Computing equilibrium quantities through invertible non-equilibrium processes

by

Kangxin Liu

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

New York University

August, 2025

Glen M. Hocky

Dedication

To my family: Hua Shi (Anbing Gong), Dongwen Liu

Acknowledgements

I am thankful first and foremost to my advisor Glen M. Hocky, who accepted me as second generation group member and gave me a great opportunity to pursue research in development and application of non-equilibrium methods. I really appreciate it that Glen gave me freedom and time to solve challenging problems and to discover new ideas. I believe that this invaluable experience makes me not afraid of difficulties in research as well as in life. Working in the Hocky group at NYU, I have not only grown as a scientist but also become a better person.

I am grateful to our collaborators, Eric Vanden-Eijnden and Grant M. Rotskoff, who proposed the original idea where my research was built upon and gave me thought-provoking notes on the proof of the theory. I am also thankful to our collaborators, Stefano Martiniani and Mathias Casiulis, who gave me instructive suggestions on the project. I would also like to show my gratitude to Jonathan Weare, who allowed me to audit his lectures and deepened my understanding of my research.

I owe many thanks to all the members in Hocky group, who have shared enriching moments during my PhD career. They gave me many valuable suggestions not only in my research but also in my life. I am deeply impressed by their shiny points such as kindness, diligence and enthusiasm. I am also thankful to other friendly members in the theory suite.

My heartfelt appreciation goes to my dissertation committee: Mark E. Tuckerman, Yingkai Zhang, Nathaniel J. Traaseth and Stefano Martiniani for their insightful guidance, constructive feedback and instructive suggestions throughout my doctoral studies.

I acknowledge with great gratitude the financial support that made my research possible, including funding from the Department of Energy (DOE), National Institutes of Health (NIH), National Science Foundation (NSF), the Simons Center for Computational Physical Chemistry through a Graduate Fellowship, and the MacCracken Fellowship. I am also grateful to the members in NYU HPC team especially Shenglong Wang for the consistent computing resources and technical assistance.

On my personal note, I am deeply grateful to my parents, Mr. Hua Shi (Anbing Gong) and Mrs. Dongwen Liu, for their unconditional love and support in my life. I would definitely not be able to pursue my PhD study without their effort in return for nothing. I am also thankful to all teachers during my education, especially Huai Sun, who led me to the world of Molecular Dynamic simulation and enhanced sampling techniques, as well as Yuko Okamoto, who greatly inspired in terms of my study in the theory of enhanced sampling techniques.

Finally, I would like to say thank you to all who have played a role in my education and in my life. This dissertation is a reflection of your effort.

Abstract

Many methods to accelerate sampling of molecular configurations are based on the idea that temperature can be used to accelerate rare transitions. These methods typically compute equilibrium properties at a target temperature using reweighting or through Monte Carlo exchanges between replicas at higher temperatures. Our work built on studies showing the potential of invertible non-equilibrium dynamics when estimating equilibrium properties at target equilibrium density. An unbiased non-equilibrium importance sampling (NEIS) estimator with smaller variance compared to equilibrium vanilla estimator was proposed and tested in simple mathematical models. We have shown that the non-equilibrium method can be extended to molecular systems and it can be further improved by combining it with sophisticated enhanced sampling techniques such as umbrella sampling when dealing with free energy surface with multiple minima. We have developed generalized unbiased NEIS estimators with given trajectory data, and have therefore clarified the nature of the unbiased property of NEIS estimators. With the correct way of computing variance of unbiased NEIS estimators, we have shown that there exists an optimal NEIS estimator with given trajectory. Finally, we showed the connection between density propagation method and unbiased NEIS estimators, targeted at estimating the volume of energy basins. We also introduce an "umbrella sampling"-like dynamics that could be potentially useful to compute basin volume in high dimensions.

Contents

List of Figures

1 Introduction

1.1 Molecular Dynamic simulation

In 1957 the technique of molecular dynamics (MD) was first introduced by the pioneering work of B. J. Alder and T. E. Wainwright [1] to study the phase transition for a hard sphere system. Later in 1968 and 1969, refined and consistent force field was proposed by two Nobel-prize-awarded work of Lifson, Warshel and Levitt [2, 3], which built a foundation for further MD simulations. Fast forwarding to 1977, the first MD simulation of protein was performed in the Karplus group [4], where the dynamics of bovine pancreatic trypsin inhibitor were studied by solving the equation of motion under an empirically fit potential energy function.

Thanks to the increasingly powerful computer hardware, especially graphics processing units (GPUs), MD simulations of complex molecular systems become feasible [5, 6]. Over the years, MD simulations have been extensively used in various fields of research [7, 8], including examples from my colleagues at NYU such as understanding the assembly pathways of colloidal system [9, 10], studying large biomolecular complexes [11], discovering the influence of force in biochemical systems [12, 13], studying the mechanism of proton transportation [14], characterizing local environments in different polymorphs of molecular crystals [15], and in designing new drugs [16].

As an analytical solution of equation of motion is not possible for any realistic system, we instead resort to "sampling" possible structures, which allows us to compute average observables

using these sampled data points [17, 18]. MD simulations offer an efficient approach to sampling molecular configurations. In a typical MD simulation, interactions between particles in a system are modeled by force field (representing the potential energy as a function of positions). Trajectories of particles are obtained by iteratively incrementing the solutions to kinematic equations of motion using a finite time step Δt, The accuracy of the numerical solution to equation of motion depends on time step Δt, and more accurate numerical solution are achieved with smaller time step at the cost of more expensive computation cost. Simply solving the equations of classical mechanics will result in generating samples from the so-called microcanonical ensemble, where total energy is constant and each configuration is equally likely [19].

We are most interested in studying processes occurring at constant temperature, where the probability density of seeing a given configuration at equilibrium is given by the Boltzmann distribution $\rho(\boldsymbol{x}) = \exp(-\beta\mathscr{H}(\boldsymbol{x}))/Q(\beta)$. Here $\boldsymbol{x}$ is the coordinate in $6N$-dimensional phase space of the system containing N particles in 3D, $\mathscr{H}$ is the Hamiltonian (potential + kinetic energy) of the system, $\beta = 1/k_{\mathrm{B}}T$, k_{B} is Boltzmann's constant, T is the absolute temperature, and $Q(\beta)$ is the "partition function" which is the integral over all possible configurations that normalizes the distribution. As such, the equations of motion are supplemented by adding a "thermostat" such that energy is effectively exchanged with a constant temperature "bath", keeping the average kinetic energy constant. If sampling is *ergodic*, meaning that all configurations are visited with frequencies consistent with that equilibrium density, then from N samples generated by MD, we can compute the average of an observable O that is approximately equal to its time average:

$$\langle O\rangle = \frac{1}{Q(\beta)}\int_{\Gamma} O(\boldsymbol{x})\exp(-\beta\mathscr{H}(\boldsymbol{x}))\,d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} O(\boldsymbol{x}(t_i)), \tag{1.1}$$

where the integral over Γ represents a sum over all possible phase space points.

At constant temperature, a quantity termed the free energy determines the direction of spontaneous processes, such that the system starting in any particular configuration will evolve such

that the free energy is minimized on average, at which point the system explores the equilibrium conformational ensemble. The free energy of the system at equilibrium can be computed as $F = -k_{\mathrm{B}}T \ln Q$.

For many physical processes of interest, we are interested in the equilibrium difference in free energy between two "states" of the system, where here state is used in the colloquial sense and refers to two sets of configurations of the system that we think are interesting-for example, we could study the free energy difference between a protein bound to a ligand and a protein and ligand separated in solution. The free energy of a state can be defined by calculating the contribution of configurations in that state to the partition function using a characteristic function $\chi_\alpha(\boldsymbol{x})$, with:

$$F_\alpha = -k_{\mathrm{B}}T \ln\left(\frac{1}{Q(\beta)}\int_\Gamma \chi_\alpha(\boldsymbol{x}) \exp\left(-\beta \mathscr{H}(\boldsymbol{x})\right) d\boldsymbol{x}\right) = -k_{\mathrm{B}}T \ln \langle \chi_\alpha \rangle. \tag{1.2}$$

Hence, the free energy difference between between two states at equilibrium can be computed from the negative log of the ratio of how often each state is seen at equilibrium:

$$\Delta F = F_B - F_A = -k_{\mathrm{B}}T \ln\left(\frac{\langle \chi_B \rangle}{\langle \chi_A \rangle}\right). \tag{1.3}$$

A typical molecular system usually contains huge amount of particles with the dimension of phase space being $6N \gg 1$, and hence a collective variable (CV), which maps $6N$-dimensional phase space to the space of particular interest with low dimensions (usually less than 3), is used to capture critical features of microstates along trajectories from MD simulations.

After selecting a CV $\boldsymbol{S}(\boldsymbol{x})$, it is convenient to find the "potential of mean force" (PMF) or the free energy surface (FES) as a function along that CV [17]. $F(\boldsymbol{s};\beta)$ at the coordinate $\boldsymbol{s}$ in CV space serves as an average of all possible microstates that have a CV value equal to $\boldsymbol{s}$ at the given temperature T, and is equivalent to defining the characteristic function above χ_α using a delta

function:

$$\exp\left(-\beta F\left(\boldsymbol{s};\beta\right)\right) \equiv \frac{1}{Q\left(\beta\right)} \int_{\Gamma} \delta\left(\boldsymbol{S}\left(\boldsymbol{x}\right) - \boldsymbol{s}\right) \exp\left(-\beta \mathscr{H}\left(\boldsymbol{x}\right)\right) d\boldsymbol{x}. \tag{1.4}$$

This FES helps give a physically motivated visualization of what the most important configurations are, but can mask hidden states and barriers when marginalizing over too many dimensions. Hence it is important to select CVs that capture the most important degrees of freedom for the process of interest [20–29].

