method, the errors associated with the barrier height and well symmetry estimates are listed in Tables S1 and S2, respectively for full-length trajectories ($L = 100\%$) but varying lag times. In all cases, the barrier height errors were quite small at $\tau = \delta t$ but the well symmetry error was considerably higher than the rest for CWO (by a factor of 17) at $L = 100\%$ and $\tau = \delta t$. At the $L = 0.01\%$ and $\tau = 1000\delta t$, CWO estimates of well symmetry improved relative to the other algorithms. The well symmetry errors were generally negligible in most cases except for CWO methods. However, barrier height estimates worsen in most cases as the trajectory length decreases or as lag time increases. At $L = 0.1\%$, all algorithms perform

similarly in barrier height estimates.

The above analysis implies that most algorithms above (except for CWO) were robust for estimating thermodynamic quantities (here the free energy difference between the minima) than kinetic ones (here the barrier height, which is related to activation energy in chemical kinetics literature). We also examined the performance of these methods in terms of estimating MFPT associated with the transition from the energy minimum at $x = -1$ to the energy minimum at $x = 1$ (i.e., from the left to the right well). The analytical MFPT is 34.3. This value is estimated from Relation (2.4), using $G(x) = U(x)$ (from Relations (5.1) and $D = k_B T = 0.59$. To estimate the MFPT from the simulations, we employ Relations (2.5), which requires an estimate for the free energies (G_i) as well as D_i. Table S3 provides an estimate for $D(x)$ (averaged over x) from generator matrices that are estimated by DA, WA, QOG, CWO, EM, MLE, QP, and PA. Generally, the estimates were influenced more by τ than the trajectory lengths. Unlike the free energy estimates that tend to be more accurate for lower τ values, the diffusion constant estimates were more accurate for longer lag times. The DA and QOG algorithms generated the closest estimates to the analytical diffusion, using long lag times, though they still overestimate by approximately 80%. The QP and PA algorithms were the least accurate estimators at the long lag times, but all algorithms gave results within the same order of magnitude.

MFPT estimates based on D_i and G_i estimates are shown in Table I for varying lag times and trajectory lengths. As compared to the analytical value (34.4), all estimates were somewhat lower than expected, which is the direct result of the overestimation of $D(x)$. Similar to $D(x)$ and contrary to $G(x)$, the estimates were more accurate at longer lag times. The MFPT estimates were also affected by the trajectory length. For $L = 100\%$ and $L = 10\%$, the difference was not significant between the MFPT estimates; however, at $L = 1\%$ most methods start to show more deviation from the actual MFPT values as compared to $L = 10\%$ and $L = 100\%$ estimates. The algorithms with the closest MFPT estimates were DA, WA, and QOG, but even the closest values were still half of the analytic MFPT. Similar to diffusion, QP and PA had the least accurate MFTP estimates, but they were still within a reasonable range compared to the other algorithms.

The generator matrix contains both thermodynamic and kinetic information as the above analysis clearly illustrates. The eight different generator matrix estimators that we used here had their own pros and cons. The CWO method was a relatively poor barrier height estimator as compared to the other methods used here, especially at long lag times. On the other hand, it's well symmetry estimation surpassed other algorithms when the trajectory length was short. Furthermore, CWO

TABLE I. MFPT predicted using various estimates of the generator matrix including: DA, WA, QOG, CWO, EM, MLE, QP, and PA. Increasing values of τ are seen along the columns while decreasing trajectory length (L) are represented along the rows. The analytical MFPT is 34.3.

τ	1	10	100	1000	L (%)
DA	0.55	1.71	6.09	17.01	100
	0.56	1.77	6.28	17.41	10
	0.47	1.46	5.23	15.07	1
	0.33	1.02	3.94	11.22	0.1
WA	0.62	1.91	6.36	16.33	100
	0.64	1.97	6.56	16.77	10
	0.53	1.62	5.42	14.34	1
	0.36	1.13	4.01	10.71	0.1
CWO	0.62	1.84	6.60	15.22	100
	0.64	1.90	6.82	15.57	10
	0.52	1.56	5.69	13.90	1
	0.36	1.10	4.39	10.89	0.1
QOG	0.54	1.70	6.01	16.11	100
	0.56	1.76	6.19	16.53	10
	0.47	1.46	5.17	13.22	1
	0.33	1.02	3.90	10.01	0.1

