Published in final edited form as:

SIAM Rev Soc Ind Appl Math. 2018 ; 60(4): 909–938. doi:10.1137/16M1104329.

Trajectory stratification of stochastic dynamics

Aaron R. Dinner[1,2,*], **Jonathan C. Mattingly**[3], **Jeremy O. B. Tempkin**[1,2], **Brian Van Koten**[4], **Jonathan Weare**[1,4,†]

[1]James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA

[2]Department of Chemistry, The University of Chicago, Chicago, Illinois 60637, USA

[3]Departments of Mathematics and Statistical Science, Duke University, Durham, North Carolina 27708, USA

[4]Department of Statistics, The University of Chicago, Chicago, Illinois 60637, USA

Abstract

We present a general mathematical framework for trajectory stratification for simulating rare events. Trajectory stratification involves decomposing trajectories of the underlying process into fragments limited to restricted regions of state space (strata), computing averages over the distributions of the trajectory fragments within the strata with minimal communication between them, and combining those averages with appropriate weights to yield averages with respect to the original underlying process. Our framework reveals the full generality and flexibility of trajectory stratification, and it illuminates a common mathematical structure shared by existing algorithms for sampling rare events. We demonstrate the power of the framework by defining strata in terms of both points in time and path-dependent variables for efficiently estimating averages that were not previously tractable.

I. INTRODUCTION

Computer simulation is a powerful tool for the study of physical processes. Specifically, stochastic simulation methods have broad applicability in modeling physical systems in a variety of fields including chemistry, physics, climate science, engineering, and economics [1, 2]. In many practical applications, the statistical properties of the process of interest are approximated by averages over many independent realizations of trajectories of the process, or, in the case of ergodic properties, by averages taken over a single very long trajectory of the process. However, for many systems, the most interesting events occur infrequently and are therefore very difficult to observe by direct numerical integration of the equations governing the dynamics. For example, in chemistry, the conformational changes responsible for the function of many molecules and, in climate science, extreme events like severe droughts and violent hurricanes, occur on timescales orders of magnitude longer than the timestep for numerical integration. This basic observation has motivated the development of numerous techniques aimed at enhancing the sampling of rare events of interest without

* dinner@uchicago.edu . † weare@uchicago.edu .

sacrificing statistical fidelity (see [3] for an account within the context of molecular simulation).

In this article, we depart from standard enhanced sampling approaches and develop a general mathematical and computational framework for the estimation of statistical averages involving rare trajectories of stochastic processes. Our approach can be viewed as a form of stratified sampling, long a cornerstone of experimental design in statistics (e.g., [4]). In stratified sampling, a population is divided into subgroups (strata), averages within those strata are computed separately, and then averages over the entire state space are assembled as weighted sums of the strata averages. Stratification also has a long history in computer simulations of condensed-phase systems as umbrella sampling (US) [3, 5–8]. The key idea behind any stratified sampling strategy is that, when the strata are chosen appropriately, their statistics can be obtained accurately with relatively low effort and combined to estimate the average of interest with (much) less overall effort than directly sampling the stochastic process to the same statistical precision. Here we show that the trajectories of an arbitrary discrete-time Markov process (including many dynamics with memory, so long as they can be written as a suitable mapping) can also be stratified: they can be decomposed into fragments restricted to regions of trajectory space (strata), averages over the distributions of trajectory fragments within the strata can be computed with limited communication between them, and those averages can be combined in a weighted fashion to yield a very broad range of statistics that characterize the dynamics.

These basic features are at the core of the existing nonequilibrium umbrella sampling (NEUS) method [9–11], which forms the starting point for our development. NEUS was originally introduced to estimate stationary averages with respect to a given, possibly irreversible, stochastic process [9]. Starting in [10, 11] it was observed that the general NEUS approach was applicable to certain dynamic averages as well. The basic NEUS approach has been been applied and further developed in subsequent articles [12–15] and in the Exact Milestoning scheme [16], which was derived from the Milestoning method [17] but is very similar in structure to NEUS. At its most basic level, NEUS relies on duplication of states in rarely visited regions of space and subsequent forward evolution of the duplicated states. In this way it is similar to a long list of so-called “trajectory splitting” techniques [18–26] that are also able to compute averages of dynamic quantities. Like NEUS, splitting techniques also often involve a decomposition of state space into regions. Unlike NEUS however, in most splitting techniques bias is removed through the use of a separate weight factor for each individual sample (rather than for an entire region), and the computational effort expended in each region is not controlled directly. What makes the NEUS method unique among splitting techniques is that it is also a trajectory stratification strategy.

Our goal in this article is to provide a clear and general mathematical framework for trajectory stratification that builds upon the NEUS method. In the process we clearly delineate the range of statistics that can be estimated by NEUS, including more general quantities than previously computed. Our analysis of the underlying mathematical structure of US [27, 28] has already facilitated the derivation of a central limit theorem for US and a detailed understanding of its error properties. Here, our framework reveals unanticipated

connections between the equilibrium and nonequilibrium US methods and places the nonequilibrium algorithm within the well-studied family of stochastic approximation methods [29]. The analysis leads to a practical scheme that departs dramatically from currently available alternatives. We demonstrate the use of trajectory stratification to compute a hitting time distribution as well as to compute the expectation of a path-dependent functional that gives the relative normalization constants for two arbitrary, user-specified un-normalized probability densities.

II. A UNIFIED FRAMEWORK

In this section we present a framework that reveals the unified structure underlying umbrella sampling in both the equilibrium and nonequilibrium case. In Section II A, we review the equilibrium approach [27, 28] to introduce terminology and the central eigenproblem in a context where the analogies to traditional umbrella sampling descriptions [3, 5–8] are readily apparent. In Section II B, we present the nonequilibrium version of the algorithm and show how this interpretation results in a flexible scheme for computing dynamic averages. As for its equilibrium counterpart, an eigenproblem lies at the core of the nonequilibrium method. This eigenproblem however, involves a matrix that depends on the desired eigenvector, introducing the need for a self-consistent iteration. In Section III, we give a precise description of the fixed-point problem solved by this iteration and show that the algorithm is an example of a stochastic approximation strategy [29]. In Section IV we specialize our development to the context of steady-state averages that motivated the original development of NEUS [9].

A. Averages with Respect to a Specified Density

Our presentation in this section follows [27]. We view umbrella sampling as a method to compute averages of the form

$$\int_{x \in \mathbb{R}^d} f(x)\pi(dx), \tag{1}$$

where π is a known probability distribution and d is the dimension of the underlying system (e.g., the total number of position coordinates for all atoms in a molecular system). For example, π might be the canonical distribution, $\pi(dx) \propto e^{-\beta V(x)}dx$ where V is a potential energy function, β is an inverse temperature, and f might be 1 on some set A and 0 elsewhere. In this case, $-\beta^{-1} \log \int f(x)\pi(dx)$ can be regarded as the free energy of the set A.

Note that in our notation π is a probability measure on $\mathbb{R}^d$ and dx is an infinitesimal volume element in $\mathbb{R}^d$. If the distribution π has a density function $p(x)$ then $\pi(A) = \int_{x \in A} p(x)dx$ and, in particular, $\pi(dx) = p(x)dx$. This more general notation is useful when we move to our description of the nonequilibrium umbrella sampling scheme. As an aid to the reader, we choose to introduce it in the simpler setting of this section.

Consistent with traditional implementations of US [3, 6], we divide the computation of the average in (1) into a series of averages over local subsets of space. More precisely, instead of directly computing averages with respect to π, we compute averages with respect to π

probability distributions, $\boldsymbol{\pi}_j$, each of which concentrates probability in a restricted region of space (relative to $\boldsymbol{\pi}$ itself) with the goal of eliminating or reducing barriers to efficient sampling associated with $\boldsymbol{\pi}$. So that general averages with respect to $\boldsymbol{\pi}$ can be assembled, the $\boldsymbol{\pi}_j$ satisfy $\pi = \sum_{j=1}^{n} z_j \pi_j$ for a set of weights z_j to be defined in a moment.

To obtain the restricted distributions $\boldsymbol{\pi}_j$ we can set

$$\pi_j(dx) = \frac{\psi_j(x)\pi(dx)}{\int_{y \in \mathbb{R}^d} \psi_j(y)\pi(dy)}, \tag{2}$$

where the ψ_j are non-negative user defined functions satisfying $\sum_{j=1}^{n} \psi_j(x) = 1$ for all x (this last requirement is relaxed in [27]). For example, one might choose $\psi_j(x) = 1_{A_j}(x) / \sum_{\ell=1}^{n} 1_{A_\ell}(x)$, where the A_j are a collection of sets covering the space to be sampled, and, for any set A_j, the function $1_{A_j}(x)$ is 1 if $x \in A_j$ and 0 otherwise.

Note that $\pi = \sum_{j=1}^{n} z_j \pi_j$ is satisfied with

$$z_j = \int_{x \in \mathbb{R}^d} \psi_j(x)\pi(dx) \tag{3}$$

and that the average (1) with respect to $\boldsymbol{\pi}$ can be reconstructed using the equation

$$\int_{x \in \mathbb{R}^d} f(x)\pi(dx) = \sum_{j=1}^{n} z_j \langle f \rangle_j, \tag{4}$$

with

$$\langle f \rangle_j = \int_{x \in \mathbb{R}^d} f(x)\pi_j(dx)\,. \tag{5}$$

Here z_j is the statistical weight associated with each distribution $\boldsymbol{\pi}_j$ and $\langle f \rangle_j$ are the averages of the observable f against $\boldsymbol{\pi}_j$. From (4) we see that if we can sample from the $\boldsymbol{\pi}_j$ and compute the z_j then we can compute averages with respect to $\boldsymbol{\pi}$. Since $\boldsymbol{\pi}_j$ is known explicitly in this case, it can be sampled by standard means (e.g., Langevin dynamics or Metropolis Monte Carlo [3]).

Our key observation underpinning the equilibrium umbrella sampling method is that the z_j themselves are functions of averages with respect to the local distributions $\boldsymbol{\pi}_j$:

$$z_j = \sum_{i=1}^{n} z_i F_{ij} \quad \text{and} \quad \sum_{j=1}^{n} z_j = 1, \tag{6}$$

where

$$F_{ij} = \int_{x \in \mathbb{R}^d} \psi_j(x) \pi_i(dx) . \tag{7}$$

The matrix F is stochastic (i.e., has non-negative entries with rows that sum to 1) and (6), which is written in matrix-vector form as

$$z^{\mathrm{T}} F = z^{\mathrm{T}} \text{ and } \sum_{j=1}^{n} z_j = 1, \tag{8}$$

is an eigenproblem that can be solved easily for the vector z.

We now have a stratification scheme for computing the target average in (1) by sampling from the distributions $\boldsymbol{\pi}_j$. Operationally, the main steps are as follows.

1. Assemble F defined in (7) (or the alternative in Appendix A below) and $\langle \hat{A} \rangle_j$ defined in (5) by sampling from $\boldsymbol{\pi}_j$ defined in (2).
2. Solve the eigenvector equation (8) for z defined in (3).
3. Compute the desired expectation via (4).

The efficiency of this *equilibrium* US scheme has been analyzed in detail elsewhere [27, 28]. Roughly, the benefit of US is due to the facts that averages with respect to the $\boldsymbol{\pi}_j$ are often sufficient to solve for all desired quantities, and one can choose ψ_j so that averages with respect to the $\boldsymbol{\pi}_j$ converge much more quickly than averages with respect to $\boldsymbol{\pi}$ itself. It is this basic philosophy that we extend in Section II B to the computation of dynamic averages.

B. Averages with Respect to a Given Markov Process

The mathematical description of the nonequilibrium umbrella sampling scheme that follows reveals how the stratification strategy developed for the equilibrium case in Section II A can be extended to compute nearly arbitrary dynamic statistics. Our interest in this section is computing averages over trajectories of some specified Markov process, $X^{(t)}$. This process can be time-inhomogenous, i.e., given the value of $X^{(t)}$, the distribution of $X^{(t+1)}$ can depend on the value of t. We compute averages of trajectories evolved up to a first exit time of the process $(t, X^{(t)})$ from a user specified set of times and positions, D—i.e., trajectories terminate when they first leave the set D. We consider averages over trajectories of $X^{(t)}$ run until time

$$\tau = \min\left\{ t > 0 : \left(t, X^{(t)}\right) \notin D \right\} \tag{9}$$

for a set $D \in \mathbb{N} \times \mathbb{R}^d$. In the first numerical example in Section V, D is a set of times and positions for which we would like to compute an escape probability. In the second numerical example, D restricts only the times over which we simulate. The averages are of the form

$$\mathbf{E}\left[\sum_{t=0}^{\tau - 1} f\left(t, X^{(t)}\right) \right]. \tag{10}$$

We note that the average in (10) is not completely general, in order to streamline the developments below. Without any modification, we can compute averages similar to (10) but with the argument $(t,X^{(t)})$ in the definitions of τ and f replaced by $(t,X^{(t-1)},X^{(t)})$. On the other hand, expectations with $(t,X^{(t)})$ replaced by $(t,X^{(t-m)},\ldots,X^{(t-1)},X^{(t)})$ for $m \geq 2$ cannot be obtained immediately. These and many more general expectations can, however, be accommodated by applying the algorithm to an enlarged process (e.g., $(t,X^{(t-m)}, \ldots, X^{(t-1)},X^{(t)})$) at the cost of storing copies of the enlarged process. For many expectations, this cost is quite manageable. Finally, we require that $\mathbf{E}[\tau] < \infty$. The limit $\tau \to \infty$ is considered in Section IV.

Below we show that expectations of time-dependent functions can be decomposed as a weighted sum of expectations computed over restricted subsets of the full space and, in turn, how the statistical weights can be computed as expectations over these subsets, mirroring the basic structure of the equilibrium scheme described in Section II A. However, as we discuss in Section III, the algorithm for computing these local expectations departs significantly from the equilibrium case because their form is not known *a priori* in the nonequilibrium setting.