1.2 Enhanced Sampling

Despite the power and wide application of MD simulation, one critical challenge of sampling of rare events limits the efficiency of MD simulations [17, 18]. Simulations will inevitably be trapped in a local basins for some amount of time due to free energy barriers $\left(\beta \Delta F^{\ddagger} \gg 1\right)$, where . For many processes of interest in biomolecular simulation, the transition of interest can be dominated by a sufficiently high barrier from a "reactant" to "product" so that it will effectively never be observed in unbiased simulations at room temperature. Therefore, various enhanced sampling techniques have been proposed and developed over the decades.

Most enhanced sampling techniques follow the idea of importance sampling (IS) [30], i.e. drawing sample points from a starting probability density $\rho_0\left(\boldsymbol{x}\right)$ which is easy to be sampled and giving an unbiased estimation of a test function $\phi\left(\boldsymbol{x}\right)$ of interest at the target probability density $\rho_1\left(\boldsymbol{x}\right)$ by

$$\begin{aligned} \left\langle \phi \right\rangle_1 &\equiv \int_{\Gamma} \phi\left(\boldsymbol{x}\right) \rho_1\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\ &= \int_{\Gamma} \frac{\phi\left(\boldsymbol{x}\right) \rho_1\left(\boldsymbol{x}\right)}{\rho_0\left(\boldsymbol{x}\right)} \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\ &\equiv \left\langle \frac{\phi \rho_1}{\rho_0} \right\rangle_0 . \end{aligned} \tag{1.5}$$

If the starting probability $\rho_0\left(\boldsymbol{x}\right)$ and the target probability density $\rho_1\left(\boldsymbol{x}\right)$ are known up to a normalization constant, then a self-normalized expression is used to give a slightly biased estimation

of the test function $\phi(\boldsymbol{x})$ but with possibly smaller variance [31, 32]:

$$\langle \phi \rangle_1 = \frac{\left\langle \frac{\phi \rho_1}{\rho_0} \right\rangle_0}{\left\langle \frac{\rho_1}{\rho_0} \right\rangle_0}. \tag{1.6}$$

Enhanced sampling techniques can be briefly classified in two categories. (**a**) Collective variable (CV) based method that adds biased potentials to accelerate slow modes of motion recognized by CVs. Some available methods include umbrella sampling (US) [33], metadynamics (metaD) [34], well-tempered metadynamics (WTmetaD) [35], on the fly probability enhanced sampling (OPES) [36], adiabatic free energy dynamics (AFED) [37, 38], temperature accelerated molecular dynamics (TAMD) [39], infinite switch simulated tempering in force (FISST) [40]. Obviously the efficiency of such kind of enhanced sampling methods largely depends on how good the chosen CVs can capture the slow modes of motion, which is another challenging topic. (**b**) Temperature based method that utilizes high mobility at high temperature to overcome large free energy barriers. Some typical methods are replica exchange molecular dynamics (REMD) [41], replica exchange with solute tempering (REST) [42, 43].

1.2.1 Umbrella Sampling

Umbrella sampling performs localized simulations around some CV values $\{\boldsymbol{s}_i\}_{i=1,\ldots,K}$ respectively by adding a biased harmonic potential centered at a particular CV coordinate $\boldsymbol{s}_i$ in the ith biased simulation.

$$U_i(\boldsymbol{x}) = U(\boldsymbol{x}) + \frac{1}{2} k_i \left| \boldsymbol{S}(\boldsymbol{x}) - \boldsymbol{s}_i \right|^2, \tag{1.7}$$

where $U(\boldsymbol{x})$ is the unbiased potential energy of interest, $U_i(\boldsymbol{x})$ is the biased potential energy in the biased ith simulation, k_i is the spring constant in the ith biased simulation that controls the spread of biased sampling, $\boldsymbol{S}(\boldsymbol{x})$ is the defined CV that maps position $\boldsymbol{x}$ to CV space $\boldsymbol{s}$ and the dimension of CV space is usually 1 or 2 in practice.

1.2.2 Reweighting Techniques

When k_i is large, free energy in CV space around $\boldsymbol{s}_i$ in the ith biased simulation is approximately harmonic, and sampling around $\boldsymbol{s}_i$ in CV space is efficient. Then biased free energy surface can be estimated locally in each biased simulation. Some reweighting techniques [44–48] are used to combine these local biased data to obtain the overall unbiased free energy surface.

1.2.2.1 Weighted Histogram Analysis Method (WHAM)

One common reweighting technique is the so-called Weighted Histogram Analysis Method (WHAM) [44, 49, 50]. The idea of WHAM is to maximize the overall probability of sampling the specific trajectories.

Firstly, the biased probability density at the ith biased simulation in CV space $\rho_i^b(\boldsymbol{s})$ is assumed to have the form:

$$\rho_i^b(\boldsymbol{s}) = c_i \rho^u(\boldsymbol{s}) \exp(-\beta_0 w_i(\boldsymbol{s})), \tag{1.8}$$

where $\rho^u(\boldsymbol{s})$ is the unbiased probability density in CV space, $w_i(\boldsymbol{s}) = \frac{1}{2} k_i |\boldsymbol{s} - \boldsymbol{s}_i|^2$ is the biased potential added in the ith biased simulation, and $c_i = \left(\int_\Gamma \rho^u(\boldsymbol{s}) \exp(-\beta_0 w_i(\boldsymbol{s})) \, d\boldsymbol{s} \right)^{-1}$ is the normalization factor of the biased probability density.

Secondly, the CV space is discretized uniformly for simplicity and analytical biased probability densities reduce to averaged probabilities over these windows:

$$\rho_{i,k}^b \equiv \rho_i^b(\boldsymbol{s}_k) = c_i \rho^u(\boldsymbol{s}_k) \exp(-\beta_0 w_i(\boldsymbol{s}_k)) \equiv c_i \rho_k^u \exp(-\beta_0 w_{i,k}), \tag{1.9}$$

where $c_i \propto \left(\sum_k \rho_k^u \exp(-\beta_0 w_{i,k}) \right)^{-1}$ is the numerical normalization factor over the windows in the ith biased simulation.

Then the overall probability of obtaining $n_{i,k}$ samples within the window k in the ith biased

simulation can be expressed as follows.

$$\mathbb{P}\left(\left\{n_{i,k}\right\}_{i,k} \mid \left\{\rho_k^u\right\}_k\right) = \frac{\left(\sum_{i,k} n_{i,k}\right)!}{\prod_{i,k}\left(n_{i,k}!\right)} \prod_{i,k}\left(\rho_{i,k}^b\right)^{n_{i,k}} = \frac{\left(\sum_{i,k} n_{i,k}\right)!}{\prod_{i,k}\left(n_{i,k}!\right)} \prod_{i,k}\left(c_i \rho_k^u \exp\left(-\beta_0 w_{i,k}\right)\right)^{n_{i,k}} \quad (1.10)$$

Optimizing the overall probability $\mathbb{P}\left(\left\{n_{i,k}\right\}_{i,k} \mid \left\{\rho_k^u\right\}_k\right)$ with respect to parameters $\left\{\rho_k^u\right\}_k$ is the same as optimizing the logarithm of the overall probability $\ln \mathbb{P}\left(\left\{n_{i,k}\right\}_{i,k} \mid \left\{\rho_k^u\right\}_k\right)$:

$$\begin{aligned} \frac{\partial}{\partial \rho_k^u} \ln \mathbb{P} &= \sum_i \sum_j n_{i,j} \frac{\partial \ln c_i}{\partial \rho_k^u} + \sum_i n_{i,k} \frac{1}{\rho_k^u} \\ &= -\sum_i N_i c_i \exp\left(-\beta_0 w_{i,k}\right) + \frac{N_k}{\rho_k^u} = 0, \end{aligned} \quad (1.11)$$

where $N_i = \sum_k n_{i,k}$ is the total number of sampling points in the ith biased simulation and $N_k = \sum_k n_{i,k}$ is the total number of sampling points within the kth window among all the biased simulations. This leads to the so-called WHAM equations:

$$\begin{cases} \rho_k^u = \dfrac{N_k}{\sum_i N_i c_i \exp\left(-\beta_0 w_{i,k}\right)} \\ c_i^{-1} = \displaystyle\sum_k \rho_k^u \exp\left(-\beta_0 w_{i,k}\right) \end{cases} \quad (1.12)$$

The WHAM equations can be solved self-consistently:

$$\begin{cases} \rho_k^{(n+1)} = \dfrac{N_k}{\sum_i N_i c_i^{(n)} \exp\left(-\beta_0 w_{i,k}\right)} \\ c_i^{(n)} = \left(\displaystyle\sum_k \rho_k^{(n)} \exp\left(-\beta_0 w_{i,k}\right)\right)^{-1}, \end{cases} \quad (1.13)$$

where $\rho_k^{(n)}, c_i^{(n)}$ are the values of updated unbiased probability at kth window and the corresponding normalization factor in the ith biased simulation in the nth iteration respectively. $\left\{\rho_k^{(0)}\right\}_k$ are initialized as all ones as a fair initial guess and iteration steps until it meets some termination

conditions such as change of all estimated unbiased probabilities is within some tolerance.