τ	1	10	100	1000	L (%)
EM	0.55	1.71	5.99	15.40	100
	0.56	1.76	6.19	15.93	10
	0.47	1.46	5.13	13.49	1
	0.33	1.01	3.82	9.81	0.1
MLE	0.54	1.71	5.28	14.29	100
	0.57	1.77	5.64	14.82	10
	0.48	1.46	4.48	11.82	1
	0.34	0.92	2.96	8.56	0.1
QP	0.55	1.71	5.23	13.35	100
	0.56	1.77	5.40	14.18	10
	0.47	1.46	4.33	9.98	1
	0.33	1.01	3.05	11.17	0.1
PA	0.55	1.74	5.53	13.00	100
	0.57	1.80	5.75	13.35	10
	0.47	1.49	4.64	10.22	1
	0.33	1.03	3.25	6.83	0.1

performed similarly to other methods for estimating diffusion constant and MFTP. The EM, MLE, and PA algorithms was generally robust for estimating both thermodynamic and kinetic properties, with accurate predictions of barrier height for all lag times using the full trajectory. Additionally, the diffusion and MFPT estimates for EM, MLE, and PA were in general agreement with the other algorithms. It is important to remember that the PA algorithm was able to achieve these results while requiring significantly less computing time.

In these results, there were certain features that were broadly true. For instance, more data lead to better estimates; although different algorithms had different sensitivity levels to the amount of data used. An interesting observation is that free energy estimates were generally more accurate for shorter lag times but the diffusion constant estimates were more accurate for longer lag times.

The overestimation of diffusion constant (and subsequent underestimation of the MFPT) can be attributed to the correlatedness of the data generated by the overdamped langevin equation. In a typical molecular dynamics simulations, we also expect to observe such correlated behavior, which would result in similar correlatedness in the data. The longer lag times reduced correlation in the empirical transition matrix. On the other hand, longer lag times result in empirical transition matrices that have more non-zero off-diagonal elements and are farther away from the generator matrix, which in this case is expected to be tridiagonal.

VII. CONCLUSIONS

We implemented eight algorithms for overcoming the "embeddability problem" in MSMs. The efficiency of these algorithms were tested on a 1D bistable toy model. The relative performance of each algorithm were compared with regards to their ability to handle differing lag times and trajectory lengths. We found that depending on whether one was interested in free energy calculations or kinetic characterization of a process, different algorithms may behave differently. For thermodynamic characterization of a process, EM, MLE, and PA showed robust results for different lag times. PA was able to achieve comparable results in kinetic information estimation to EM and MLE despite having a significantly lower computational cost. CWO performed the worst when all the data was used, though it may perform better when trajectories are short and lag times are long.

In terms of kinetic estimations such as MFPT and diffusion constant, the simplest algorithm, DA, gave marginally closer estimates, though all algorithms performed similarly. In general, in the abundance of data, all algorithms provide reasonable results; however, the lag time dependence should be checked by repeating the analysis for various lag times to ensure the reliability of the results.

The development of the novel PA algorithm, while conceptually similar to the QP algorithm, improved upon the long lag time errors of QP in free energy estimation. PA also had similarly accurate estimates for barrier height and well symmetry to more computationally expensive algorithms like EM and MLE. This represents a viable approach to extracting embedded information for large transition matrices when computational resources are at a premium.

Overall, this exploration provides the interested reader a foundation for dealing with kinetic and thermodynamic analysis of equilibrium MD trajectories, particularly in the context of MSM.

Through our analysis we hope to shed some light on the approaches as well as the information which can be garnered from direct analysis of MD data. More data is almost always preferable but in situations where this is not possible, overcoming the embeddability problem is a possibility for estimating the thermodynamic and kinetic quantities reliably.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgement

This research is supported by the National Science Foundation under Awards 1940188, 1945465, 1934985, 1940124, 1940276, and 1940223. This research is also supported by the Arkansas High Performance Computing Center which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

---

[1] H. Berendsen and S. Hayward, "Collective protein dynamics in relation to function," Current opinion in structural biology, vol. 10, no. 2, pp. 165–169, 2000.

[2] P. Eastman, J. Swails, J. Chodera, R. McGibbon, Y. Zhao, K. Beauchamp, L.-P. Wang, A. Simmonett, M. Harrigan, C. Stern, R. Wiewiora, B. Brooks, and V. Pande, "Openmm 7: Rapid development of high performance algorithms for molecular dynamics," PLoS computational biology, vol. 13, no. 7, p. e1005659, 2017.