1. The Index Process—The US scheme in Section II A used the basis functions ψ_j to stratify the sampling of the distribution π by decomposing averages with respect to π into averages with respect to the more easily sampled π_j. To arrive at an analogous partitioning of state space for the nonequilibrium case, we introduce an *index process* $\mathcal{J}^{(t)}$ that takes values in $\{1,2,\ldots,n\}$ and (roughly) labels the point $(t,X^{(t)})$ in time and space, $\mathbb{N} \times \mathbb{R}^d$. Our objective is to generate fragments of trajectories of $X^{(t)}$ consistent with specific values of $\mathcal{J}^{(t)}$ thereby breaking the coupled process $(X^{(t)}, \mathcal{J}^{(t)})$ into separate regions corresponding to a given value of $\mathcal{J}^{(t)}$ (see panel A of Figure 1).

The idea of discretizing a process $X^{(t)}$ according to the value of some user-specified index process is not new in computational statistical mechanics. For example, in our notation, given a partition of state space $A_1,A_2,\ldots,A_n$, the Milestoning procedure [17] and some Markov State Modeling procedures [30] correspond to an index process that marks the pairs of sets (A_i,A_j) for $i \neq j$ between which $X^{(t)}$ last transitioned. In the Milestoning method, the pairs of sets are considered unordered, so that a transition from A_j to A_i immediately following a transition from A_i to A_j does not correspond to a change in $\mathcal{J}^{(t)}$, and $\mathcal{J}^{(t)}$ can assume $n = \binom{m}{2}$ distinct values. The original presentation of NEUS on the other hand corresponds to a process $\mathcal{J}^{(t)}$ which marks the index of the set A_j containing $X^{(t)}$. For accurate results, the Milestoning procedure requires that the index process $\mathcal{J}^{(t)}$ itself be Markovian. Even under the best circumstances, that assumption is only expected to hold approximately. It is not required by the NEUS algorithm. Our presentation below reveals the full flexibility in the choice of $\mathcal{J}^{(t)}$ within NEUS. That flexibility is essential in the generalized setting of this article.

In the developments below we require that $\mathcal{J}^{(t)}$ is chosen so that the *joint process* $(X^{(t)}, \mathcal{J}^{(t)})$ is Markovian. This assumption allows that trajectories can be continued beyond a single transition event (before τ) without additional information about the history of $X^{(t)}$ or $\mathcal{J}^{(t)}$. We

do not assume that $J^{(t)}$ alone is Markovian and in general it is not. Our assumption implies no practical restriction on the underlying Markov process $X^{(t)}$. When $X^{(t)}$ is non-Markovian, additional variables can often be appended to $X^{(t)}$ to yield a new Markov process to which the developments below can be applied. A version of this idea is applied in Section V C where we append a variable representing a nonequilibrium work to an underlying Markov process.

2. The Eigenproblem—Given a specific choice of index process $J^{(t)}$, the nonequilibrium umbrella sampling algorithm stratifies trajectories of $X^{(t)}$ according to their corresponding values of $J^{(t)}$. That is, for each possible value of the index process, NEUS generates segments of trajectories of $X^{(t)}$ between the times that $J^{(t)}$ transitions to and from $J = j$. To make this idea more precise, we need to carefully describe the distribution sampled by these trajectory fragments:

$$\pi_j(t, dx) = \frac{\mathbf{P}\left[t < \tau, X^{(t)} \in dx, J^{(t)} = j\right]}{z_j}, \tag{11}$$

where

$$z_j = \sum_{t=0}^{\infty} \mathbf{P}\left[t < \tau, J^{(t)} = j\right]. \tag{12}$$

For each j, $\boldsymbol{\pi}_j$ is the distribution of time and position pairs $(t, X^{(t)})$ conditioned on $J^{(t)} = j$ and $t < \boldsymbol{\tau}$. We call the $\boldsymbol{\pi}_j$ *restricted distributions*. We have reused the notations $\boldsymbol{\pi}_j$ and z_j from our account of the equilibrium umbrella sampling scheme to emphasize the analogous roles played by those objects in both sections. Note that here we are treating time as an additional random variable. Also note that in these definitions as well as in the formulas below, **P** and **E** represent probabilities and expectations with respect to the original, unbiased $X^{(t)}$ and $J^{(t)}$. We assume that $z_j > 0$ for all j since we can remove the index j from consideration if $z_j = 0$. The z_j are all finite because $\sum_{j=1}^{n} z_j = \mathbf{E}[\tau]$, which we assume is finite.

Observe that

$$\begin{aligned}
\mathbf{E}\left[\sum_{t=0}^{\tau-1} f\left(t, X^{(t)}\right)\right] &= \sum_{t=0}^{\infty} \mathbf{E}\left[f\left(t, X^{(t)}\right), t < \tau\right] \\
&= \sum_{j=1}^{n} \sum_{t=0}^{\infty} \int_{x \in \mathbb{R}^d} f(t, x) \\
&\times \mathbf{P}\left[t < \tau, X^{(t)} \in dx, J^{(t)} = j\right] \\
&= \sum_{j=1}^{n} z_j \langle f \rangle_j,
\end{aligned} \tag{13}$$

where

$$\langle f \rangle_j = \sum_{t=0}^{\infty} \int_{x \in \mathbb{R}^d} f(t,x) \pi_j(t, dx) \,. \tag{14}$$

Thus we have a decomposition of (10) analogous to the decomposition of (1) in (4). Also as in the equilibrium case, the z_j can be computed from averages with respect to the $\boldsymbol{\pi}_j$. To see this, observe that for any t we can write

$$\begin{aligned} &\sum_{i=1}^{n} \mathbf{P}\left[t+1 < \tau, J^{(t+1)} = j, J^{(t)} = i\right] \\ &= \mathbf{P}\left[t+1 < \tau, J^{(t+1)} = j\right]. \end{aligned} \tag{15}$$

Summing this expression over t we obtain

$$\begin{aligned} &\sum_{i=1}^{n} \sum_{t=0}^{\infty} \mathbf{P}\left[t+1 < \tau, J^{(t+1)} = j, J^{(t)} = i\right] \\ &= \sum_{t=0}^{\infty} \mathbf{P}\left[t < \tau, J^{(t)} = j\right] - \mathbf{P}\left[J^{(0)} = j\right]. \end{aligned} \tag{16}$$

These expressions are all bounded by $\mathbf{E}[\boldsymbol{\tau}]$ and are therefore finite. Expression (16) can be rewritten as an affine eigenequation

$$z^{\mathrm{T}} G + a^{\mathrm{T}} = z^{\mathrm{T}}, \tag{17}$$

where z is defined in (12),

$$G_{ij} = \frac{\sum_{t=0}^{\infty} \mathbf{P}\left[t+1 < \tau, J^{(t+1)} = j, J^{(t)} = i\right]}{z_i}, \tag{18}$$

and

$$a_j = \mathbf{P}\left[J^{(0)} = j\right]. \tag{19}$$

Equation (17) is the analog of (8) in Section II A. Here, the matrix element G_{ij} stores the expected number of transitions from $J = i$ to $J = j$, normalized by the expected number of time steps with $J = i$. Note that the matrix G is substochastic; that is, it has non-negative entries and rows that sum to a number less than or equal to one.

To complete the analogy with the umbrella sampling scheme described in Section II A, we need to show that the elements of the matrix G are expressible as expectations over the $\boldsymbol{\pi}_j$. Indeed,

$$G_{ij} = \frac{1}{z_i}\int_{x\in\mathbb{R}^d}\sum_{t=0}^{\infty}\mathbf{P}_{t,x,i}\left[t+1<\tau, J^{(t+1)}=j\right]$$
$$\times\mathbf{P}\left[t<\tau, X^{(t)}\in dx, J^{(t)}=i\right] \tag{20}$$
$$=\sum_{t-0}^{\infty}\int_{x\in\mathbb{R}^d}\mathbf{P}_{t,x,i}\left[t+1<\tau, J^{(t+1)}=j\right]\pi_i(t,dx)$$

where $\mathbf{P}_{t,x,i}$ is used to denote probabilities with respect to X initialized at time and position (t,x) and conditioned on $J^{(t)} = i$ and $t < \tau$. Note that in the first line we have appealed to the Markovian assumption on $(X^{(t)}, J^{(t)})$. Had we instead assumed that $J^{(t)}$ alone was Markovian, we could have ignored the x dependence in (20).

Just as for the umbrella sampling algorithm described in Section II A, we arrive at a procedure for computing (10) via stratification:

1. Assemble G_{ij} defined in (18) and $\langle f\rangle_j$ defined in (14) by sampling from the π_j defined in (11).
2. Solve the affine eigenvector equation (17) for z defined in (12).
3. Compute

$$\mathbf{E}\left[\sum_{t=0}^{\tau-1} f\left(t, X^{(t)}\right)\right] = \sum_{j=1}^{n} z_j\langle f\rangle_j \tag{21}$$

via (13).

Relative to the scheme in Section II A, sampling the restricted distributions π_j requires a more complicated procedure. This is the subject of Section III. In Section III, instead of G, we choose to work with the matrix

$$\bar{G}_{ij} = \frac{\sum_{\ell=0}^{\infty}\mathbf{P}\left[S^{(\ell+1)}<\tau, J^{\left(S^{(\ell+1)}\right)}=j, J^{\left(S^{(\ell)}\right)}=i\right]}{\sum_{\ell=0}^{\infty}\mathbf{P}\left[J^{\left(S^{(\ell)}\right)}=j, S^{(\ell)}<\tau\right]}, \tag{22}$$

where

$$S^{(\ell)} = \min\left\{s > S^{(\ell-1)}: J^{(s)} \neq J^{\left(S^{(\ell-1)}\right)}\right\} \tag{23}$$

is the time of the ℓth change in the value of $J^{(t)}$ for a given realization of the coupled process $(X^{(t)}, J^{(t)})$. Likewise, instead of z, we choose to work with the weights

$$\bar{z}_j = \sum_{\ell=0}^{\infty} \mathbf{P}\left[J^{\left(S^{(\ell)}\right)} = j, S^{(\ell)} < \tau\right]. \tag{24}$$

We show in Appendix B that $\bar{G}$ is related to G by the identity

$$\bar{G}_{ij} = \begin{cases} G_{ij}/\left(1 - G_{ii}\right), & j \neq i \\ 0, & j = i, \end{cases} \tag{25}$$

and that $\bar{z}$ is related to z by

$$\bar{z}_j = z_j\left(1 - G_{jj}\right). \tag{26}$$

Therefore, knowledge of G implies knowledge of $\bar{G}$ and $\bar{z}$, and the algorithm detailed in the next section could also be expressed in terms of G and z at the cost of additional factors of $1–G_{jj}$ in several formulas. Moreover, identities (17), (25), and (26) imply

$$\bar{z}^{\mathrm{T}} = \bar{z}^{\mathrm{T}}\bar{G} + a^{\mathrm{T}}; \tag{27}$$

that is, $\bar{z}$ and $\bar{G}$ solve the same affine eigenproblem as z and G. We emphasize $\bar{G}$ and $\bar{z}$ over G and z only to simplify the presentation and interpretation of the algorithm in Section III.

To give an appealing intuitive interpretation of $\bar{G}$, we note that for $i \neq j$,

$$\begin{aligned} \bar{z}_i\bar{G}_{ij} &= z_i G_{ij} \\ &= \sum_{\ell=0}^{\infty} \mathbf{P}\left[S^{(\ell+1)} < \tau, J^{\left(S^{(\ell+1)}\right)} = j, J^{\left(S^{(\ell)}\right)} = i\right]. \end{aligned} \tag{28}$$

We refer to this quantity as the *net probability flux* from $J = i$ to $J = j$; it is the expected number of transitions of the process $J(t)$ from $J = i$ to $J = j$ before time τ. The matrix $\bar{G}$ stores the relative probabilities of transitions to different values of J before time τ and $\bar{z}_j$ is the expected number of transitions into $J = j$ before time τ.

Finally, we remark that rapid convergence of the scheme in practice rests upon the choice of $J(t)$. Roughly, one should choose the index process so that the variations in estimates of the required averages with respect to the π_j (e.g., estimates of the G_{jj}) are small. In practice, this requires that transitions between values of $J(t)$ are frequent, which is the analog of selecting the biases in equilibrium US to limit the range of the free energy over each subset of state space (see [27, 28]). In Section V we describe this and other important implementation details in the context of particular applications.

III. A GENERAL NEUS FIXED-POINT ITERATION

In this section we present a detailed algorithm for computing (10) by the stratification approach outlined in Section II B. To accomplish this one must be able to generate samples from the restricted distributions $\boldsymbol{\pi}_j(t, dx)$. In NEUS, the restricted distributions are sampled by introducing a set of Markov processes

$$\mathscr{Y}_j^{(r)} = \left(T_j^{(r)}, Y_j^{(r)}, I_j^{(r)}\right) \tag{29}$$

called *excursions* whose values are triples of a time $T_j^{(r)}$, a position $Y_j^{(r)}$, and a value of the index process $I_j^{(r)}$. To avoid confusion, we consistently use the variable r for the time associated with an excursion $\mathscr{Y}_j^{(r)}$ and the variable t for the time associated with the process $(t, X^{(t)}, J^{(t)})$.

Roughly speaking, each excursion is a finite segment of a trajectory of the process $(t, X^{(t)}, J^{(t)})$ with $J = j$. These segments are stopped either on reaching time τ or at the first time when $J \neq j$. To be precise, excursions are generated as follows:

1. Draw an initial time and position pair $\left(T_j^{(0)}, Y_j^{(0)}\right)$ from the distribution $\bar{\pi}_j(s, dy)$ specified below or from an estimate of that distribution. Set $\mathscr{Y}_j^{(0)} = \left(T_j^{(0)}, Y_j^{(0)}, j\right)$.
2. Set $T_j^{(r+1)} = T_j^{(r)} + 1$, and generate $\left(Y_j^{(r+1)}, I_j^{(r+1)}\right)$ from the distribution of $\left(X^{\left(T_j^{(r+1)}\right)}, J^{\left(T_j^{(r+1)}\right)}\right)$ conditioned on $X^{\left(T_j^{(r)}\right)} = Y_j^{(r)}$ and $J^{\left(T_j^{(r)}\right)} = j$.
3. Stop on reaching time τ or when $J \neq j$. That is, stop when r reaches

$$\rho_j = \min\left\{r \geq 0 : I_j^{(r)} \neq j \text{ or } \left(T_j^{(r)}, Y_j^{(r)}\right) \notin D\right\}. \tag{30}$$

The excursions $\mathscr{Y}_j^{(r)}$ are illustrated in Figure 1 for a particular choice of index process.