1.2.2.2 Eigenvector Method for Umbrella Sampling (EMUS)

Another method with novel mathematical results is so-called eigenvector method for umbrella sampling (EMUS) [46]. The idea of EMUS is establishing connections between biased results and unbiased results through some mathematical derivations and recognizing an eigenvector problem regarding the "overlap" matrices.

Firstly, several quantities are defined to simplify following derivations:

$$\begin{cases} \rho_0\left(\boldsymbol{x}\right) = \dfrac{\exp\left(-\beta_0 \mathscr{H}\left(\boldsymbol{x}\right)\right)}{\int_\Gamma \exp\left(-\beta_0 \mathscr{H}\left(\boldsymbol{x}\right)\right) d\boldsymbol{x}} \\ \psi_i\left(\boldsymbol{x}\right) = \exp\left(-\beta_0 \dfrac{1}{2} k_i \left|\boldsymbol{S}\left(\boldsymbol{x}\right) - \boldsymbol{s}_i\right|^2\right) \\ \rho_i\left(\boldsymbol{x}\right) = \dfrac{\rho_0\left(\boldsymbol{x}\right)\psi_i\left(\boldsymbol{x}\right)}{\int_\Gamma \rho_0\left(\boldsymbol{x}\right)\psi_i\left(\boldsymbol{x}\right)\, d\boldsymbol{x}}, \end{cases} \tag{1.14}$$

where $\rho_0\left(\boldsymbol{x}\right)$ is the unbiased probability density, $\psi_i\left(\boldsymbol{x}\right)$ is the biased factor introduced by biased harmonic potential centered at the CV $\boldsymbol{s}_i$ in the ith biased simulation, and $\rho_i\left(\boldsymbol{x}\right)$ is the biased probability density in the $i$the biased simulation.

Secondly, a connection between mean value $\left\langle\phi\right\rangle_0$ of a test function $\phi\left(\boldsymbol{x}\right)$ under unbiased probability density $\rho_0\left(\boldsymbol{x}\right)$ and mean value $\left\langle\phi\right\rangle_i$ of the test function under biased probability $\rho_i\left(\boldsymbol{x}\right)$

in the ith biased simulation can be established by mathematical derivations:

$$\begin{aligned}
\langle\phi\rangle_0 &\equiv \int_\Gamma \phi(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \\
&= \int_\Gamma \phi(\boldsymbol{x})\,\frac{\sum_i \psi_i(\boldsymbol{x})}{\sum_k \psi_k(\boldsymbol{x})}\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \\
&= \sum_i \left\{ \int_\Gamma \phi^*(\boldsymbol{x})\,\psi_i(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\} \\
&= \sum_i \left\{ \frac{\int_\Gamma \phi^*(\boldsymbol{x})\,\psi_i(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x}}{\int_\Gamma \psi_i(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x}} \int_\Gamma \psi_i(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\} \\
&= \sum_i \left\{ \langle\phi^*\rangle_i \int_\Gamma \psi_i(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\} \\
&= \sum_i \left\{ \langle\phi^*\rangle_i\, z_i \right\} \sum_j \left\{ \int_\Gamma \psi_j(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\},
\end{aligned} \tag{1.15}$$

where $\phi^* \equiv \frac{\phi}{\sum_i \psi_i}$ and $z_i \equiv \frac{\int_\Gamma \psi_i(\boldsymbol{x})\rho_0(\boldsymbol{x})\,d\boldsymbol{x}}{\sum_j \left\{\int_\Gamma \psi_j(\boldsymbol{x})\rho_0(\boldsymbol{x})\,d\boldsymbol{x}\right\}}$ with obviously $\sum_i z_i = 1$. To cancel out the latter term in the last line, let $\phi(\boldsymbol{x}) = 1$ in Eq. 1.15, and we get

$$1 = \langle 1\rangle_0 = \sum_i \left\{ \langle 1^*\rangle_i\, z_i \right\} \sum_j \left\{ \int_\Gamma \psi_j(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\} \tag{1.16}$$

Combining Eq. 1.15 and Eq. 1.16, we have the key expression in EMUS:

$$\langle\phi\rangle_0 = \frac{\sum_i z_i \langle\phi^*\rangle_i}{\sum_i z_i \langle 1^*\rangle_i} \tag{1.17}$$

Next step we express the constants $\{z_i\}_i$ by taking $\phi(\boldsymbol{x}) = \psi_j(\boldsymbol{x})$ in Eq. 1.15:

$$\begin{aligned}
\left\langle\psi_j\right\rangle_0 &\equiv \int_\Gamma \psi_j(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \\
&\equiv z_j \sum_k \left\{ \int_\Gamma \psi_k(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\} \\
&= \sum_i \left\{ \left\langle\psi_j^*\right\rangle_i z_i \right\} \sum_k \left\{ \int_\Gamma \psi_k(\boldsymbol{x})\,\rho_0(\boldsymbol{x})\,d\boldsymbol{x} \right\},
\end{aligned}$$

which gives

$$z_j = \sum_i z_i F_{ij}, \tag{1.18}$$

where $F_{ij} \equiv \left\langle \psi_j^* \right\rangle_i$ is the overlap matrix.

Eq. 1.18 suggests that the vector of normalization constants $\boldsymbol{z}$ is the left eigenvector of the overlap matrix $\mathbf{F}$ with eigenvalue 1. If there is a sufficient overlap between any of two biased simulations, there is a unique solution to this eigenvector problem since the overlap matrix $\mathbf{F}$ is a stochastic matrix $\sum_j F_{ij} = 1$ by definition and a stochastic matrix $\mathbf{J}$ has a unique eigenvector with eigenvalue one if it is irreducible: for any possible grouping of indices into two distinct sets A and B, there exists some $i \in A$ and $j \in B$, such that $J_{ij} \neq 0$.

In practice, the entries of the overlap matrix F_{ij} and the averages $\langle \phi^* \rangle_i$ and $\langle 1^* \rangle_i$ are estimated by the time averages:

$$\begin{cases} \bar{\phi}^* = \dfrac{1}{T} \displaystyle\sum_{t=0}^{T-1} \dfrac{\phi\left(\boldsymbol{X}_t^i\right)}{\sum_k \psi_k\left(\boldsymbol{X}_t^i\right)} \\ \bar{1}^* = \dfrac{1}{T} \displaystyle\sum_{t=0}^{T-1} \dfrac{1}{\sum_k \psi_k\left(\boldsymbol{X}_t^i\right)} \\ \bar{F}_{ij} = \dfrac{1}{T} \displaystyle\sum_{t=0}^{T-1} \dfrac{\psi_j\left(\boldsymbol{X}_t^i\right)}{\sum_k \psi_k\left(\boldsymbol{X}_t^i\right)}, \end{cases} \tag{1.19}$$

where $\boldsymbol{X}_t^i$ is the coordinate of the sample point at time t in the ith biased simulation. Then the vector of the estimated normalization constants $\boldsymbol{z}^{\mathrm{EMUS}}$ is the numerical solution to $z_j^{\mathrm{EMUS}} = \sum_i z_i^{\mathrm{EMUS}} \bar{F}_{ij}$ with $\sum_i z_i^{\mathrm{EMUS}} = 1$ using QR factorization as in Golub and Meyer [51]. If the samples $\boldsymbol{X}_t^i$ are independent, multistate Bennett acceptance ratio estimator (MBAR) [45] is the nonparametric maximum-likelihood estimator of $\boldsymbol{z}$ [52].

1.3 Nonequilibrium Importance Sampling (NEIS) estimator

Statistical estimation through the process of equilibrium dynamics is based on the principle of detailed balance. Given an invertible dynamics

$$\dot{\boldsymbol{X}}\left(t,\boldsymbol{x}\right)=\boldsymbol{b}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right), \tag{1.20}$$

where $\boldsymbol{X}\left(0,\boldsymbol{x}\right)=\boldsymbol{x}\in\Gamma\subset\mathbb{R}^{d}$ is a microstate propagated in time to $\boldsymbol{X}\left(t,\boldsymbol{x}\right)$ through the vector field $\boldsymbol{b}$. Since the dynamical process is invertible, we have

$$\boldsymbol{X}\left(s,\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)=\boldsymbol{X}\left(s+t,\boldsymbol{x}\right),\forall s,t\in\mathbb{R}. \tag{1.21}$$

An equilibrium dynamics means that some probability density $\rho\left(\boldsymbol{x}\right)$ is preserved under time evolution of such equilibrium dynamics. Then the expectation value of an arbitrary test function $\phi\left(\boldsymbol{x}\right)$ under the target probability density $\rho\left(\boldsymbol{x}\right)$, denoted by $\langle\phi\rangle$, can be estimated by a time average along the equilibrium trajectory if the process is ergodic [17]. Converging this estimate is difficult due to the challenge of observing rare events, due to the fact that the expectation $\langle\phi\rangle$ is dominated at the states $\boldsymbol{x}$ which are rare under the probability density $\rho\left(\boldsymbol{x}\right)$ and are infrequently visited by the equilibrium dynamics.

Several importance sampling techniques [53–58] based on nonequilibrium dynamics, which are potentially more versatile than equilibrium ones, have been developed and shown success over the decades.