[3] A. Götz, M. Williamson, D. Xu, D. Poole, S. Le Grand, and R. Walker, "Routine microsecond molecular dynamics simulations with amber on gpus. 1. generalized born," Journal of Chemical Theory and Computation, vol. 8, no. 5, pp. 1542–1555, 2012.

[4] B. Towles, J. Grossman, B. Greskamp, and D. Shaw, "Unifying on-chip and inter-node switching within the Anton 2 network," ACM SIGARCH Computer Architecture News, vol. 42, no. 3, pp. 1–12, 2014.

[5] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," Journal of the Royal Statistical Society Series B-Statistical Methodology, vol. 72, no. 3, pp. 269–342, 2010.

[6] Y. Ephraim and N. Merhav, "Hidden Markov processes," IEEE Transactions on Information Theory, vol. 48, no. 6, pp. 1518–1569, 2002.

[7] V. Pande, K. Beauchamp, and G. Bowman, "Everything you wanted to know about Markov state models but were afraid to ask," Methods, vol. 52, no. 1, pp. 99–105, 2010.

[8] G. Bowman, V. Voelz, and V. Pande, "Taming the complexity of protein folding," Current Opinion in Structural Biology, vol. 21, no. 1, pp. 4–11, 2011.

[9] S. Sisson, "Transdimensional Markov chains: A decade of progress and future perspectives," Journal of the American Statistical Association, vol. 100, no. 471, pp. 1077–1089, 2005.

[10] D. Shukla, C. Hernandez, J. Weber, and V. Pande, "Markov state models provide insights into dynamic modulation of protein function," Accounts of Chemical Research, vol. 48, no. 2, pp. 414–422, 2015.

[11] G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande, "Progress and challenges in the automated construction of Markov state models for full protein systems," Journal of Chemical Physics, vol. 131, no. 12, 2009.

[12] N. Djurdjevac, M. Sarich, and C. Schütte, "Estimating the eigenvalue error of Markov state models," Multiscale Modeling & Simulation, vol. 10, no. 1, pp. 61–81, 2012.

[13] J.-H. Prinz, J. Chodera, and F. Noé, "Spectral rate theory for two-state kinetics," Physical Review X, vol. 4, no. 1, p. 011020, 2014.

[14] P. Gagniuc, Markov Chains: From Theory To Implementation And Experimentation. Wiley, 2017.

[15] G. Hummer, "Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations," New Journal of Physics, vol. 7, no. 1, p. 34, 2005.

[16] F. Noé, I. Horenko, C. Schütte, and J. Smith, "Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states," The Journal of Chemical Physics, vol. 126, no. 15, p. 155102, 2007.

[17] D. Lando and T. M. Skødeberg, "Analyzing rating transitions and rating drift with continuous observations," Journal of Banking & Finance, vol. 26, no. 2–3, pp. 423–444, 2002.

[18] R. C. Bernardi, M. C. R. Melo, and K. Schulten, "Enhanced Sampling Techniques In Molecular Dynamics Simulations Of Biological Systems," Biochimica Et Biophysica Acta-General Subjects,

vol. 1850, pp. 872–877, May 2015.

[19] S. Kumar, D. Bouzida, R. Swendsen, P. Kollman, and J. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method," Journal of Computational Chemistry, vol. 13, no. 8, pp. 1011–1021, 1992.

[20] G. Torrie and J. Valleau, "Non-Physical Sampling Distributions In Monte-Carlo Free-Energy Estimation - Umbrella Sampling," Journal of Computational Physics, vol. 23, no. 2, pp. 187–199, 1977.

[21] B. Roux, "The calculation of the potential of mean force using computer-simulations," Computer Physics Communications, vol. 91, pp. 275–282, Sep 1995.

[22] S. Kumar, J. Rosenberg, D. Bouzida, R. Swendsen, and P. Kollman, "Multidimensional free-energy calculations using the weighted histogram analysis method," Journal of Computational Chemistry, vol. 16, no. 11, pp. 1339–1350, 1995.

[23] A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," Physical Review Letters, vol. 100, no. 2, 2008.

[24] A. Laio and F. L. Gervasio, "Metadynamics: A Method To Simulate Rare Events And Reconstruct The Free Energy In Biophysics, Chemistry And Material Science," Reports On Progress In Physics, vol. 71, no. 12, 2008.

[25] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," Wiley Interdisciplinary Reviews-Computational Molecular Science, vol. 1, no. 5, pp. 826–843, 2011.