For the excursions $\mathscr{Y}_j^{(r)}$ to sample the restricted distribution $\boldsymbol{\pi}_j(t, dx)$, we must take the initial distribution $\bar{\pi}_j(s, dy)$ to be the distribution of times s and positions y at which the process $(t, X^{(t)}, J^{(t)})$ transitions from a state $J^{(s-1)} = i$ with $i \neq j$ to state $J^{(s)} = j$ (see Section III A and Appendix C). We call these distributions the *flux distributions*.

In general, the flux distributions $\bar{\pi}_j(s, dy)$ are not known *a priori* and must be computed approximately. In the NEUS algorithm, we begin with estimates of the flux distributions and the matrix $\bar{G}$. We then compute excursions initialized from these estimates of the flux distributions. From the excursions and the current estimate of $\bar{G}$, we compute statistics which are used to improve the estimates of both the flux distributions and $\bar{G}$. Thus, NEUS is an iteration designed to produce successively better estimates of the flux distributions and $\bar{G}$ simultaneously.

In Section III B, we derive a fixed-point equation solved by $\bar{G}$ and the flux distributions, and we motivate NEUS as a self-consistent iteration for solving this equation. In Section III C, we describe the complete NEUS algorithm in detail and interpret it as a stochastic approximation algorithm [29] for solving the fixed-point equation derived in Section III B. In the Supplementary Material, we analyze a simple four-site Markov model to clearly illustrate the structure of this self-consistent iteration and the terminology of the framework.

A. The Flux Distributions

Before deriving the fixed-point problem and the corresponding stochastic approximation algorithm, we define the flux distributions $\bar{\pi}_j(s, dy)$ precisely. We let

$$\bar{\pi}_j(s, dy) = \frac{\sum_{\ell=0}^{\infty} \mathbf{P}\left[S^{(\ell)} = s, s < \tau, X^{(s)} \in dy, J^{(s)} = j\right]}{\bar{z}_j} \tag{31}$$

be the distribution of time and position pairs $\left(S^{(\ell)}, X^{\left(S^{(\ell)}\right)}\right)$ conditioned on $J^{\left(S^{(\ell)}\right)} = j$. With this definition of $\bar{\pi}_j(s, dy)$, an excursion $\mathscr{Y}_j^{(r)}$ samples the restricted distribution $\boldsymbol{\pi}_j(t, dx)$ in the sense that

$$\begin{aligned} \pi_j(t, dx) &= \frac{\bar{z}_j}{z_j} \mathbf{P}\left[t < \rho_j + T_j^{(0)}, Y_j^{\left(t - T_j^{(0)}\right)} \in dx\right] \\ &= \frac{\bar{z}_j}{z_j} \sum_{s=0}^{t} \int_{y \in \mathbb{R}^d} \mathbf{P}_{s,y,j}\left[t < \sigma(s) \wedge \tau, X^{(t)} \in dx\right] \bar{\pi}_j(s, dy), \end{aligned} \tag{32}$$

where

$$\sigma(s) = \min\left\{r > s : J^{(r)} \neq J^{(s)}\right\} \tag{33}$$

and ρ_j is defined in (30). We prove (32) in Appendix C.

Given (32), we may express any average over $\boldsymbol{\pi}_j$ as an average over $\bar{\pi}_j$. For example,

$$\bar{G}_{ij} = \sum_{s=0}^{\infty} \int_{y \in \mathbb{R}^d} \mathbf{P}_{s,y,i}\left[J^{(\sigma(s))} = j, \sigma(s) < \tau\right] \bar{\pi}_i(s, dy). \tag{34}$$

Moreover, from (13), we can express general averages as

$$\mathbf{E}\left[\sum_{t=0}^{\tau-1} f\left(t, X^{(t)}\right)\right] = \sum_{j=1}^{n} \bar{z}_j \left\langle \bar{f} \right\rangle_j, \tag{35}$$

where

$$\begin{aligned} \left\langle \bar{f} \right\rangle_j = & \sum_{s=0}^{\infty} \int_{y \in \mathbb{R}^d} \sum_{t=s}^{\infty} \int_{x \in \mathbb{R}^d} f(t,x) \\ & \times \mathbf{P}_{s,y,j}\left[t < \sigma(s) \wedge \tau, X^{(t)} \in dx\right] \bar{\pi}_j(s, dy)\,. \end{aligned} \tag{36}$$

We use these facts in our interpretation of the NEUS algorithm in Section III B.

Instead of working directly with the flux distributions, we find it convenient to express both the fixed-point problem and the algorithm in terms of the probability distribution of time and position pairs $(t, X^{(t)})$ conditioned on observing a transition from $J = i$ to $J = j$ at time t, i.e., in terms of

$$\begin{aligned} & \gamma_{ij}(s, dy) \\ & = \frac{1}{\bar{z}_i \bar{G}_{ij}} \\ & \times \sum_{\ell=0}^{\infty} \mathbf{P}\left[s = S^{(l+1)} < \tau, J^{\left(S^{(l)}\right)} = i, J^{(s)} = j, X^{(s)} \in dy\right] \\ & = \frac{1}{\bar{G}_{ij}} \sum_{r=0}^{\infty} \int_{w \in \mathbb{R}^d} \mathbf{P}_{r,w,i} \\ & \left[s = \sigma(r), s < \tau, X^{(s)} \in dy, J^{(s)} = j\right] \times \bar{\pi}_i(r, dw) \end{aligned} \tag{37}$$

which is defined only for $s > 0$. To simplify notation, we let γ denote the set of all conditional distributions γ_{ij}. Recall from (28) that $\bar{z}_i \bar{G}_{ij}$ is the net probability flux from $J = i$ to $J = j$. The following simple but key identity relates γ to the flux distributions $\bar{\pi}_j$:

$$\bar{\pi}_j(s, dy) = \frac{1}{\bar{z}_j} \begin{cases} \sum_{i \neq j} \bar{z}_i \bar{G}_{ij} \gamma_{ij}(s, dy), & \text{if } s > 0 \\ a_j \mathbf{P}\left[X^{(0)} \in dy \mid J^{(0)} = j\right] & \text{if } s = 0. \end{cases} \tag{38}$$

The $s > 0$ term is the contribution from transitions into state $J = j$ from the neighboring state $J = i$, and the $s = 0$ term accounts for the initial $t = 0$ contribution of the underlying process when $J = j$. We emphasize that both the fixed-point problem and the iteration that we define below could be expressed in terms of the flux distributions $\bar{\pi}_j$ instead of γ. We choose to express them in terms of γ because the resulting formalism more naturally captures the implementation of the method used to generate our numerical results in Section V.

B. The Fixed-Point Problem

We now derive the fixed-point problem. Our goal is to find an expression of the form

$$(\mathcal{G}(\bar{G},\gamma),\Gamma(\bar{G},\gamma)) = (\bar{G},\gamma) \tag{39}$$

that characterizes the desired matrix $\bar{G}$ and collection of probability measures γ as the fixed-point of a pair of maps $\mathcal{G}(\tilde{G},\tilde{\gamma})$ and $\Gamma(\tilde{G},\tilde{\gamma})$ that take as arguments approximations $\tilde{G}$ of $\bar{G}$ and $\tilde{\gamma}$ of γ and return, respectively, a new substochastic matrix and a new collection of probability measures.

To this end, we define a function mapping $\tilde{G}$ and $\tilde{\gamma}$ to an approximation of the flux distribution $\bar{\pi}_j$. We denote this function by the corresponding capital letter $\bar{\Pi}_j$. Based on (27) and (38), we define

$$\begin{aligned}&\bar{\Pi}_j(s,dy;\tilde{G},\tilde{\gamma})\\ &= \frac{1}{\tilde{z}_j}\begin{cases}\sum_{i\neq j}\tilde{z}_i\tilde{G}_{ij}\tilde{\gamma}_{ij}(s,dy) & \text{if } s>0,\\ a_j\mathbf{P}\left[X^{(0)}\in dy \mid J^{(0)}=j\right] & \text{if } s=0,\end{cases}\end{aligned} \tag{40}$$

where $\tilde{z}$ solves the equation $\tilde{z}^{\mathrm{T}} = \tilde{z}^{\mathrm{T}}\tilde{G} + a^{\mathrm{T}}$. The matrices $\tilde{G}$ that we consider are strictly substochastic. We assume that $\tilde{G}$ is also irreducible, in which case the solution $\tilde{z}$ exists and is unique. To motivate the definition above, we observe that for the exact values $\bar{G}$ and γ, $\bar{\pi}_j(s,dy) = \bar{\Pi}_j(s,dy;\bar{G},\gamma)$ by (38). Moreover, given $\tilde{G}$ and samples from $\tilde{\gamma}$, one can generate samples from $\bar{\Pi}_j(s,dy;\tilde{G},\tilde{\gamma})$; see Section III C. This is crucial in developing a practical algorithm to solve the fixed-point problem.

At this point we are ready to define the functions $\mathcal{G}$ and Γ appearing in (39) above. For a substochastic matrix $\tilde{G}$ and a collection of probability distributions $\tilde{\gamma} = \{\tilde{\gamma}_{ij}\}$, define the substochastic matrix

$$\begin{aligned}&\mathcal{G}_{ij}(\tilde{G},\tilde{\gamma})\\ &= \sum_{s=0}^{\infty}\int_{y\in\mathbb{R}^d}\mathbf{P}_{s,y,i}\left[J^{(\sigma(s))}=j,\sigma(s)<\tau\right]\bar{\Pi}_i(s,dy;\tilde{G},\tilde{\gamma})\end{aligned} \tag{41}$$

and the collection of probability distributions

$$\Gamma_{ij}(s,dy;\tilde{G},\tilde{\gamma}) \propto \sum_{r=0}^{\infty}\int_{w\in\mathbb{R}^d}\mathbf{P}_{r,w,i}\left[s=\sigma(r),s<\tau,X^{(s)}\in dy,J^{(s)}=j\right]\bar{\Pi}_i(r,dw;\tilde{G},\tilde{\gamma}). \tag{42}$$

Because $\bar{\Pi}_j(\bar{G},\gamma)=\bar{\pi}_j$, expressions (34) and (37) imply that $\mathcal{G}(\bar{G},\gamma)=\bar{G}$ and $\Gamma_{ij}(\bar{G},\gamma)=\gamma_{ij}$, establishing our fixed-point relation (39).

Having fully specified the fixed-point problem, we can now consider iterative methods for its solution. One approach would be to fix some $\varepsilon\in(0,1]$ and compute the deterministic fixed-point iteration

$$\begin{aligned}\tilde{G}(m+1)&=\tilde{G}(m)+\varepsilon(\mathcal{G}(\tilde{G}(m),\tilde{\gamma}(m))-\tilde{G}(m)),\ \text{and}\\ \tilde{\gamma}(m+1)&=\tilde{\gamma}(m)+\varepsilon(\Gamma(\tilde{G}(m),\tilde{\gamma}(m))-\tilde{\gamma}(m)),\end{aligned}\tag{43}$$

given initial guesses $\tilde{G}(0)$ and $\tilde{\gamma}(0)$ for G and γ, respectively. One would typically choose $\varepsilon=1$ in this deterministic iteration; we consider arbitrary $\varepsilon\in(0,1]$ to motivate the stochastic approximation algorithm developed in Section III C.

In practice, computing $\mathcal{G}$ and Γ in the right hand side of (43) requires computing averages over trajectories of $(X^{(t)},J^{(t)})$ initiated from $\bar{\Pi}_j(\tilde{G}(m),\tilde{\gamma}(m))$. While we cannot hope to compute these integrals exactly, we can construct a stochastic algorithm approximating the iteration in (43) using a finite number of sampled trajectories. The resulting scheme, which we detail in Section III C, fits within the basic stochastic approximation framework.

C. A Stochastic Approximation

In this section, we present the full NEUS algorithm and we interpret it as a stochastic approximation algorithm analogous to the deterministic fixed-point iteration (43). In NEUS, as in the fixed-point iteration, we generate a sequence of approximations $\tilde{G}(m)$ and $\tilde{\gamma}(m)$, converging to $\bar{G}$ and γ, respectively. During the mth iteration of the NEUS algorithm, we update the current approximations $\tilde{G}(m)$ and $\tilde{\gamma}(m)$ based on statistics gathered from K independent excursions $\mathcal{Y}_j^{(r)}(m)=\left(T_j^{(r)},Y_j^{(r)},I_j^{(r)}\right)$ defined according to the rules governing $\mathcal{Y}_j^{(r)}$ enumerated above with $\left(T_j^{(0)},Y_j^{(0)}\right)$ drown from $\bar{\Pi}_j(\tilde{G}(m),\tilde{\gamma}(m))$, the current (at the mth iteration of the scheme) estimate of the flux distribution $\bar{\pi}_j$.

We now state the NEUS algorithm. To simplify the expressions below, we sometimes omit the iteration number m. The algorithm proceeds as follows:

1. Choose initial approximations $\tilde{G}(0)$ and $\tilde{\gamma}(0)$ of $\bar{G}$ and γ, respectively. Fix the number K of independent excursions $\mathcal{Y}_j^{(r)}(m)$ to compute for each restricted distribution $\pi_j(t,dx)$. Choose the maximum number of new points L included in the update to the empirical approximations of the distributions $\tilde{\gamma}_{ij}(m)$.