One well-known method is the annealed importance sampling (AIS) [59], which gradually propagates a starting probability density that is easy to sample (e.g. at high temperature) to the target probability density that is hard to sample (e.g. at low temperature). Let $\rho_{0}\left(\boldsymbol{x}\right)=\exp\left(-U_{0}\right)$ be the appropriately chosen starting probability density with a reduced energy $U_{0}\left(\boldsymbol{x}\right)$

and $\rho(\boldsymbol{x}) = \exp(-U(\boldsymbol{x}))$ be the target probability density with a reduced energy $U(\boldsymbol{x})$. Then $L-1$ intermediate reduced energies $\{U_l(\boldsymbol{x})\}_{l=1}^{L-1}$ are inserted between $U_0(\boldsymbol{x})$ and $U(\boldsymbol{x})$ and the l-th intermediate reduced energy is defined as

$$U_l(\boldsymbol{x}) \equiv \left(1 - c\left(\frac{l}{L}\right)\right) U_0(\boldsymbol{x}) + c\left(\frac{l}{L}\right) U(\boldsymbol{x}), \forall l = 1, \ldots, L-1, \tag{1.22}$$

where $c : [0,1] \rightarrow [0,1]$ is a strictly monotonically increasing function with $c(0) = 0$ and $c(1) = 1$. Let $T_l(\boldsymbol{x}, \boldsymbol{y})$ be the transition kernel at the intermediate probability density $\rho_l(\boldsymbol{x})$ that satisfies detailed balance condition:

$$\rho_l(\boldsymbol{x}) T_l(\boldsymbol{x}, \boldsymbol{y}) = \rho_l(\boldsymbol{y}) T_l(\boldsymbol{y}, \boldsymbol{x}). \tag{1.23}$$

Then a sequence of sample points $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}\}$ are generated as follows. The initial sample point $\boldsymbol{x}_0$ is drawn from the starting probability density $\rho_0(\boldsymbol{x})$. Then for $l = 1, \ldots, L-1$, the sample point $\boldsymbol{x}_l$ is drawn after the transition kernel $T_l(\boldsymbol{x}_{l-1}, \boldsymbol{x}_l)$. The corresponding weight w of the sequence $\{\boldsymbol{x}, \ldots, \boldsymbol{x}_{L-1}\}$ is:

$$w = \frac{\rho_1(\boldsymbol{x}_0)}{\rho_0(\boldsymbol{x}_0)} \cdots \frac{\rho(\boldsymbol{x}_{L-1})}{\rho_{L-1}(\boldsymbol{x}_{L-1})} \propto \frac{\exp(-U_1(\boldsymbol{x}_0))}{\exp(-U_0(\boldsymbol{x}_0))} \cdots \frac{\exp(-U(\boldsymbol{x}_{L-1}))}{\exp(-U_{L-1}(\boldsymbol{x}_{L-1}))}, \tag{1.24}$$

which satisfies detailed balance condition by construction:

$$\rho_0(\boldsymbol{x}_0) T_1(\boldsymbol{x}_0, \boldsymbol{x}_1) \cdots T_{L-1}(\boldsymbol{x}_{L-2}, \boldsymbol{x}_{L-1}) w = \rho(\boldsymbol{x}_{L-1}) T_{L-1}(\boldsymbol{x}_{L-1}, \boldsymbol{x}_{L-2}) \cdots T_1(\boldsymbol{x}_1, \boldsymbol{x}_0), \tag{1.25}$$

which means that the probability density of the sequence $\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}\}$ with the weight w is the same as the probability density of the reverse sequence $\{\boldsymbol{x}_{L-1}, \ldots, \boldsymbol{x}_0\}$. Therefore an estimation of a test function $\phi(\boldsymbol{x})$ can be made from N trajectories from N independent starting points

$\left\{\boldsymbol{x}_0^{(i)}\right\}_{i=1}^N$ with the corresponding weights $\left\{w^{(i)}\right\}_{i=1}^N$ by

$$\bar{\phi} = \frac{\sum_{i=1}^N \phi\left(\boldsymbol{x}_{L-1}^{(i)}\right) w^{(i)}}{\sum_{i=1}^N w^{(i)}}. \tag{1.26}$$

When $N \to \infty$, the estimator $\bar{\phi}$ converges to the ideal expectation $\langle\phi\rangle$.

Our work in Chapter 2 follows from a related idea derived by Rotskoff and Vanden-Eijnden [60], who proposed a class of unbiased nonequilibrium importance sampling (NEIS) estimators; those methods use carefully chosen dynamics such that the probability density is transported to the rare regions of phase space and the corresponding statistical weights can be directly obtained through the invertible nonequilibrium dynamics. This differs from AIS, which requires computation of a ratio of sample means. Moreover, the unbiased nonequilibrium estimator always has a smaller variance compared to the direct sampling in equilibrium dynamics at the nontrivial cost of generating nonequilibrium trajectories.

1.3.1 Derivation of NEIS estimator

The unbiased NEIS estimator is derived as follows. Firstly, by the idea of importance sampling, the expectation value of a test function $\phi(\boldsymbol{x})$ under a target probability density $\rho(\boldsymbol{x})$ can be computed by the average of samples drawn from another probability density $\rho_{\text{ne}}(\boldsymbol{x})$ with some reweighting factors.

$$\langle\phi\rangle \equiv \int \phi(\boldsymbol{x})\,\rho(\boldsymbol{x})\; d\boldsymbol{x} = \int \frac{\phi(\boldsymbol{x})\,\rho(\boldsymbol{x})}{\rho_{\text{ne}}(\boldsymbol{x})}\rho_{\text{ne}}(\boldsymbol{x})\; d\boldsymbol{x} \equiv \left\langle\frac{\phi\rho}{\rho_{\text{ne}}}\right\rangle_{\text{ne}} \tag{1.27}$$

Samples are drawn under nonequilibrium stationary density $\rho_{\text{ne}}(\boldsymbol{x})$ by the initiate-then-propagate algorithm: starting points $\boldsymbol{x}$ are drawn under starting probability density $\rho(\boldsymbol{x})$ and then each starting point is propagated under dynamics (Eq. 1.20) forward and backward in time until the trajectory hits some appropriate fixed target sets such as the boundary of Γ or the whole

phase space $\mathbb{R}^d$. With this algorithm, the nonequilibrium average can be expressed as

$$\langle \phi \rangle_{\text{ne}} = \frac{1}{\langle \tau \rangle} \int_{\Gamma} \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \, dt \rho\left(\boldsymbol{x}\right) \, d\boldsymbol{x}, \tag{1.28}$$

where $\tau^+\left(\boldsymbol{x}\right) \geqslant 0$ and $\tau^-\left(\boldsymbol{x}\right) \leqslant 0$ are the first time when the trajectory starting at $\boldsymbol{x}$ hits the boundary $\partial\Gamma$ forward or backward in time respectively. $\langle \tau \rangle = \langle \tau^+ \rangle - \langle \tau^- \rangle$ serves as the normalization factor of the stationary nonequilibrium probability density by letting $\phi\left(\boldsymbol{x}\right) = 1$.

Separately, the nonequilibrium average can be defined as

$$\langle \phi \rangle_{\text{ne}} \equiv \int_{\Gamma} \phi\left(\boldsymbol{x}\right) \rho_{\text{ne}}\left(\boldsymbol{x}\right) \, d\boldsymbol{x}. \tag{1.29}$$

By changing variable from $\boldsymbol{X}\left(t, \boldsymbol{x}\right) \longrightarrow \boldsymbol{x}$ and $t \longrightarrow -t$, Eq. 1.28 becomes

$$\begin{aligned} \langle \phi \rangle_{\text{ne}} &= \frac{1}{\langle \tau \rangle} \int_{\Gamma} \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \, dt \rho\left(\boldsymbol{x}\right) \, d\boldsymbol{x} \\ &= \int_{\Gamma} \phi\left(\boldsymbol{x}\right) \frac{1}{\langle \tau \rangle} \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right) \, dt \, d\boldsymbol{x}. \end{aligned} \tag{1.30}$$

Then the stationary probability density $\rho_{\text{ne}}\left(\boldsymbol{x}\right)$ is recognized as

$$\rho_{\text{ne}}\left(\boldsymbol{x}\right) = \frac{1}{\langle \tau \rangle} \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right) \, dt, \tag{1.31}$$

where $J\left(t, \boldsymbol{x}\right)$ is the Jacobian of the variable transformation

$$J\left(t, \boldsymbol{x}\right) \equiv \exp\left(\int_0^t \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(s, \boldsymbol{x}\right)\right) \, ds \right), \tag{1.32}$$

for which a detailed derivation can be found in Sec. 2.8.1. The physical interpretation of the Jacobian is the factor that describes changes of phase space volume over time t through the invertible dynamics (Eq. 1.20).

A similar result was also found by Tuckerman and coworkers [61, 62], where the Jacobian was understood as the determinant of the transformation matrix from the starting coordinate $\boldsymbol{x}_0$ to the coordinate $\boldsymbol{x}_t$ after time t under the flow $\boldsymbol{b}$. Then an invariant measure of the phase space volume under the time-independent vector field $\boldsymbol{b}$ can be established as

$$\exp\left(-w\left(\boldsymbol{x}_t, t\right)\right) d\boldsymbol{x}_t = \exp\left(-w\left(\boldsymbol{x}_0, 0\right)\right) d\boldsymbol{x}_0, \tag{1.33}$$

where $w\left(\boldsymbol{x}, t\right)$ is the indefinite time integral of the compressibility $\boldsymbol{\nabla} \cdot \boldsymbol{b}$ of a non-Hamiltonian system. Here $\sqrt{g\left(\boldsymbol{x}_t, t\right)} \equiv \exp\left(-w\left(\boldsymbol{x}_t, t\right)\right)$ can be understood as the determinant of the metric tensor $G\left(\boldsymbol{x}_t, t\right)$ obtained from $G\left(\boldsymbol{x}_0, 0\right)$ via the coordinate transformation $\boldsymbol{x}_0 \rightarrow \boldsymbol{x}_t$ and $G\left(\boldsymbol{x}_0, 0\right)$ is the metric tensor that describes the geometry of the phase space.