[26] A. Laio, A. Rodriguez-Fortea, F. Gervasio, M. Ceccarelli, and M. Parrinello, "Assessing the accuracy of metadynamics," Journal of Physical Chemistry B, vol. 109, no. 14, pp. 6714–6721, 2005.

[27] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," Chemical Physics Letters, vol. 314, no. 1-2, pp. 141–151, 1999.

[28] K. Hukushima and K. Nemoto, "Exchange Monte Carlo method and application to spin Glass simulations," Journal of The Physical Society of Japan, vol. 65, no. 6, pp. 1604–1608, 1996.

[29] D. Earl and M. Deem, "Parallel tempering: Theory, applications, and new perspectives," Physical Chemistry Chemical Physics, vol. 7, no. 23, pp. 3910–3916, 2005.

[30] Y. Sugita, A. Kitao, and Y. Okamoto, "Multidimensional replica-exchange method for free-energy calculations," Journal of Chemical Physics, vol. 113, no. 15, pp. 6042–6051, 2000.

[31] H. Fukunishi, O. Watanabe, and S. Takada, "On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction,"

Journal of Chemical Physics, vol. 116, no. 20, pp. 9058–9067, 2002.

[32] M.-A. Guerry, "Some results on the embeddable problem for discrete-time Markov models in manpower planning," Communications in Statistics-Theory and Methods, vol. 43, no. 7, pp. 1575–1584, 2014.

[33] M.-A. Guerry, "On the embedding problem for discrete-time Markov chains," Journal of Applied Probability, vol. 50, no. 4, pp. 918–930, 2013.

[34] Y. Inamura, "Estimating continuous time transition matrices from discretely observed data," Bank Of Japan, pp. 06–07, 2006.

[35] T. Marada, "Quantitative credit risk modeling: Ratings under stochastic time," VU University Amsterdam, Master thesis, 2008.

[36] I. Kurylyak and I. Yukhnovskii, "The method of collective variables in the equilibrium statistical-theory of bounded systems of charged-particles .1. continuum model of an electrolyte solution occupying a half-space," Theoretical and Mathematical Physics, vol. 52, no. 1, pp. 691–699, 1982.

[37] I. Gimondi, G. A. Tribello, and M. Salvalaglio, "Building maps in collective variable space," Journal of Chemical Physics, vol. 149, no. 10, 2018.

[38] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, "Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics," Journal of Chemical Physics, vol. 126, no. 15, 2007.

[39] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schuette, and F. Noe, "Markov models of molecular kinetics: Generation and validation," Journal of Chemical Physics, vol. 134, no. 17, 2011.

[40] K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande, "Msmbuilder2: Modeling Conformational Dynamics On The Picosecond To Millisecond Scale," Journal of Chemical Theory and Computation, vol. 7, pp. 3412–3419, Oct 2011.

[41] S. Lifson and J. L. Jackson, "On the self-diffusion of ions in a polyelectrolyte solution," The Journal of Chemical Physics, vol. 36, no. 9, pp. 2410–2414, 1962.

[42] A. Ansari, "Mean first passage time solution of the smoluchowski equation: Application to relaxation dynamics in myoglobin," The Journal of Chemical Physics, vol. 112, no. 5, pp. 2516–2522, 2000.

[43] R. B. Israel, J. S. Rosenthal, and J. Z. Wei, "Finding generators for markov chains via empirical transition matrices, with applications to credit ratings," Mathematical finance, vol. 11, no. 2, pp. 245–265, 2001.

[44] A. Kreinin and M. Sidelnikova, "Regularization algorithms for transition matrices," Algo Research Quarterly, vol. 4, no. 1/2, pp. 23–40, 2001.

[45] D. Crommelin and E. Vanden-Eijnden, "Fitting timeseries by continuous-time Markov chains: A quadratic programming approach," Journal of Computational Physics, vol. 217, no. 2, pp. 782–805, 2006.

[46] M. Frank and P. Wolfe, "An algorithm for quadratic programming," Naval research logistics quarterly, vol. 3, no. 1–2, pp. 95–110, 1956.

[47] G. Maruyama, "Continuous Markov processes and stochastic equations," Rendiconti del Circolo Matematico di Palermo, vol. 4, no. 1, p. 48, 1955.

[48] S. Kullback and R. A. Leibler, "On information and sufficiency," Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79–86, 1951.