2. For each $j=1,2,\ldots,n$ generate K independent excursions

$$\mathcal{Y}_{ik}^{(r)}=\left(T_{ik}^{(r)},Y_{ik}^{(r)},I_{ik}^{(r)}\right)\text{ for }k=1,2,\ldots,K.\tag{44}$$

Let

$$\rho_{ik} = \min\left\{r \geq 0 : I_{ik}^{(r)} \neq j \text{ or } \left(T_{ik}^{(r)}, Y_{ik}^{(r)}\right) \notin D\right\} \tag{45}$$

be the length of the excursion $\mathscr{Y}_{ik}^{(r)}$ as in (30).

3. Let

$$M_{ij}(m) = \sum_{k=1}^{K} 1_{\{j\}}\left(I_{ik}^{(\rho_{ik})}\right) 1_D\left(T_{ik}^{(\rho_{ik})}, Y_{ik}^{(\rho_{ik})}\right) \tag{46}$$

be the number of i to j transitions of the index process observed while generating the excursions $\mathscr{Y}_{ik}^{(r)}(m)$. Let Let $\left\{T_{ij}^{(\ell)}\right\}_{\ell=1}^{M_{ij}(m)}$ and $\left\{Y_{ij}^{(\ell)}\right\}_{\ell=1}^{M_{ij}(m)}$ be the times $T_{ik}^{(\rho_{ik})}$ and positions $Y_{ik}^{(\rho_{ik})}$ for which $I_{ik}^{(\rho_{ik})} = j$ and $Y_{ik}^{(\rho_{ik})} \in D$.

4. Compute

$$\widehat{G}_{ij}(m) = \frac{M_{ij}(m)}{K}, \tag{47}$$

$$\widehat{\gamma}_{ij}(s, dy; m) = \begin{cases} \dfrac{1}{L \wedge M_{ij}(m)} \displaystyle\sum_{\ell=1}^{L \wedge M_{ij}(m)} \mathbf{1}_{T_{ij}^{(\ell)}}(s)\delta_{Y_{ij}^{(\ell)}}(dy) & \text{if } M_{ij}(m) > 0, \\ 0 & \text{if } M_{ij}(m) = 0, \end{cases} \tag{48}$$

and

$$\left\langle \widehat{f} \right\rangle_i(m) = \frac{1}{K} \sum_{k=1}^{K} \sum_{r=0}^{\rho_{jk}-1} f\left(T_{jk}^{(r)}(m), Y_{jk}^{(r)}(m)\right), \tag{49}$$

where $L \wedge M_{ij}(m) = \min\{L, M_{ij}(m)\}$. In Equation (48), δ_x represents the Dirac delta function centered at position x.

5. Replace the deterministic iteration (43) by the approximation

$$\widetilde{G}_{ij}(m+1) = \widetilde{G}_{ij}(m) + \varepsilon_m\left(\widehat{G}_{ij}(m) - \widetilde{G}_{ij}(m)\right) \tag{50}$$

and

$$\tilde{\gamma}_{ij}(m+1) = \tilde{\gamma}_{ij}(m) + \varepsilon_m\left(\hat{\gamma}_{ij}(m) - \tilde{\gamma}_{ij}(m)\right)\left(\frac{1_{\left\{M_{ij}(m) > 0\right\}}}{I_{ij}(m)}\right) \tag{51}$$

where

$$I_{ij}(m) = \frac{1}{m+1} \sum_{\ell=0}^{m} 1_{\left\{M_{ij}(\ell) > 0\right\}} \tag{52}$$

and $\varepsilon_m > 0$ satisfies

$$\sum_{m=1}^{\infty} \varepsilon_m = \infty \text{ and } \sum_{m=1}^{\infty} \varepsilon_m^2 < \infty . \tag{53}$$

6. Update the expectations

$$\left\langle \tilde{f} \right\rangle_i (m+1) = \left\langle \tilde{f} \right\rangle_i (m) + \varepsilon_m \left(\left\langle \tilde{f} \right\rangle_i (m) - \left\langle \tilde{f} \right\rangle_i (m) \right). \tag{54}$$

7. Once the desired level of convergence has been reached, compute

$$\mathbf{E}\left[\sum_{t=0}^{\tau-1} f\left(t, X^{(t)}\right) \right] \approx \sum_{j=1}^{n} \tilde{z}_j(m) \left\langle \tilde{f} \right\rangle_i (m), \tag{55}$$

where the vector $\tilde{z}(m)$ solves $\tilde{z}^{\mathrm{T}}(m) = \tilde{z}^{\mathrm{T}}(m)\widetilde{G}(m) + a^{\mathrm{T}}$.

We now interpret NEUS as a stochastic approximation algorithm analogous to the deterministic fixed-point iteration (43). First, we observe that $\widehat{G}(m)$ approximates $\mathcal{G}(\widetilde{G}(m), \tilde{\gamma}(m))$ in the following sense. Suppose we were to compute a sequence $\widehat{G}(n)$, $\widehat{G}(n+1), \ldots, \widehat{G}(n+k-1)$ as in NEUS, except holding the values of $\widetilde{G}(n)$ and $\tilde{\gamma}(n)$ fixed. We would then have that $\mathbf{E}[\widehat{G}(n+i)] = \mathcal{G}(\widetilde{G}(n), \tilde{\gamma}(n))$, and that each of the $\widehat{G}(n+i)$ were independent (conditionally on $\widetilde{G}(n)$ and $\tilde{\gamma}(n)$). A Law of Large Numbers would therefore apply and we could conclude that

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=0}^{k-1} \widehat{G}(n+i) = \mathcal{G}(\widetilde{G}(n), \tilde{\gamma}(n)) . \tag{56}$$

The distribution $\gamma_{ij}(m)$ approximates $\Gamma_{i,j}(\widetilde{G}(m), \tilde{\gamma}(m))$ in a similar sense. Therefore, the NEUS iteration (50) is a version of the deterministic fixed-point iteration (43) but with a shrinking sequence ε_m instead of a fixed ε and with random approximations instead of the exact values of $\mathcal{G}$ and Γ. The conditions (53) on the sequence ε_m are common to most stochastic approximation algorithms [29]; they ensure convergence of the iteration when $\mathcal{G}$ and Γ can only be approximated up to random errors.

We remark that in practice the empirical measures $\tilde{\gamma}(m)$ are stored as lists of time and position pairs. The update in (50) allows the number of pairs stored in these lists to grow with each iteration. This can lead to impractical memory requirements for the method. We therefore limit the size of each list $\tilde{\gamma}_{ij}(m)$ to a fixed maximum value by implementing a selection step in which the points that have been stored for the most iterations are removed

to make room for the points in the updates of $\tilde{\gamma}_{ij}(m)$ when this maximum is exceeded. Also, in our numerical experiments in Section V, we use $\varepsilon_m = 1/(m+1)$ in which case,

$$\widetilde{G}_{ij}(m) = \frac{1}{m+1}\sum_{\ell=0}^{m} \widehat{G}_{ij}(\ell) \tag{57}$$

and

$$\tilde{\gamma}_{ij}(m) = \frac{1}{\sum_{\ell=0}^{m} \mathbf{1}_{\left\{M_{ij}(\ell)>0\right\}}} \sum_{\ell=0}^{m} \hat{\gamma}_{ij}(\ell)\,. \tag{58}$$

This and other details of our implementation are explained in Section V.

The implementation detailed above borrows ideas from several earlier modifications of the basic NEUS algorithm. The use of a linear system solve for the weights z was introduced in [11]. In the scheme presented above, the number of samples, K, of the process $Y_j^{(r)}$ is fixed at the beginning of each iteration of the scheme. In this aspect, the implementation above is similar to the Exact Milestoning approach presented in [16]. With the number of samples of $Y_j^{(r)}$ fixed, the total amount of computational effort, as measured in number of time steps of the process $X^{(t)}$, becomes a random variable (with expectation $K\mathbf{E}[\sigma(S^{(\ell)})]$). In practical applications, it may be advantageous to fix the total computational effort expended per iteration in each $J = j$. An alternative version of the NEUS scheme is therefore to fix the total computational effort expended (or similarly the number of numerical integration steps) and allow the number of samples, K, to be a random number. In our tests (not shown here), neither implementation showed a clear advantage provided that a sufficient number of samples, K, was generated to compute the necessary transition statistics.

It is also important to note that if the number of points used in the representation of $\tilde{\gamma}$ is restricted (as it typically has to be in practice), any of the implementations of NEUS that we have described has a systematic error that decreases as the number of points increases *or* as the work per iteration increases. Earlier implementations of NEUS [9–12, 14] computed transition statistics that were normalized with respect to the simulation time spent associated with each $J = j$ rather than the number of samples of $Y_j^{(r)}$ generated. This implementation choice leads to a scheme with a systematic error that vanishes only as the number of points allowed in the representation of $\tilde{\gamma}$ grows, regardless of the work performed per iteration.

IV. ERGODIC AVERAGES

In this section we consider the calculation of ergodic averages with respect to a general (not necessarily time-homogenous) Markov process. We also describe the simplifications that occur when the target Markov process is time-homogenous as in the original NEUS algorithm.

In order to ensure that the definitions in this section are sensible, we require that

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{P}\left[X^{(t)}\in dx, J^{(t)}=i\right] \tag{59}$$

exists as a probability distribution on $\mathbb{R}^d\times\{1,2,\ldots,n\}$ and let

$$\pi(dx)=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{P}\left[X^{(t)}\in dx\right]. \tag{60}$$

This general ergodicity requirement allows processes $X^{(t)}$ with periodicities or time dependent forcing.

Our goal is to compute ergodic averages of the form

$$\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{E}\left[f\left(X^{(t)}\right)\right]=\int_{x\in\mathbb{R}^d}f(x)\pi(dx). \tag{61}$$

To that end, we fix a deterministic time horizon $\tau>0$ in (12) and (18); the condition $t<\tau$ can thus be written as an upper bound of $\tau-1$ on the summation index. If we divide both sides of (17) by τ and take the limit $\tau\to\infty$, we obtain the equation

$$z^{\mathrm{T}}G=z^{\mathrm{T}} \tag{62}$$

where now

$$z_j=\lim_{\tau\to\infty}\frac{1}{\tau}\sum_{t=0}^{\tau-1}\mathbf{P}\left[J^{(t)}=j\right]. \tag{63}$$

and

$$G_{ij}=\lim_{\tau\to\infty}\frac{\sum_{t=0}^{\tau-2}\mathbf{P}\left[J^{(t+1)}=j,J^{(t)}=i\right]}{\sum_{t=0}^{\tau-1}\mathbf{P}\left[J^{(t)}=i\right]}. \tag{64}$$

Note that the matrix G is now stochastic and that $\sum_{j=1}^{n}z_j=1$. We can rewrite the ergodic average of f as

$$\int_{x\in\mathbb{R}^d}f(x)\pi(dx)=\sum_{j=1}^{n}z_j\langle f\rangle_j, \tag{65}$$

where

$$\langle f\rangle_j=\int_{x\in\mathbb{R}^d}f(x)\pi_j(dx) \tag{66}$$

and we represent the large τ limit of the position marginal distribution of π_j defined in (11) as

$$\pi_j(dx) = \lim_{\tau \to \infty} \sum_{t=0}^{\tau - 1} \pi_j(t, dx)\,. \tag{67}$$

These formulas indicate that the only modification of the algorithm in Section III that is required to compute a long-time average is to set $\tau = \infty$ in the definition of the processes $Y_t(\tilde{G}, \tilde{\gamma})$, to set $a = 0$ in (40), and let $\tilde{z}$ solve $\tilde{z}^{\mathrm{T}} = \tilde{z}^{\mathrm{T}}\tilde{G}$ with $\sum_{j=1}^{n} \tilde{z}_j = 1$. In other words, the algorithm seamlessly transitions from solving the initial value problem to solving the infinite time problem as τ becomes large.

When the joint process $(X^{(t)}, J^{(t)})$ is time-homogenous and stationary and our goal is to compute the average of a position dependent observable $f(x)$ with respect to the stationary distribution π of $X^{(t)}$, the above relations can be further simplified. In this case,

$$\pi_j(dx) = \frac{1}{z_j} \lim_{t \to \infty} \mathbf{P}\left[X^{(t)} \in dx, J^{(t)} = j\right], \tag{68}$$

where z_j defined in (63) becomes

$$z_j = \lim_{t \to \infty} \mathbf{P}\left[J^{(t)} = j\right]. \tag{69}$$

The matrix G in (64) can now be written

$$G_{ij} = \lim_{t \to \infty} \mathbf{P}\left[J^{(t+1)} = j \mid J^{(t)} = i\right] \tag{70}$$

and the vector $\langle f \rangle_j$ defined in (66) becomes

$$\langle f \rangle_j = \int_{x \in \mathbb{R}^d} f(x)\pi_j(dx)\,. \tag{71}$$

These simplifications lead to a version of the original NEUS method [9] that employs a direct method for solving for the weights similar to the scheme in [11].

In [11] and [10] the basic NEUS approach was extended to the estimation of transition rates between sets for a stationary Markov process. Implicit in this extension was the observation that any algorithm that can efficiently compute averages with respect to the stationary distribution of a time-homogenous Markov process can be applied to computing dynamic averages more generally by an enlargement of the state space, i.e., by applying the scheme to computing stationary averages for a higher dimensional time-homogenous Markov process. This idea is also central to Exact Milestoning [16], which extends the original Milestoning procedure [17] to compute steady-state averages with respect to a time-homogenous Markov process and is very similar in structure to steady-state versions of NEUS.

V. NUMERICAL EXAMPLES

Here we illustrate the flexibility of the generalized algorithm with respect to both the means of restricting the trajectories (the choice of the $J(t)$ process) and the averages that can be calculated. Specifically, in Section V A we discuss our choice of the $J(t)$ process. In Section V B we show how finite-time hitting probabilities can be calculated by discretizing the state space according to both time and space. In Section V C we show how free energies can be obtained by discretizing the state space according to time and the irreversible work.