One important property from the definition of the Jacobian is

$$\begin{aligned}
J\left(s+t, \boldsymbol{x}\right) &\equiv \exp\left(\int_0^{s+t} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{x}\right)\right) \, du\right) \\
&= \exp\left(\int_0^{t} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{x}\right)\right) \, du\right) \exp\left(\int_t^{s+t} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{x}\right)\right) \, du\right) \\
&= \exp\left(\int_0^{t} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{x}\right)\right) \, du\right) \exp\left(\int_0^{s} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u+t, \boldsymbol{x}\right)\right) \, du\right) \\
&= \exp\left(\int_0^{t} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{x}\right)\right) \, du\right) \exp\left(\int_0^{s} \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(u, \boldsymbol{X}\left(t, \boldsymbol{x}\right)\right)\right) \, du\right) \\
&\equiv J\left(t, \boldsymbol{x}\right) J\left(s, \boldsymbol{X}\left(t, \boldsymbol{x}\right)\right).
\end{aligned} \tag{1.34}$$

With the expression of the stationary nonequilibrium probability density (Eq. 1.31), Eq. 1.27

becomes

$$
\begin{aligned}
\langle \phi \rangle &= \left\langle \frac{\phi \rho}{\rho_{\mathrm{ne}}} \right\rangle_{\mathrm{ne}} \\
&= \frac{1}{\langle \tau \rangle} \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right)}{\rho_{\mathrm{ne}}\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right)} dt \rho\left(\boldsymbol{x}\right) \, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right)}{\int_{\tau^{-}(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^{+}(\boldsymbol{X}(t,\boldsymbol{x}))} \rho\left(\boldsymbol{X}\left(s, \boldsymbol{X}\left(t, \boldsymbol{x}\right)\right)\right) J\left(s, \boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \, ds} dt \rho\left(\boldsymbol{x}\right) \, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho\left(\boldsymbol{X}\left(s, \boldsymbol{x}\right)\right) J\left(s, \boldsymbol{x}\right) \, ds} dt \rho\left(\boldsymbol{x}\right) \, d\boldsymbol{x} \\
&= \left\langle \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho\left(\boldsymbol{X}\left(s, \boldsymbol{x}\right)\right) J\left(s, \boldsymbol{x}\right) \, ds} dt \right\rangle .
\end{aligned}
\tag{1.35}
$$

With the equality in Eq. 1.35, we can define the so-called unbiased NEIS estimator as in Eq. 1.36. For each starting points $\boldsymbol{x}$ drawn under target probability density $\rho\left(\boldsymbol{x}\right)$, the estimation of a test function $\phi\left(\boldsymbol{x}\right)$ is computed based on the trajectory generated by $\boldsymbol{x}$ under an invertible nonequilibrium dynamics (Eq. 1.20). This unbiased nonequilibrium estimator is unbiased by the construction itself and is valid for any target probability density $\rho\left(\boldsymbol{x}\right)$ and any invertible dynamics (Eq. 1.20). Moreover, it is self-normalized and it only requires knowledge of the starting density $\rho\left(\boldsymbol{x}\right)$ up to a normalization constant.

$$
\phi_{\mathrm{neq}}\left(\boldsymbol{x}\right) \equiv \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho\left(\boldsymbol{X}\left(s, \boldsymbol{x}\right)\right) J\left(s, \boldsymbol{x}\right) \, ds} dt \tag{1.36}
$$

Furthermore, the equality in Eq. 1.35 establishes a connection between the expectation value of a test function $\phi\left(\boldsymbol{x}\right)$ under equilibrium probability density $\rho\left(\boldsymbol{x}\right)$ and the expectation value of a test function $\phi\left(\boldsymbol{x}\right)$ under a nonequilibrium process. Then it can be proved that the unbiased NEIS estimator (Eq. 1.36) has a smaller variance than that of the direct estimator by Jensen's

inequality [63].

$$
\begin{aligned}
\operatorname{Var}\left[\phi_{\mathrm{neq}}\right] &\equiv \left\langle\left|\phi_{\mathrm{neq}}\right|^2\right\rangle - \left\langle\phi_{\mathrm{neq}}\right\rangle^2 = \left\langle\left|\phi_{\mathrm{neq}}\right|^2\right\rangle - \langle\phi\rangle^2 \\
&= \left\langle\left|\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})}\rho\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\,dt\right|^2\right\rangle - \langle\phi\rangle^2 \\
&\leqslant \left\langle\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi^2\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})}\rho\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\,dt\right\rangle - \langle\phi\rangle^2 \\
&= \left\langle\phi^2\right\rangle - \langle\phi\rangle^2 \equiv \operatorname{Var}\left[\phi\right]
\end{aligned}
\tag{1.37}
$$

Note that during the proof in Eq. 1.37, we take it as granted that the variance of this unbiased NEIS estimator is defined as in the first line of Eq. 1.37. However, this is not always true as we will see in Ch. 3, this trivial and naive definition of variance will lead to some unphysical results when we generalize unbiased NEIS estimators.

In practice, an estimation is made from trajectories generated by starting points $\{\boldsymbol{x}_i\}_{i=1}^N$ drawn under target probability density $\rho\left(\boldsymbol{x}\right)$:

$$
\langle\phi\rangle \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{\tau^-(\boldsymbol{x}_i)}^{\tau^+(\boldsymbol{x}_i)}\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}_i\right)\right)\rho\left(\boldsymbol{X}\left(t,\boldsymbol{x}_i\right)\right)J\left(t,\boldsymbol{x}_i\right)\,dt}{\int_{\tau^-(\boldsymbol{x}_i)}^{\tau^+(\boldsymbol{x}_i)}\rho\left(\boldsymbol{X}\left(s,\boldsymbol{x}_i\right)\right)J\left(s,\boldsymbol{x}_i\right)\,ds},
\tag{1.38}
$$

and the estimation is exact when $N \to \infty$. Since the processes of generating trajectory from the starting point $\boldsymbol{x}_i$ and the nonequilibrium estimation $\phi_{\mathrm{neq}}\left(\boldsymbol{x}_i\right)$ at the starting point $\boldsymbol{x}_i$ are independent, these processes can be done in parallel on high performance computing resources to decrease time to solution.

As described, it is not stated what dynamics would work well for this estimator, nor which would easily permit the calculation of $J\left(t,\boldsymbol{x}_i\right)$. The idea from Ref. [60] is to construct a dynamics that takes one from an easy to sample high-temperature distribution and then descends to the regions of phase space one cares about. They proposed what we term a "quench" dynamics, which

for $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}^{2d}$, is given by:

$$\begin{cases} \dot{\boldsymbol{q}} = \mathbf{M}^{-1}\boldsymbol{p} \\ \dot{\boldsymbol{p}} = -\boldsymbol{\nabla} U(\boldsymbol{q}) - \gamma \boldsymbol{p} \end{cases}, \tag{1.39}$$

where $\boldsymbol{q} \in \mathbb{R}^d$ are the positions of system, $\boldsymbol{p} \in \mathbb{R}^d$ are the momenta of system, $\mathbf{M}$ is the diagonal mass matrix, $\dot{}$ is the time derivative, $\boldsymbol{\nabla} \equiv \left[\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_d}\right]^T$ is the position gradient and $\gamma \in \mathbb{R}$ is the friction constant.

There are several advantages of using "quench" dynamics (Eq. 1.39).

- The functional form of the "quench" dynamics is simple, which is just Langevin dynamics without random force term. Hence it is easy to implement.

- The corresponding Jacobian of the "quench" dynamics is simple and does not depend on the starting coordinate $\boldsymbol{x}$.

$$J(t, \boldsymbol{x}) = \exp(-d\gamma t), \tag{1.40}$$

  where d is the degrees of freedom of system.

- The "quench" dynamics is dissipative. Compared to equilibrium dynamics, high energy region and low energy region is more likely to be sampled under "quench" dynamics in backward time or forward time respectively.

$$\dot{E} = \boldsymbol{\nabla} U(\boldsymbol{q}) \cdot \dot{\boldsymbol{q}} + \mathbf{M}^{-1}\boldsymbol{p} \cdot \dot{\boldsymbol{p}} = -\gamma |\boldsymbol{p}|^2 \leqslant 0 \tag{1.41}$$

- When starting points $\boldsymbol{x}$ are drawn uniformly and let $\Gamma = \{\boldsymbol{x} : \mathscr{H}(\boldsymbol{x}) \leqslant E_{\max} < +\infty\}$. Then volume of phase space $V(E)$ can be estimated up to a constant.

$$\frac{V(E)}{V(E_{\max})} \approx \frac{1}{N} \sum_{i=1}^{N} \exp\left(-d\gamma \left(\tau^E(\boldsymbol{x}_i) - \tau^-(\boldsymbol{x}_i)\right)\right), \tag{1.42}$$

where $\tau^E(\boldsymbol{x})$ is the first time when total energy of system $\mathscr{H}\left(\boldsymbol{X}\left(\tau^E(\boldsymbol{x}), \boldsymbol{x}\right)\right) = E$ starting at $\boldsymbol{x}$ and evolved under "quench" dynamics (Eq. 1.39). Moreover, $\tau^E(\boldsymbol{x})$ is unique by intermediate value theorem [64] since the "quench" dynamics is dissipative.