A. One Choice of the $J(t)$ Process

Rapid convergence of the scheme outlined in Section III rests on the choice of $J(t)$. Perhaps the most intuitive choice is

$$J^{(t)} = \sum_{j=1}^{n} j \mathbf{1}_{A_j}\left(t, X^{(t)}\right) \tag{72}$$

where the subsets $A_1, A_2, \ldots, A_n$ partition $\mathbb{N} \times \mathbb{R}^d$. Indeed, earlier steady-state NEUS implementations [9–12, 14] employed an analogous rule using a partition of the space variable (the time variable was not stored or partitioned). However, even with an optimal choice of the subsets $A_1, A_2, \ldots, A_n$, (72) has an important disadvantage: in many situations, $X^{(t)}$ frequently recrosses the boundary between neighboring subsets A_i and A_j, which slows convergence. Fortunately, there are many alternative choices of $J(t)$ that approximate the choice in (72) while mitigating this issue. We give one simple and intuitive alternative which we use in the numerical examples that follow.

Let ψ_j be a set of non-negative functions on $\mathbb{N} \times \mathbb{R}^d$ for which $\sum_{j=1}^{n} \psi_j = 1$. The ψ_j are generalizations of the functions 1_{A_j} in that they serve to restrict trajectories to regions of state space. In practice, given a partition of space $A_1, A_2, \ldots, A_n$, the ψ_j can be chosen to be smoothed approximations of the functions 1_{A_j} Given a trajectory of $X^{(t)}$, the rule defining $J(t)$ is as follows. Initially, choose $J^{(0)} \in \{1, 2, \ldots, n\}$ with probabilities proportional to $\{\psi_1(0, X^{(0)}), \psi_2(0, X^{(0)}), \ldots, \psi_n(0, X^{(0)})\}$. At later times $J(t)$ evolves according to the rule

1. If $\psi_{J(t-1)}\left(t, X^{(t)}\right) > 0$ then $J^{(t)} = J^{(t-1)}$.
2. Otherwise sample $J(t)$ independently from $\{1, 2, \ldots, n\}$ according to probabilities $\{\psi_1(t, X^{(t)}), \psi_2(t, X^{(t)}), \ldots, \psi_n(t, X^{(t)})\}$.

While transitions out of $J(t) = i$ occur when $X^{(t)}$ leaves the support of ψ_i, transitions back into $J(t) = i$ can only occur outside of the support of ψ_j. Thus, this transition rule allows one to separate in space the values of $X^{(t)}$ at which $J(t)$ transitions away from i from those where $J(t)$ transitions into i, mitigating the recrossing issues mentioned above.

In our examples, we discretize time and only one additional "collective variable" (a dihedral angle in Section V B and the nonequilibrium work in Section V C). Here we denote the collective variable by ϕ, and we discretize it within some interval of values $[a,b]$ (though it

may take values outside this interval). In both examples $[a, b]$ is evenly discretized into a set of points $\left\{a + k(b-a)/m_\phi\right\}_{k=0}^{m_\phi}$ for some integer m_ϕ. Letting ϕ_j be any of the points in that discretization, we set

$$\psi_j(t,x) \propto \begin{cases} \left[1 - \frac{1}{\Delta_\phi}\left|\phi(x) - \phi_j\right|\right] \mathbf{1}_{[a,b]} & \text{if } \left|\phi(x) - \phi_j\right| \le \Delta_\phi \text{ and } t \in \left[t^j_{start}, t^j_{end}\right) \\ 0 & \text{otherwise} \end{cases} \tag{73}$$

where Δ_ϕ is some fixed value controlling the width of the support of ψ_j, and the indicator $\mathbf{1}_{[a,b]}$ restricts the terminal functions. Recall that the ψ_j are required to sum to 1. We choose t^j_{start} and t^j_{end} to equally divide the interval $[0, \tau)$, where, in our examples, τ is a fixed time horizon. The function ψ_j is largest when $t \in \left[t^j_{start}, t^j_{end}\right)$ and $\phi(x) = \phi_j$. The supports of the various ψ_j correspond to products of overlapping intervals in the ϕ variable, but non-overlapping intervals in time. The fact that ψ_j depends on time is essential in our examples.

B. Finite-Time Hitting Probability

In this section we compute the probability, $P_{BA}(\tau_{\max})$, of hitting a set B before a separate set A and before a fixed time $\tau_{\max} > 0$ given that the system is at a point $X^{(0)} \notin A \cup B$ at time $t = 0$. In the case where $X^{(0)}$ and B are separated by a large free energy barrier while $X^{(0)}$ and A are not, computing $P_{BA}(\tau_{\max})$ can be challenging since trajectories that contribute to $P_{BA}(\tau_{\max})$ are rare in direct simulations. To compute $P_{BA}(\tau_{\max})$ via the scheme in Section III C, we let the stopping time τ be the minimum of $\tau_{\max}$ and the first time, t, at which $X^{(t-1)}$ is in either A or B, i.e., $\tau - 1 = \min\{\tau_A, \tau_B, \tau_{\max} - 1\}$ where τ_A and τ_B are the first times that enters the sets A and B respectively. Strictly speaking, to write τ in the form in (9), we need to replace $(t, X^{(t)})$ in that equation by $(t, X^{(t-1)}, X^{(t)})$. The set D corresponding to our choice of τ is then $D = \{(t, x, y): t < \tau_{\max}, x \notin (A \cup B)\}$. As we have already mentioned, this can be done without further modification of the scheme. Then $f(t, X^{(t)})$ in (10) is

$$f\left(t, X^{(t)}\right) = 1_B\left(X^{(t)}\right). \tag{74}$$

The system that we simulate is the alanine dipeptide (CH_3-CONH-$C^\alpha H(C^\beta H_3)$-CONH-CH_3) in vacuum modeled by the CHARMM 22 force field [31]. We use the default Langevin integrator [32] implemented in LAMMPS [33], with a temperature of 310 K, a timestep of 1 fs and a damping coefficient of $30 ps^{-1}$. The SHAKE algorithm is used to constrain all bonds to hydrogens [34]. We consider the system to be in set A if $-150° < \phi < -100°$ and in set B if $30° < \phi < 100°$ (Figure 2). We discretize time into intervals of $t_{end} - t_{start} = 10^3$ time steps with a terminal time of $\tau_{\max} = 10^4$ time steps. We use the rule outlined in Section V A for the evolution of $X^{(t)}$ with the ψ_j of the form in (73). The ϕ_j in (73) are chosen from the set $\{-100°, -74°, -48°, -22°, 4°, 30°\}$ with $[a, b] = [100°, 30°]$ and $\Delta_\phi = 20°$.

We generate the initial point $X^{(0)}$ by running an unbiased simulation at 310 K and choosing a single point $X^{(0)}$ between the sets A and B. The vector a defined in (19) is

$$a_j = \frac{\psi_j\left(0, X^{(0)}\right)}{\sum_{i=1}^{n} \psi_i\left(0, X^{(0)}\right)}. \tag{75}$$

Note that the initial condition at $\mathcal{J}^{(0)}$ can be drawn from an ensemble of configurations with minimal changes to the algorithm, but we restrict our attention to the initial condition consisting of a single point. To evaluate the performance of the algorithm in Section III C, we choose two points from our direct simulation, one at $\phi = -58.0°$ and one at $\phi = -91.0°$. The former is chosen to allow the NEUS results to be compared with results from unbiased direct simulations, while the latter provides a more challenging test because P_{BA} becomes small when $X^{(0)}$ is close to A.

We set $K = 100$ and $L = 1$ and perform a total of 10^4 iterations (about 7.2 μs of dynamics) of the scheme in Section III C for each starting point. Each step of the process $\mathcal{Y}_j^{(r)}(\tilde{G}(m), \tilde{\gamma}(m))$ corresponds to 10 time steps of the physical model. The $\tilde{\gamma}_{ij}$ are represented as lists of time and position pairs with associated weights. We cap the maximum size of those lists at 25 entries. If $\overline{\Pi}_j(s, dy; \tilde{G}(m), \tilde{\gamma}(m))$ by the following. With probability a_j/z_j, set $S = 0$ and select Y from $\mathbf{P}[X^{(0)} \in dy | \mathcal{J}^{(0)} = j]$, or with the remaining probability select an index I proportional to the flux $\tilde{z}_i \tilde{G}_{ij}$ and then select (S, Y) from the list of weighted samples comprising $\tilde{\gamma}_{Ij}(m)$. For each j we compute $f_j = P_{BA}^j = M_{jB}/(mK)$ where M_{jB} is the total number of transition events of $X_j^{(r)}$ into B observed after m iterations (mK is the total number of excursions in state j after m iterations). The estimate of $P_{BA}(\tau_{\max})$ after m iterations is then computed as $P_{BA}\left(\tau_{\max}\right) = \sum_{j=1}^{n} P_{BA}^j \tilde{z}_j(m)$.

To assess the efficiency of the trajectory stratification, we also estimate $P_{BA}(\tau_{\max})$ by integrating an ensemble of $n = 10^6$ unbiased dynamics trajectories for $\tau_{\max}$ time steps from the initial point $X^{(0)}$. In this case, $P_{BA}(\tau_{\max}) \approx N_B/N$, where N_B is the number of trajectories that hit set B before set A. To assess the accuracy of the NEUS result, we perform 10 independent NEUS calculations. In each NEUS simulation, we estimate the value of P_{BA} as the average over the final 1000 iterations of each simulation and compute the mean of this estimate over 10 independent NEUS simulations. We obtain $P_{BA}(\tau_{\max}) \approx 4.43 \times 10^{-4}$ from NEUS and $P_{BA}(\tau_{\max}) \approx 4.12 \times 10^{-4}$ from direct simulation for the starting point at $\phi = -58.0°$ (Figure 3). In this case, the NEUS result is within the 95% confidence interval [3.72×10^{-4}, 4.52×10^{-4}] (estimated as $\pm 1.96\sqrt{p(1-p)/n}$, where p is the estimate of P_{BA} from the direct simulation) for the direct simulation estimate given the number of samples. We obtain $P_{BA}(\tau_{\max}) \approx 2.78 \times 10^{-8}$ from NEUS for the starting point at $\phi = -91.0°$, consistent with the fact that none of the unbiased trajectories reached B before A in this case. From the same data (for either NEUS or direct simulation), one can easily assemble estimates of $P_{BA}(t)$ for any $t \leq \tau_{\max}$ by counting only those transitions into B that occur before t time steps. Up to a normalization, $P_{BA}(t)$ is the cumulative distribution function for the time that it takes $X^{(t)}$ to enter B conditioned on not entering A. Estimates of this cumulative distribution function compiled from the NEUS and direct simulation data are plotted in Figure 4. The NEUS results show excellent agreement with the results from the direct simulation.

Spatiotemporal plots of the weights computed from the converged NEUS calculations and the direct simulations are shown in Figure 5. For both starting points, the stratification scheme is able to efficiently sample events with weights spanning 12 orders of magnitude. When $X^{(0)}$ is close to the boundary of set A, accurate estimation of the very small probability $P_{BA}(\tau_{\max})$ depends sensitively on the ability to realize a set of very rare trajectories, ruling out the use of direct simulation.

C. Free Energy Differences via the Jarzynski Equation

In this section, we show how a specific choice of the $X^{(t)}$ process enables us to stratify a path-dependent variable, specifically, the accumulated work appearing in the Jarzynski equation [8, 35]. For a statistical model defined by a density proportional to $\exp[-V(x)]$ (e.g., $V(x)$ is a potential function or a log-likelihood), the normalization constant is $Q = \int e^{-V(x)} dx$. In fields ranging from statistics to chemistry, a ratio of normalization constants is often used to compare models [36, 37]. Subject to certain conditions [35, 38], the Jarzynski equation relates the ratio of normalization constants to an average over paths of a time-dependent process, $X^{(t)}$:

$$\frac{Q_t}{Q_0} = \mathbf{E}\left[\exp\left(-W^{(t)}\right)\right] \tag{76}$$

where

$$W^{(t)} = \sum_{\ell=0}^{t-1} V\left(\ell+1, X^{(\ell)}\right) - V\left(\ell, X^{(\ell)}\right), \quad W^{(0)} = 0 \tag{77}$$

and we refer to $\Delta F = -\log(Q_t/Q_0)$ as the free energy difference. For example, for a small time discretization parameter, dt, a suitable choice of dynamics is

$$X^{(t+1)} = X^{(t)} - \frac{\partial V\left(t+1, X^{(t)}\right)}{\partial x} dt + \sqrt{2dt}\xi_t \tag{78}$$

where ξ_t is a standard Gaussian random variable and $X^{(0)}$ is drawn from $p_0 \propto \exp[-V(0,x)]$.

Formula (76) suggests a numerical procedure for estimating free energy differences in which one simulates many trajectories of $X^{(t)}$, evaluates the work $W^{(t)}$ for each, and then uses this sample to compute the expectation on the right hand side of (76) approximately. This approach has been particularly useful in the context of single-molecule laboratory experiments [39, 40]. A well-known weakness of this strategy in the fast-switching (small t) regime is large statistical errors result from the fact that low-work trajectories contribute significantly to the expectation but are infrequently sampled [39, 41–44].

The quantity that we seek to compute is the free energy difference between a particle in a double-well potential that is additionally harmonically restrained with spring constant $k = 20$ near $x = -1$ and a particle in the same potential restrained near $x = 1$. The model is adapted from the one presented in [36]. Setting $\tau = 501$, for $t < \tau$ we define

$$V(t,x) = 5\left(x^2 - 1\right)^2 + 3x + k(x - (2tdt - 1))^2 \tag{79}$$

where $dt = 0.001$. We show $V(0, x)$, $V(\tau - 1, x)$ and $V(x; k = 0)$ in Figure 6. The process $X^{(t)}$ evolves according to (78).