- The "quench" dynamics can be generalized to quench a subset of interest among the whole phase space. For instance, a solute-"quench" dynamics, which only quenches the solute part of the system, is given by:

$$\begin{cases} \dot{\boldsymbol{q}}_{\text{solute}} = \mathbf{M}_{\text{solute}}^{-1}\boldsymbol{p}_{\text{solute}} \\ \dot{\boldsymbol{p}}_{\text{solute}} = -\boldsymbol{\nabla}_{\text{solute}}U(\boldsymbol{q}) - \gamma\boldsymbol{p}_{\text{solute}} \\ \dot{\boldsymbol{q}}_{\text{solvent}} = \mathbf{M}_{\text{solvent}}^{-1}\boldsymbol{p}_{\text{solvent}} \\ \dot{\boldsymbol{p}}_{\text{solvent}} = -\boldsymbol{\nabla}_{\text{solvent}}U(\boldsymbol{q}) \end{cases}, \tag{1.43}$$

and the corresponding Jacobian $J(t, \boldsymbol{x}) = \exp\left(-d_{\text{solute}}\gamma t\right)$, where $d_{\text{solute}} \ll d$ is the degree of freedom of the solute part.

1.3.2 Classification of NEIS estimator

Following work from Cao and Vanden-Eijnden [65], we will classify unbiased NEIS estimators into two groups based on integral limit with respect to time. As shown in Eq. 1.44, when integral limits are two finite values (not necessarily positive or negative), the NEIS estimators are termed *finite time* nonequilibrium importance sampling (FTNEIS) estimators, and when integral limits are negative infinity and positive infinity, the NEIS estimators are termed *infinite time* nonequilibrium importance sampling (ITNEIS) estimators.

$$\begin{cases} \phi_{\text{neq}}^{\text{FTNEIS}}(\boldsymbol{x}) = \displaystyle\int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}(t,\boldsymbol{x})\right)\rho\left(\boldsymbol{X}(t,\boldsymbol{x})\right)J(t,\boldsymbol{x})}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}(s,\boldsymbol{x})\right)J(s,\boldsymbol{x})\,ds}\,dt \\ \phi_{\text{neq}}^{\text{ITNEIS}}(\boldsymbol{x}) = \displaystyle\int_{-\infty}^{+\infty} \frac{\phi\left(\boldsymbol{X}(t,\boldsymbol{x})\right)\rho\left(\boldsymbol{X}(t,\boldsymbol{x})\right)J(t,\boldsymbol{x})}{\int_{-\infty}^{+\infty}\rho_0\left(\boldsymbol{X}(s,\boldsymbol{x})\right)J(s,\boldsymbol{x})\,ds}\,dt \end{cases}, \tag{1.44}$$

where $\rho(\boldsymbol{x})$ is the target probability density and $\rho_0(\boldsymbol{x})$ is the starting probability density from which starting points $\boldsymbol{x}$ are drawn. Note that in the original NEIS estimator (Eq. 1.36), target probability should be the same as starting probability and an importance sampling (IS)-like estimator $\phi^{\mathrm{IS}} \equiv \frac{\phi\rho_1}{\rho_0}$ is used to estimate a test function $\phi(\boldsymbol{x})$ in an arbitrary target probability density $\rho_1(\boldsymbol{x})$ from trajectories generated by starting points drawn under starting probability density $\rho_0(\boldsymbol{x})$ with invertible dynamics (Eq. 1.20). Besides, if simulations are restrained within a lower boundary and an upper boundary and invertible dynamics is carefully chosen so that there is at most one intersection point between each boundary and the trajectory, then the unbiased ITNEIS estimator reduces to the original unbiased NEIS estimator (Eq. 1.36).

There are three critical properties, which have been proved in Eq. 1.21, Eq. 1.34 and Sec. 2.8.1, that will be frequently used in the proof regarding NEIS estimators.

- $\boldsymbol{X}(s, \boldsymbol{X}(t, \boldsymbol{x})) = \boldsymbol{X}(s+t, \boldsymbol{x})$
- $J(s, \boldsymbol{X}(t, \boldsymbol{x}))\, J(t, \boldsymbol{x}) = J(s+t, \boldsymbol{x})$
- $\boldsymbol{y} \equiv \boldsymbol{X}(t, \boldsymbol{x}) \Rightarrow d\boldsymbol{y} = J(t, \boldsymbol{x})\, d\boldsymbol{x}$

The following proof shows that FTNEIS estimators and ITNEIS estimators are unbiased.

$$
\begin{aligned}
\left\langle \phi_{\mathrm{neq}}^{\mathrm{FTNEIS}} \right\rangle_0 &\equiv \int_\Gamma \phi_{\mathrm{neq}}^{\mathrm{FTNEIS}}(\boldsymbol{x})\, \rho_0(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_\Gamma \rho_0(\boldsymbol{x}) \int_{\tau^-}^{\tau^+} \frac{\phi(\boldsymbol{X}(t, \boldsymbol{x}))\, \rho_1(\boldsymbol{X}(t, \boldsymbol{x}))\, J(t, \boldsymbol{x})}{\int_{t-\tau^+}^{t-\tau^-} \rho_0(\boldsymbol{X}(s, \boldsymbol{x}))\, J(s, \boldsymbol{x})\, ds}\, dt\, d\boldsymbol{x} \\
&= \int_\Gamma \phi(\boldsymbol{y})\, \rho_1(\boldsymbol{y}) \int_{\tau^-}^{\tau^+} \frac{\rho_0(\boldsymbol{X}(-t, \boldsymbol{y}))}{\int_{t-\tau^+}^{t-\tau^-} \rho_0(\boldsymbol{X}(s, \boldsymbol{X}(-t, \boldsymbol{y})))\, J(s, \boldsymbol{X}(-t, \boldsymbol{y}))\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi(\boldsymbol{y})\, \rho_1(\boldsymbol{y}) \int_{\tau^-}^{\tau^+} \frac{\rho_0(\boldsymbol{X}(-t, \boldsymbol{y}))\, J(-t, \boldsymbol{y})}{\int_{t-\tau^+}^{t-\tau^-} \rho_0(\boldsymbol{X}(s-t, \boldsymbol{y}))\, J(s-t, \boldsymbol{y})\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi(\boldsymbol{y})\, \rho_1(\boldsymbol{y}) \int_{\tau^-}^{\tau^+} \frac{\rho_0(\boldsymbol{X}(-t, \boldsymbol{y}))\, J(-t, \boldsymbol{y})}{\int_{\tau^-}^{\tau^+} \rho_0(\boldsymbol{X}(-s, \boldsymbol{y}))\, J(-s, \boldsymbol{y})\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi(\boldsymbol{y})\, \rho_1(\boldsymbol{y})\, d\boldsymbol{y} \equiv \langle \phi \rangle_1
\end{aligned}
\tag{1.45}
$$

$$
\begin{aligned}
\left\langle \phi_{\mathrm{neq}}^{\mathrm{ITNEIS}} \right\rangle_0 &\equiv \int_\Gamma \phi_{\mathrm{neq}}^{\mathrm{ITNEIS}}\left(\boldsymbol{x}\right) \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \rho_0\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds}\, dt\, d\boldsymbol{x} \\
&= \int_\Gamma \phi\left(\boldsymbol{y}\right) \rho_1\left(\boldsymbol{y}\right) \int_{-\infty}^{+\infty} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\boldsymbol{y}\right)\right)}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{X}\left(-t,\boldsymbol{y}\right)\right)\right) J\left(s,\boldsymbol{X}\left(-t,\boldsymbol{y}\right)\right)\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi\left(\boldsymbol{y}\right) \rho_1\left(\boldsymbol{y}\right) \int_{-\infty}^{+\infty} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\boldsymbol{y}\right)\right) J\left(-t,\boldsymbol{y}\right)}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s-t,\boldsymbol{y}\right)\right) J\left(s-t,\boldsymbol{y}\right)\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi\left(\boldsymbol{y}\right) \rho_1\left(\boldsymbol{y}\right) \int_{-\infty}^{+\infty} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\boldsymbol{y}\right)\right) J\left(-t,\boldsymbol{y}\right)}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(-s,\boldsymbol{y}\right)\right) J\left(-s,\boldsymbol{y}\right)\, ds}\, dt\, d\boldsymbol{y} \\
&= \int_\Gamma \phi\left(\boldsymbol{y}\right) \rho_1\left(\boldsymbol{y}\right)\, d\boldsymbol{y} \equiv \left\langle \phi \right\rangle_1
\end{aligned}
\tag{1.46}
$$

Another significant contribution from Ref. [65] is the finding of the optimal flow condition regarding unbiased ITNEIS estimators with given starting probability density $\rho_0\left(\boldsymbol{x}\right)$ and given target probability density $\rho_1\left(\boldsymbol{x}\right)$:

$$
\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J^{\mathrm{opt}}\left(t,\boldsymbol{x}\right)\, dt = \int_{-\infty}^{+\infty} \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J^{\mathrm{opt}}\left(t,\boldsymbol{x}\right)\, dt. \tag{1.47}
$$

Under the optimal flow condition (Eq. 1.47), the performance of the unbiased ITNEIS estimator is no worse than that of the equilibrium counterpart under target probability density $\rho_1\left(\boldsymbol{x}\right)$ by Jensen's inequality [63].

$$
\begin{aligned}
\mathrm{Var}\left[\phi_{\mathrm{neq}}^{\mathrm{ITNEIS}}\right] &\equiv \left\langle \left| \phi_{\mathrm{neq}}^{\mathrm{ITNEIS}} \right|^2 \right\rangle_0 - \left\langle \phi_{\mathrm{neq}}^{\mathrm{ITNEIS}} \right\rangle_0^2 = \left\langle \left| \phi_{\mathrm{neq}}^{\mathrm{ITNEIS}} \right|^2 \right\rangle_0 - \left\langle \phi \right\rangle_1^2 \\
&= \int_\Gamma \rho_0\left(\boldsymbol{x}\right) \left(\frac{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \right)^2 d\boldsymbol{x} - \left\langle \phi \right\rangle_1^2 \\
&\leqslant \int_\Gamma \rho_0\left(\boldsymbol{x}\right) \frac{\int_{-\infty}^{+\infty} \phi^2\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds}\, d\boldsymbol{x} - \left\langle \phi \right\rangle_1^2 \\
&= \left\langle \phi^2 \right\rangle_1 - \left\langle \phi \right\rangle_1^2 \equiv \mathrm{Var}\left[\phi\right]
\end{aligned}
\tag{1.48}
$$

In Chapter 2 we will show how the quench dynamics and NEIS estimators can be used on

molecular systems to estimate free energies and partition functions. In subsequent chapters, we will make use of the preceding definitions to expand upon these approaches, and also to estimate basin volumes through trajectories.