The reader may be concerned that the expectation in (76) is not immediately of the general form in (10) suitable for an application of NEUS. We apply NEUS as described in Section II B to the augmented process $Z^{(t)} = (X^{(t)}, W^{(t)})$. To compute the expectation of the left hand side of (76) via NEUS, we compute the expectation in (10) with

$$f\left(t, Z^{(t)}\right) = \begin{cases} \exp\left(-W^{(t)}\right) & \text{if } t = \tau - 1 \\ 0 & \text{if } t \neq \tau - 1. \end{cases} \tag{80}$$

The index process $\mathcal{J}^{(t)}$ marks transitions between regions of the time t and accumulated work $W^{(t)}$ variables. We discretize the work space in overlapping subsets using the pyramid form in (73). We use 100 subsets with centers evenly spaced on the interval [−35.0,35.0] with a width of $\Delta_\phi = 0.6$. We discretize time into 5 discrete nonover-laping subsets every 100 time steps for a total of 500 subsets. We cap the maximum size of the list representation of $\left\{\tilde{\gamma}_{ij}\right\}$ at 50 entries using the same scheme as in Section V B.

To assess the accuracy of the NEUS result, we perform 10 independent NEUS simulations. For both NEUS and direct simulations, we prepare an ensemble of 1000 starting states $X^{(0)}$ by performing an unbiased simulation with fixed potential $V(0, x)$ for 10^6 steps, saving every 1000 steps. The direct fast-switching simulations start from each of these points and comprise 500 steps of integration forward in time; each trajectory contributes equally to the left hand side of (76). For the NEUS simulations, the vector a is constructed as in (75), and trajectories are initialized at $\mathcal{J}^{(0)}$ by drawing uniformly from this ensemble. We set $K = 100$ and $L = 1$, and we perform 500 iterations. Each step in K corresponds to a single step of (78). As in Section V B, we sample only in the restricted distributions where there is at least one point stored in $\tilde{\gamma}$ from which to restart the dynamics.

The estimated ΔF produced from data generated in the last 50 iterations of NEUS is 5.89 (the units are chosen to absorb temperature factors above), which is in excellent agreement with the reference value of 5.94, in contrast to the estimate from direct simulation (Figure 7). The left panel of Figure 8 shows the weights along the time and work axes. In the right panel of Figure 8 we plot histogram approximations of the density $P_W(w)$ of $W^{(\tau-1)}$ along with the weighted density proportional to $P_W(w)\exp(-w)$. The separation of the peaks of this distribution highlight how NEUS is able to effectively sample the low work tails that contribute significantly to the expectation in the Jarzynski relation in (76) but are rarely accessed by the switching procedure in the unbiased simulations.

VI. CONCLUSIONS

We describe a trajectory stratification framework for the estimation of expectations with respect to arbitrary Markov processes. The basis for this framework is the nonequilibrium

umbrella sampling method (NEUS) originally introduced to compute steady state averages. Our development highlights the structural similarities between the nonequilibrium and equilibrium US algorithms and places the NEUS method within the general context of stochastic approximation. These connections have practical implications for further optimizing the procedure and point the way to a more in depth convergence analysis that will be the subject of future work.

Our development reveals that the basic trajectory stratification approach can be useful well beyond the estimation of stationary averages for time-homogenous Markov processes. This flexibility is demonstrated in two examples, both involving an expectation over trajectories of finite duration. In the first example, we show that the probability of first hitting a set within a finite time can be efficiently computed via stratification even when the dynamics start close to a competing absorbing state. In our second example, we use NEUS to stratify a process according to a path-dependent variable, the accumulated work in a nonequilibrium process appearing in the Jarzynski equation. The result is a novel and effective scheme for estimating free energy differences by enhancing sampling of the tails of the accumulated work distribution.

Our general framework also suggests new and exciting applications of trajectory stratification. For example, with little modification, these methods can be applied to sequential data assimilation applications where the goal is to approximate averages with respect to the conditional distribution of a hidden signal $X(t)$ given sequentially arriving observations (i.e., with respect to the posterior distribution). In high-dimensional settings (e.g., weather forecasting) the only practical alternatives are limited to providing information about only the mode of the posterior distribution (i.e., variational methods) or involve uncontrolled and often unjustified approximations (i.e., Kalman-type schemes). The approach that we present here opens the door to efficient data assimilation, machine learning, and, more generally, new forms of analysis of complex dynamics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors would like to thank David Aristoff, James Dama, Jianfeng Lu, Charles Matthews, Erik Thiede, Omiros Papaspiliopoulos, and Eric Vanden-Eijnden for helpful discussions. This research is supported by the National Institutes of Health (NIH) Grant Number 5 R01 GM109455-02. Computational resources were provided by the University of Chicago Research Computing Center (RCC).

Appendix A: An Alternative F

Here we present an alternative construction of the stochastic matrix F (Section II A) that more closely aligns with the nonequilibrium version of the algorithm presented in Section II B. Suppose that one has available a transition distribution $p(dy \mid x)$ for a Markov chain that preserves (or nearly preserves) the target density, π, in the sense that

$$\pi(dy) = \int_{x \in \mathbb{R}^d} p(dy \mid x)\pi(dx)\,. \tag{A1}$$

For example, $p(dy \mid x)$ might be the transition density for a number of steps of a Langevin dynamics integrator. We can again express the z_i as the solution to an eigenproblem (8) where now

$$F_{ij} = \int_{y \in \mathbb{R}^d} \int_{x \in \mathbb{R}^d} \psi_j(y) p(dy \mid x)\pi_i(dx)\,. \tag{A2}$$

Note that when $\psi_i(x) = \mathbf{1}_{Ai}$ for some partition of space $\{A_i\}$, and $p(dy \mid x)$ is reversible with respect to π, the entry F_{ij} can be estimated by evolving samples according to $p(dy \mid x)$, rejecting any proposed samples that lie outside of A_i (so that π_i is preserved), and then counting the number of times the chain attempts transitions from set A_i to set A_j. For a closely related approach to approximating certain nonequilibrium quantities see [45].

Appendix B: Expressions for $\bar{G}$ and $\bar{z}$

In this appendix we establish the identities

$$\bar{G}_{ij} = \begin{cases} \dfrac{G_{ij}}{1 - G_{ii}}, & i \neq j \\ 0, & i = j \end{cases} \quad \text{and} \quad \bar{z}_j = \left(1 - G_{jj}\right) z_j \tag{B1}$$

appearing in (22) and (24). First, note that the equality $\bar{z}_i \bar{G}_{ij} = z_i G_{ij}$ for $i \neq j$ (which follows immediately from the definitions of $\bar{z}$, $\bar{G}$, z, and G) together with $1 - G_{jj} = \bar{z}_j / z_j$ implies the expression for $\bar{G}$ in terms of G. It remains then only to establish the expression for $\bar{z}$ in terms of z and G. To that end, notice that

$$\begin{aligned} z_j &= \sum_{t=0}^{\infty} \mathbf{P}[J(t) = j, t < \tau] \\ &= \mathbf{P}\left[J^{(0)} = j\right] + \sum_{t=0}^{\infty} \mathbf{P}\left[t + 1 < \tau, J^{(t+1)} = j, J^{(t)} = j\right] \\ &+ \sum_{t=0}^{\infty} \mathbf{P}\left[t + 1 < \tau, J^{(t+1)} = j, J^{(t)} \neq j\right] \\ &= a_j + z_j G_{jj} \\ &+ \sum_{t=0}^{\infty} \sum_{\ell=0}^{\infty} \mathbf{P}\left[S^{(\ell+1)} < \tau, S^{(\ell+1)} = t + 1, J^{\left(S^{(\ell+1)}\right)} = j\right] \\ &= a_j + z_j G_{jj} + \sum_{\ell=0}^{\infty} \mathbf{P}\left[S^{(\ell+1)} < \tau, J^{\left(S^{(\ell+1)}\right)} = j\right] \\ &= z_j G_{jj} + \tilde{z}_j \end{aligned} \tag{B2}$$

so that

$$\frac{\tilde{z}_j}{z_j} = \left(1 - G_{jj}\right). \tag{B3}$$

Appendix C: Excursions sample the restricted distributions

Here, we establish (32). We have

$$\begin{aligned} z_j \pi_j(t, dx) &= \mathbf{P}\left[t < \tau, X^{(t)} \in dx, J^{(t)} = j\right] \\ &= \mathbf{P}\left[J^{(0)} = j, t < \sigma(0) \wedge \tau, X^{(t)} \in dx\right] + \sum_{s=1}^{t} \mathbf{P} \\ &\left[J^{(s)} = j, J^{(s-1)} \neq j, t < \sigma(s) \wedge \tau, X^{(t)} \in dx\right] \\ &= \sum_{s=0}^{t} \sum_{\ell=0}^{\infty} \mathbf{P}\left[s = S^{(\ell)}, t < \sigma(s) \wedge \tau, X^{(t)} \in dx\right] \\ &= \bar{z}_j \sum_{s=0}^{t} \int_y \mathbf{P}_{s,y,j}\left[t < \sigma(s) \wedge \tau, X^{(t)} \in dx\right] \bar{\pi}_j(s, dy) \\ &= \bar{z}_j \mathbf{P}\left[t < \rho_j + T_j^{(0)}, Y_j^{\left(t - T_j^{(0)}\right)} \in dx\right]. \end{aligned} \tag{C1}$$

References

[1]. Asmussen S and Glynn PW, Stochastic Simulation: Algorithms and Analysis (Springer, 2007).

[2]. Gardiner CW, Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer, 2009).

[3]. Frenkel D and Smit B, Understanding Molecular Simulation (Academic Press, 2002).

[4]. Neyman J, Journal of the Royal Statistical Society 97, 558 (1934).

[5]. Torrie GM and Valleau JP, Journal of Computational Physics 23, 187 (1977).

[6]. Pangali C, Rao M, and Berne BJ, J. Chem. Phys 71, 2975 (1979).

[7]. Chandler D, Introduction to Modern Statistical Mechanics (Oxford University Press, 1987).

[8]. Leliévre T, Rousset M, and Stoltz G, Free Energy Computations: A Mathematical Perspective (Imperial College Press, 2010).

[9]. Warmflash A, Bhimalapuram P, and Dinner AR, J. Chem. Phys 127, 154112 (2007). [PubMed: 17949137]

[10]. Dickson A, Warmflash A, and Dinner AR, J. Chem. Phys 131, 154104 (2009). [PubMed: 20568844]

[11]. Vanden-Eijnden E and Venturoli M, J. Chem. Phys 131, 044120 (2009). [PubMed: 19655850]

[12]. Dickson A, Warmflash A, and Dinner AR, J. Chem. Phys 130, 074104 (2009). [PubMed: 19239281]

[13]. Dickson A and Dinner AR, Annual review of physical chemistry 61, 441 (2010).

[14]. Dickson A, Maienschein-Cline M, Tovo-Dwyer A, Hammond JR, and Dinner AR, J. Chem. Theory Comput 7, 2710 (2011). [PubMed: 26605464]

[15]. Xu X, Rice SA, and Dinner AR, Proceedings of the National Academy of Sciences 110, 3771 (2013).

[16]. Bello-Rivas JM and Elber R, J. Chem. Phys 142, 094102 (2015). [PubMed: 25747056]

[17]. Faradjian AK and Elber R, J. Chem. Phys 120, 10880 (2004). [PubMed: 15268118]

[18]. Glasserman P, Heidelberger P, Shahabuddin P, and Zajic T, "A look at multilevel splitting," in Monte Carlo and Quasi-Monte Carlo Methods 1996: Proceedings of a conference at the University of Salzburg, Austria, 7 9–12, 1996, edited by Niederreiter H, Hellekalek P, Larcher G, and Zinterhof P (Springer New York, New York, NY, 1998) pp. 98–108.

[19]. Huber GA and Kim S, Biophys. J 70, 97 (1996). [PubMed: 8770190]

[20]. Haraszti Z and Townsend JK, ACM Trans. Model. Comput. Simul 9, 105 (1999).

[21]. van Erp TS, Moroni D, and Bolhuis PG, J. Chem. Phys 118, 7762 (2003).

[22]. Allen RJ, Warren PB, and ten Wolde PR, Phys. Rev. Lett 94, 018104 (2005). [PubMed: 15698138]

[23]. Johansen A, Del Moral P, and Doucet A, in Proceedings of the 6th International Workshop on Rare Event Simulation (Bramberg, 2006).

[24]. Cérou F and Guyader A, Stochastic Analysis and Applications 25, 417 (2007).

[25]. Guttenberg N, Dinner AR, and Weare J, J. Chem. Phys 136, 234103 (2012). [PubMed: 22779577]

[26]. Hairer M and Weare J, Commun. Pure Appl. Math 67, 1995 (2014).

[27]. Thiede E, Van Koten B, Weare J, and Dinner AR, J. Chem. Phys 145, 084115 (2016). [PubMed: 27586912]

[28]. Dinner AR, Thiede E, Van Koten B, and Weare J, arxiv 1705.08445 (2017).

[29]. Kushner HJ and Yin GG, Stochastic Approximations and Recursive Algorithms and Applications, second edition ed. (Springer, 2003).

[30]. Schütte C, Noé F, Lu J, Sarich M, and Vanden-Eijnden E, J. Chem. Phys 134, 204105 (2011). [PubMed: 21639422]

[31]. MacKerell AD Jr., Bashford D, Bellott M, Dunbrack JRL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wio J´Yin rkiewiczKuczera, D., and Karplus M, J. Phys. Chem. B 102, 3586 (1998). [PubMed: 24889800]

[32]. Schneider T and Stoll E, Phys. Rev. B 17, 1302 (1978).

[33]. Plimpton S, J. Comp. Phys 117, 1 (1995).

[34]. Ryckaert J, Ciccotti G, and Berendsen JC, J. Comp. Phys 23, 327 (1977).

[35]. Jarzynski C, Phys. Rev. Lett 78, 2690 (1997).

[36]. Chipot C and Pohorille A, Free Energy Simulations (Springer, 2007).

[37]. Kass RE and Raftery AE, Journal of the American Statistical Association 90, 773 (1995).

[38]. Neal RM, Stat. Comput 11, 125 (2001).

[39]. Hummer G, J. Chem. Phys 114, 7330 (2001).

[40]. Hummer G and Szabo A, Biophys. J 85, 5 (2003). [PubMed: 12829459]

[41]. Ytreberg FM and Zuckerman DM, J. Chem. Phys 120, 10876 (2004). [PubMed: 15268117]

[42]. Oberhofer H, Dellago C, and Geissler PL, J. Phys. Chem. B 109, 6902 (2005). [PubMed: 16851777]

[43]. Jarzynski C, Phys. Rev. E 73, 046105 (2006).