2 | Computing equilibrium free energies through a nonequilibrium quench

This chapter has been adapted from Ref. [66].

2.1 Abstract

Many methods to accelerate sampling of molecular configurations are based on the idea that temperature can be used to accelerate rare transitions. These methods typically compute equilibrium properties at a target temperature using reweighting or through Monte Carlo exchanges between replicas at higher temperatures. A recent paper [60] demonstrated that accurate equilibrium densities of states can also be computed through a nonequilibrium "quench" process, where sampling is performed at a higher temperature to encourage rapid mixing and then quenched to lower energy states with dissipative dynamics. Here we provide an implementation of the quench dynamics in LAMMPS and evaluate a new formulation of nonequilibrium estimators for the computation of partition functions or free energy surfaces (FESs) of molecular systems. We show that the method is exact for a minimal model of N-independent harmonic springs, and use these analytical results to develop heuristics for the amount of quenching required to obtain accurate sampling. We then test the quench approach on alanine dipeptide, where we show that it gives an FES that is accurate near the most stable configurations using the quench approach but

disagrees with a reference umbrella sampling calculation in high FE regions. We then show that combining quenching with umbrella sampling allows the efficient calculation of the free energy in all regions. Moreover, by using this combined scheme, we obtain the FES across a range of temperatures at no additional cost, making it much more efficient than standard umbrella sampling if this information is required. Finally, we discuss how this approach can be extended to solute tempering and demonstrate that it is highly accurate for the case of solvated alanine dipeptide without any additional modifications.

2.2 Introduction

A major challenge in molecular dynamics (MD) simulations is poor sampling of conformational landscapes because free energy barriers that are large relative to $k_{\mathrm{B}}T$ are traversed at rates much lower than the duration of a typical simulation [17, 67]. A wide variety of approaches relying on sampling of equilibrium distributions have been proposed [68] to circumvent this problem, which can be generally classified into (**a**) those which seek to lower free energy barriers by adding a bias or changing the potential that is being sampled [33, 35], or (**b**) those that use higher temperatures of all or some degrees of freedom to accelerate transitions [41, 42]. Other approaches harness nonequilibrium fluctuation theorems to estimate equilibrium free energies by averaging over many realizations of a nonequilibrium transformation [57, 69, 70]. However, these non-equilibrium approaches have not been widely adopted for chemical problems because they are difficult to converge, due to the large variance in work performed and the largest contributions to the equilibrium average being dominated by rare fluctuations.

Ref. [60] proposes a class of estimators based on an exact reweighting of the samples gathered during a nonequilibrium process that follows a dissipative dynamical scheme, starting with configurations that are well sampled from an equilibrium density, e.g. the Boltzmann distribution (see also Ref. [71], Ch. 5 and Refs. [72] and [65]). Conceptually, facile sampling at high tempera-

ture allows mixing between free energy basins and then the process of 'quenching' allows one to map out the lower energy portion of the free energy basins. This method possesses several advantages; it is unbiased and only requires the knowledge of the starting probability density ρ_0 up to a constant, unlike the annealed importance sampling [59] (AIS) method, which requires computing a posterior ratio of sample means. Furthermore, it was proven in Ref. [60] that the estimator has lower variance than a direct estimator and simulations can be run in parallel, which makes the methodology naturally suited to the architecture of high-performance computing clusters. The aim of this chapter is to investigate how to use this method to compute partition functions and FESs for molecular systems.

To this end, we derive a formulation of these nonequilibrium quench estimators for molecular systems and critically assess the efficacy of this approach. We demonstrate that the method is exact for a harmonic system where analytical results are available, and, in doing so, we obtain heuristic rules for the required amount of sampling. We then demonstrate on the simple test system of alanine dipeptide that, because the method emphasizes low free energy regions, it is not a competitive approach for computing full free-energy surfaces. Nevertheless, we find that a combination of quenching with umbrella sampling (US) provides a highly efficient way to compute a full FES for this system, simultaneously giving FESs at many different temperatures. Finally, we expand upon the quench method and show that it can be used as a nonequilibrium solute-tempering approach with highly accurate results for a solvated peptide. The quenching dynamics are implemented in LAMMPS [73, 74], and analysis methods provided as open source Python scripts, meaning that our method can be easily deployed on other problems.

2.3 Theory and Methods

2.3.1 An unbiased non-equilibrium estimator from invertible non-equilibrium dynamics

Let us denote the microstate of the system in phase space by the vector $\boldsymbol{x} = (\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}^{2d}$ and $\boldsymbol{q}, \boldsymbol{p} \in \mathbb{R}^d$, where $d \equiv 3n$ is the number of degrees of freedom (DOF) in the n-atom system. An average observable property ϕ of the system can by computed as an integral over all possible configurations, weighted by the equilibrium probability density [17, 67],

$$\langle \phi \rangle = \frac{\int d\boldsymbol{x}\, \phi(\boldsymbol{x})\, \rho(\boldsymbol{x})}{\int d\boldsymbol{x}\, \rho(\boldsymbol{x})} \tag{2.1}$$

A typical example of $\rho(\boldsymbol{x})$ would be the Boltzmann distribution,

$$\rho_\beta(\boldsymbol{x}) = \frac{e^{-\beta \mathscr{H}(\boldsymbol{x})}}{\int d\boldsymbol{x}\, e^{-\beta \mathscr{H}(\boldsymbol{x})}} \equiv \frac{e^{-\beta \mathscr{H}(\boldsymbol{x})}}{Q(\beta)}, \tag{2.2}$$

where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann's constant, T is the temperature, and $\mathscr{H}$ is the Hamiltonian (total energy function) of the system [17, 67]. Here, $Q(\beta)$ is the canonical partition function.

To compute equilibrium averages using molecular dynamics, we typically replace the expectation (2.1) by a time average along a trajectory $\boldsymbol{X}^e(t)$ along which configurations appear in proportion to $\rho_\beta(\boldsymbol{x})$. This can be done using Markov chain Monte Carlo or molecular dynamics with thermostat [17, 67]. In this case,

$$\langle \phi \rangle \approx \frac{1}{\tau} \int_0^\tau dt \phi(\boldsymbol{X}^e(t)) \tag{2.3}$$

with equality in the limit as $\tau \to \infty$.

As an alternative to running a very long trajectory, if we already had configurations well

sampled from $\rho(\boldsymbol{x})$, then we could also compute this same average by an "initiate-and-propagate" procedure, where we draw starting points $\boldsymbol{X}_i^e(0)$ from $\rho(\boldsymbol{x})$ and propagate our equations of motion to get N trajectories of length τ_{short}, $\{\boldsymbol{X}_i^e(t)\}$, computing observables as an average over the trajectory and over initial configurations,

$$\langle\phi\rangle \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tau_{\text{short}}}\int_0^{\tau_{\text{short}}} dt\phi\left(\boldsymbol{X}_i^e(t)\right) \tag{2.4}$$

with equality in the limit as $N \to \infty$ for any $\tau_{\text{short}} > 0$. The advantage of such a procedure would be that each trajectory can be simulated independently, making the algorithm trivially parallelizable.

Now suppose we wanted to do such a procedure, but the dynamics do not sample the stationary distribution ρ, using e.g. the differential equation

$$\dot{\boldsymbol{X}}(t) = \boldsymbol{b}(\boldsymbol{X}(t)). \tag{2.5}$$

where $\boldsymbol{b}(\boldsymbol{x})$ is a vector-field to be specified that does not preserves $\rho(\boldsymbol{x})$, i.e. $\boldsymbol{\nabla}\cdot(\boldsymbol{b}(\boldsymbol{x})\rho(\boldsymbol{x})) \neq 0$ where $\boldsymbol{\nabla}$ corresponds to the phase space gradient $\{\frac{\partial}{\partial q_1}, \frac{\partial}{\partial q_2}, ..., \frac{\partial}{\partial q_d}, \frac{\partial}{\partial p_1}, \frac{\partial}{\partial p_2}, ..., \frac{\partial}{\partial p_d}\}$. The change in phase space volume associated with a nonequilibrium dynamical process is quantified by a Jacobian factor,

$$J(t) = \exp\left(\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}(\boldsymbol{X}(s))\, ds\right), \tag{2.6}$$

as derived in Appendix 2.8.1.