[44]. Vaikuntanathan S and Jarzynski C, J. Chem. Phys 134, 054107 (2011). [PubMed: 21303092]

[45]. Vanden-Eijnden E and Venturoli M, The Journal of Chemical Physics 130, 194101 (2009). [PubMed: 19466815]

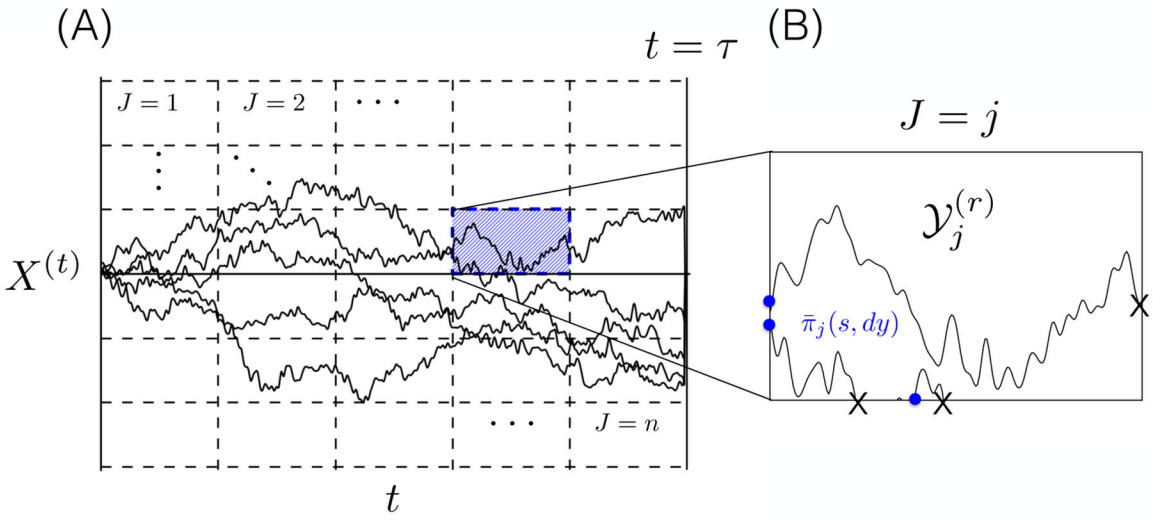

FIG. 1.

Illustration of the stratification of a process $(X^{(t)}, J^{(t)})$ (solid black lines, panel A) via the scheme outlined in Section II B. (A) The restricted distributions corresponding to each value of the index process $J^{(t)}$ are outlined as discrete regions of the $(t, X^{(t)})$ space (panel A, black dashed lines). In this depiction, the value of $J^{(t)}$ corresponds to the current cell containing $(t, X^{(t)})$ within a rectangular grid of times and positions. (B) Each of the restricted distributions $\pi_j(t, dx)$ are sampled by integrating a locally restricted dynamics $\mathcal{Y}_j^{(r)}$ (panel B, black lines). The $\mathcal{Y}_j^{(r)}$ process is generated by integrating an excursion of the unbiased process $(X^{(t)}, J^{(t)})$ corresponding to a particular fixed value of $J = j$ (panel A). As each excursion transitions from $J = i$ to $J = j$ with $j \neq i$, the dynamics are stopped and a new excursion is started at a time and point (s, y) (panel B, blue dots) drawn from the flux distribution $\bar{\pi}_j(s, dy)$.

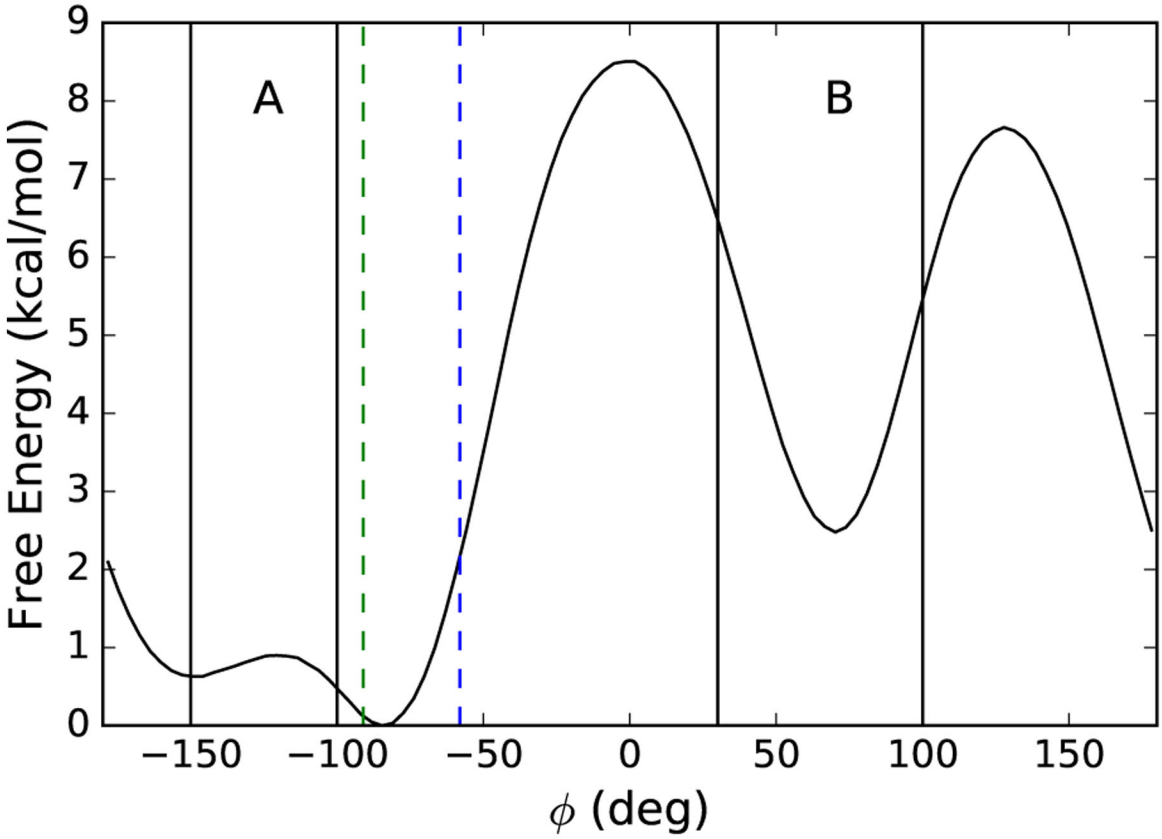

FIG. 2.

Free energy (black curve) of the alanine dipeptide projected onto the ϕ dihedral angle, with sets A and B indicated. The initial positions of $X^{(0)}$ at $\phi = -58.0°$ (blue) and $\phi = -91.0°$ (green) are shown as vertical dashed lines. The free energy is computed from the method presented in Section II A as implemented in [27].

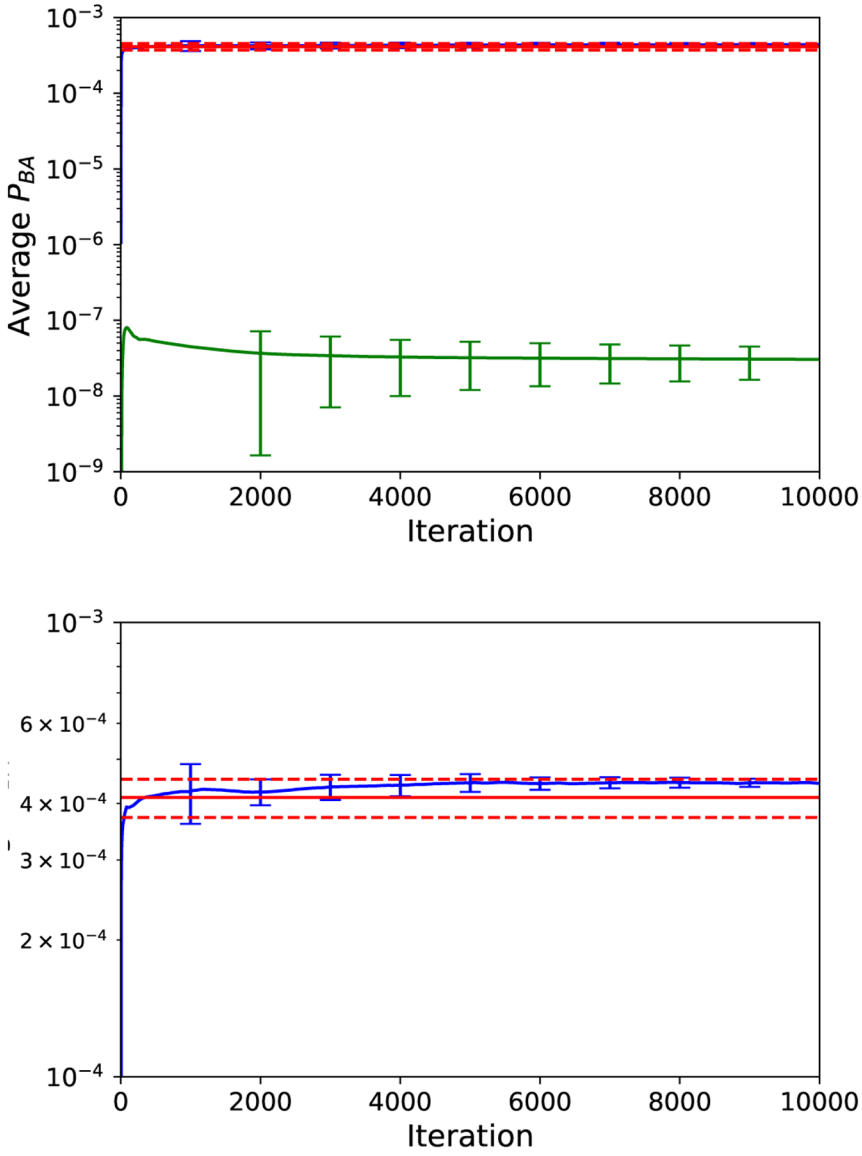

FIG. 3.

Running estimate of P_{BA} from NEUS for dynamics starting at $\phi = -58.0°$ (blue, upper curve; error bars are computed every 1000 iterations and indicate $\pm 2.262 s/\sqrt{n}$ where s is the standard error estimated from $n = 10$ independent NEUS simulations) compared to the final result from direct simulation (red solid line; dashed lines indicate $\pm 1.96\sqrt{p(1-p)/n}$, where $n = 10^6$ is the number of physically weighted trajectories generated and p is the estimate of P_{BA} from the direct simulation). Also shown is the estimate from NEUS for dynamics starting from $\phi = -91.0°$ (green, lower curve; error bars computed similarly as the blue curve). The estimate at each iteration is computed as the average of the previous 1000 iterations. Lower panel is a magnification of the upper panel.

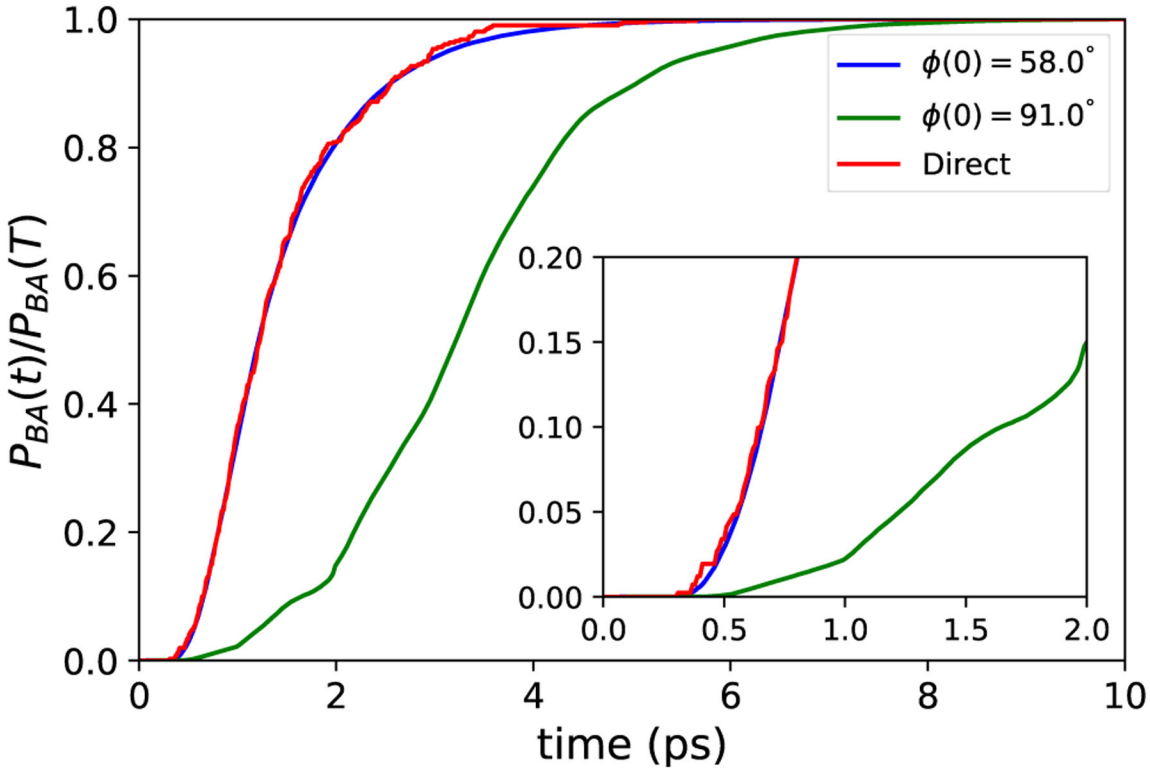

FIG. 4.

Estimate of the cumulative distribution function of the time to enter set B conditioned on not entering A from NEUS for the dynamics starting at $\phi = -58.0°$ (blue) and $\phi = -91.0°$ (green) compared to the result from the direct simulation (red). (Inset) The early time portion is shown. The estimate from each NEUS simulation at each time is computed as an average over the last 1000 iterations of the calculation and then averaged over 10 independent NEUS simulations.

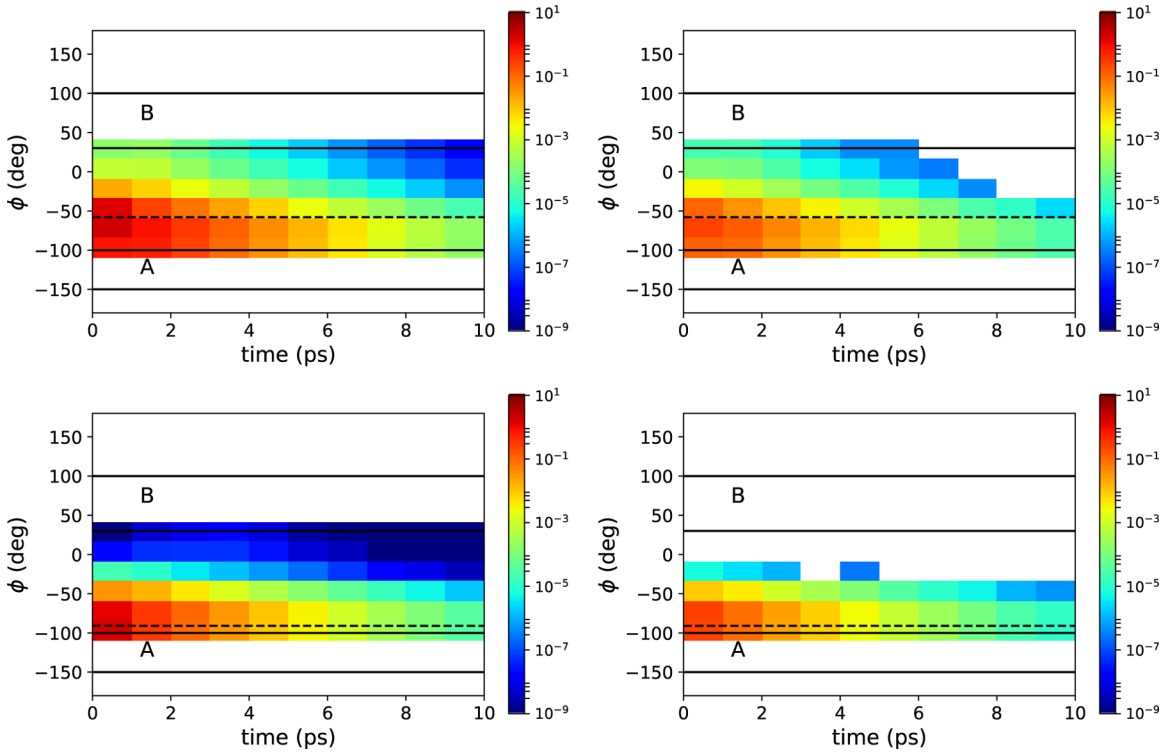

FIG. 5.

Estimates of the subset weights from NEUS (left) and direct simulations (right). Upper panels show the dynamics starting from $\phi = -58.0°$ (dashed line) and lower panels show the dynamics starting from $\phi = -91.0°$ (dashed line). White space represents subsets which were not sampled.

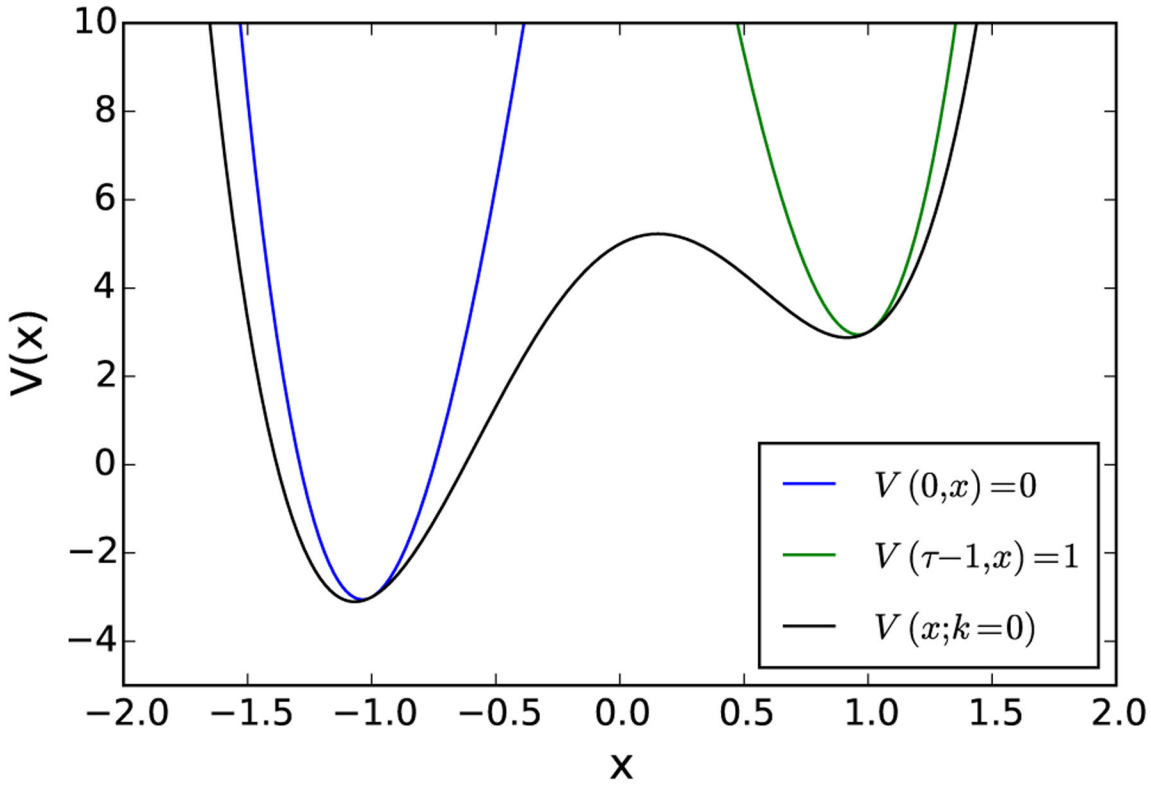

FIG. 6.

$V(0, x)$ (blue) and $V(\tau-1, x)$ (green) for the switching process used to compute Jarzynski's equality. For reference, the potential with $k = 0$ (black) is also shown.

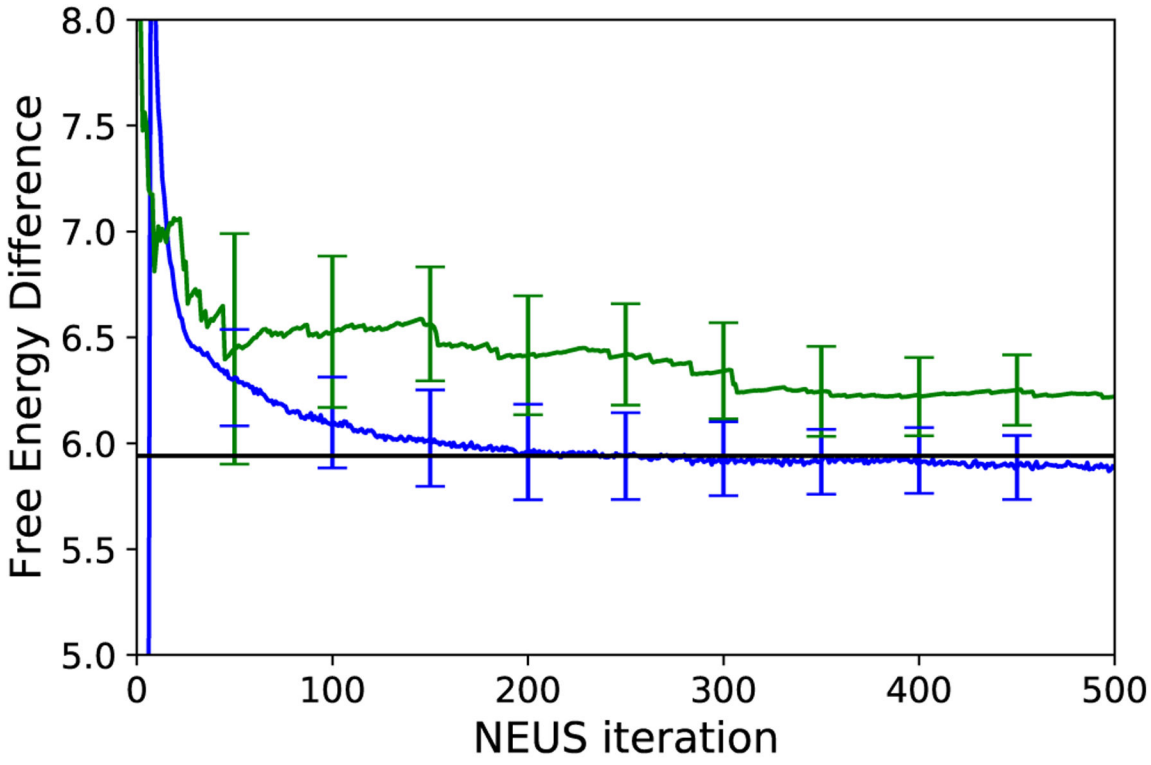

FIG. 7.

Estimate of the free energy computed from NEUS (blue; error bars are computed every 50 iterations and indicate $\pm 2.262 s/\sqrt{n}$ where s is the standard error estimated from $n = 10$ independent NEUS simulations) and from conventional fast-switching simulations (green; error bars are computed every 50 iterations and indicate $\pm 2.262 s/\sqrt{n}$ where s is the standard error estimated from $n = 10$ independent direct simulations). The value computed from numerically integrating the potentials is shown as a black line. For the direct fast-switching simulations, we scale the number of repetitions to the number of NEUS iterations that are equivalent in computational effort.

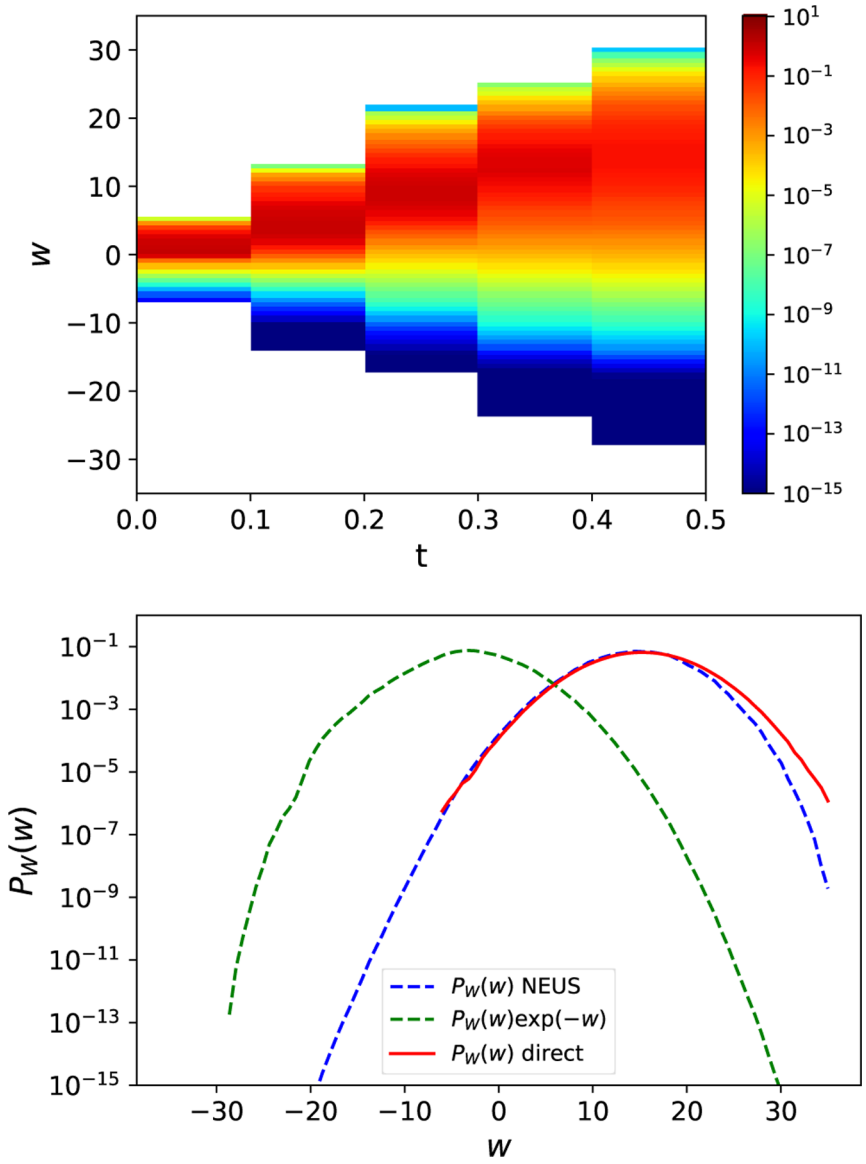

FIG. 8.

Sampling the work with NEUS. (top) The estimate of the dynamic weights, $\bar{z}_j$, from the final iteration of the NEUS calculation. White space represents subsets that are not visited in the NEUS calculation. (bottom) The probability density $P_W(w)$ of the accumulated work $W^{(\tau-1)}$ estimated from NEUS (blue dashed line), from direct integration (red solid line) and the exponentially scaled probability density proportional to $P_W(w)\exp(-w)$ estimated from the NEUS calculations (green dashed line). The estimates of $P_W(w)$ and $P_W(w)\exp(-w)$ from NEUS (blue dashed line and green dashed line respectively) at each value of $W^{(\tau-1)}$ are computed as an average over the last 10 iterations and then averaged over 10 independent NEUS simulations. The estimate of $P_W(w)$ from direct integration (red solid line) is computed as an average over 10 independent direct simulations that are equivalent in computational effort to the 10 independent NEUS simulations.