Ref. [60] makes use of the fact that Eq. 2.4 can be extended to this more general case of motion generated by $\boldsymbol{b}$ by introduction of the density scaled by this Jacobian factor, as long as points can be sampled from the initial density $\rho(\boldsymbol{x})$. In this case, estimates are computed for a subset of all phase space by propagating the non-equilibrium trajectories until they reach the boundaries of that subset of phase space, which in practice was done by terminating trajectories

at fixed maximum and minimum energy values $E_{\max}$ and $E_{\min}$. The resulting estimator over N trajectories is given by,

$$\langle \phi \rangle \approx \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\int_{\tau_i^-(E_{\max})}^{\tau_i^+(E_{\min})} dt \phi\left(\boldsymbol{X}_i\left(t\right)\right) \rho\left(\boldsymbol{X}_i\left(t\right)\right) J\left(t\right)}{\int_{\tau_i^-(E_{\max})}^{\tau_i^+(E_{\min})} dt \rho\left(\boldsymbol{X}_i\left(t\right)\right) J\left(t\right)}. \tag{2.7}$$

where $\tau_i^+\left(E_{\min}\right)$ and $\tau_i^-\left(E_{\max}\right)$ are the times that trajectory i reached the fixed energy boundaries when propagating the non-equilibrium dynamics forwards and backwards in time (we emphasize that here the integration times vary for each starting point).

As in Ref. [60], we use the equations of motion corresponding to zero temperature Langevin dynamics, which we term "quench" dynamics,

$$\begin{cases} \dot{\boldsymbol{Q}} = \mathbf{M}^{-1}\boldsymbol{P} \\ \dot{\boldsymbol{P}} = -\boldsymbol{\nabla} U\left(\boldsymbol{Q}\right) - \gamma \boldsymbol{P} \end{cases}, \tag{2.8}$$

for which the Jacobian is

$$J\left(t\right) = \exp\left(-d\gamma t\right). \tag{2.9}$$

This method is easy to implement for molecular systems by adapting the BAOAB scheme [75] (see Appendix 2.8.2).

Backwards-in-time trajectories from initial points are generated by following the same dynamical scheme using a negative γ, after reversing the initial momenta. This is derived by applying the equations of motion to the time-reversed phase space coordinates $\boldsymbol{Q}^R\left(t\right) = \boldsymbol{Q}\left(-t\right)$ and $\boldsymbol{P}^R\left(t\right) = -\boldsymbol{P}\left(-t\right)$.

Because, Eq. 2.8 is dissipative, we can use it to propagate trajectories from high energy to low energy, or from low energy to high energy using a negative γ. With this scheme, we can compute Boltzmann averages at the starting inverse temperature β_0 by first sampling from $\rho_{\beta_0}\left(\boldsymbol{x}\right) \propto \exp\left(-\beta_0 \mathscr{H}\left(\boldsymbol{x}\right)\right)$ and propagating N trajectories with our quench algorithm forwards and back-

wards in time, using the formula

$$\langle \phi \rangle_0 \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\int_{\tau_i^-}^{\tau_i^+} dt \phi \left(\boldsymbol{X}_i \left(t \right) \right) e^{-\beta_0 \mathscr{H}(\boldsymbol{X}_i(t)) - d\gamma t}}{\int_{\tau_i^-}^{\tau_i^+} dt e^{-\beta_0 \mathscr{H}(\boldsymbol{X}_i(t)) - d\gamma t}} \tag{2.10}$$

Here we no longer indicate the dependence of $\tau_i^{\pm}$ on energy for brevity. An extension of Eq. 2.10 to calculate averages at other temperatures above and below β_0 will be discussed in the next section.

We note that Eq. 2.10 is a biased estimator since it computes expectations over ρ_{β_0} conditional on $\mathscr{H}\left(\boldsymbol{x}\right) \in [E_{\min}, E_{\max}]$. To make this bias negligible, we can adjust the values of $E_{\min}$ and $E_{\max}$. To this end, notice that during a (forward) quench, the value of total energy $E_{\text{tot}} = \mathscr{H}\left(\boldsymbol{X}\left(t\right)\right)$ will decrease while $d\gamma t$ will increase, resulting in a time where the arguments of the exponentials are minimized that depends on β_0. In order to get a converged average, this time must be contained within the range (τ^-, τ^+), and so the energy levels, in particular $E_{\min}$, must be chosen such this is the case. Our method for doing so is discussed in Sec. 2.5.

2.3.2 Calculations of free energies and partition functions

As mentioned earlier, the principal challenge of computing quantities from MD simulations is that high free energy barriers at a temperature of interest prevent proper sampling of all relevant configurations with proper weights. Since it can be easier to sample at high temperature, it is tempting to sample at high temperature, and directly reweight samples to lower temperature; for example, to estimate $Q\left(\beta\right)$ we can write

$$\begin{aligned} Q\left(\beta\right) &= \int d\boldsymbol{x} e^{-\beta \mathscr{H}(\boldsymbol{x})} = \int d\boldsymbol{x} e^{-\beta \mathscr{H}(\boldsymbol{x})} \left(\frac{e^{-\beta_0 \mathscr{H}(\boldsymbol{x})}}{e^{-\beta_0 \mathscr{H}(\boldsymbol{x})}} \right) \\ &= \int d\boldsymbol{x} e^{(\beta_0 - \beta)\mathscr{H}(\boldsymbol{x})} e^{-\beta_0 \mathscr{H}(\boldsymbol{x})}. \end{aligned} \tag{2.11}$$

Hence, we can reweight samples from β_0 to evaluate the relative value of $Q(\beta)$,

$$\frac{Q(\beta)}{Q(\beta_0)} = \langle e^{(\beta_0-\beta)\mathscr{H}} \rangle_0, \tag{2.12}$$

which the central idea of free energy perturbation [67]. The challenge is that samples from $\rho_{\beta_0}(\boldsymbol{x})$ do not have good overlap with $\rho_\beta(\boldsymbol{x})$ unless $\beta \approx \beta_0$, so this estimate could have high variance in practice, which can be mitigated using simulated annealing or simulated tempering [76–78].

The same quantity in Eq. 2.12 can be computed with the quench estimator (Eq. 2.10),

$$\frac{Q(\beta)}{Q(\beta_0)} \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{\tau_i^-}^{\tau_i^+} dt e^{-\beta\mathscr{H}(\boldsymbol{X}_i(t))-d\gamma t}}{\int_{\tau_i^-}^{\tau_i^+} dt e^{-\beta_0\mathscr{H}(\boldsymbol{X}_i(t))-d\gamma t}}. \tag{2.13}$$

By quenching forwards in time, low energy samples which are more relevant at a lower temperature are generated, which should result in a much more robust calculation than reweighting from samples generated only from β_0 at equilibrium. Using a similar manipulation, we can compute the average of any observable $\phi(\boldsymbol{X})$ at β using samples generated from β_0,

$$\begin{aligned}\langle\phi\rangle &= \langle\phi e^{(\beta_0-\beta)\mathscr{H}}\rangle_0 / \langle e^{(\beta_0-\beta)\mathscr{H}}\rangle_0 \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{\tau_i^-}^{\tau_i^+} dt\phi(\boldsymbol{X}_i(t)) e^{-\beta\mathscr{H}(\boldsymbol{X}_i(t))-d\gamma t}}{\int_{\tau_i^-}^{\tau_i^+} dt e^{-\beta_0\mathscr{H}(\boldsymbol{X}_i(t))-d\gamma t}} \left(\frac{Q(\beta)}{Q(\beta_0)}\right)^{-1},\end{aligned} \tag{2.14}$$

where $Q(\beta)/Q(\beta_0)$ is computed via Eq. 2.13.

For a coordinate (possibly a vector) defined by a function $S(\boldsymbol{x})$, the FES or potential of mean force (PMF) is given by

$$F(s;\beta) = -\frac{1}{\beta}\ln(\langle\delta(S-s)\rangle), \tag{2.15}$$

where δ is the Dirac delta function [1]. Substituting in Eq.2.14 shows how we can estimate $F(s;\beta)$

[1] In real MD simulations, it is impossible to compute free energy at any particular s. Rather, block function (integrating delta function over windows) is used, and we show in Appendix 2.8.3 that free energy computed in this

using samples from β_0, which up to a constant factor gives

$$\rho\left(s;\beta\right) \propto e^{-\beta F(s;\beta)} = \left\langle \delta\left(S-s\right) e^{(\beta_0-\beta)\mathscr{H}} \right\rangle_0 . \tag{2.16}$$

Using Eq. 2.10, we then obtain our final result, which shows how the FES can be computed using quench trajectories,

$$e^{-\beta F(s,\beta)} \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{\tau_i^-}^{\tau_i^+} dt \delta\left(S\left(\boldsymbol{X}_i\left(t\right)\right)-s\right) e^{-\beta\mathscr{H}\left(\boldsymbol{X}_i(t)\right)-d\gamma t}}{\int_{\tau_i^-}^{\tau_i^+} dt e^{-\beta_0\mathscr{H}\left(\boldsymbol{X}_i(t)\right)-d\gamma t}} \tag{2.17}$$

This estimator allows us to compute the PMF at a range of temperatures β above and below β_0 using a single set of trajectories.

Because the exponential decay $e^{-d\gamma t}$ suppresses contributions at long forwards times and the exponential increase of $\mathscr{H}\left(\boldsymbol{X}_i\left(t\right)\right)$ does so for large negatives times, we also considered running simulations for fixed forwards and backwards times where $\tau^+ \gg \tau_i^+\left(E_{\min}\right)$ and $\tau^- \ll \tau_i^-\left(E_{\max}\right)$. In this case,

$$e^{-\beta F(s,\beta)} \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{\tau^-}^{\tau^+} dt \delta\left(S\left(\boldsymbol{X}_i\left(t\right)\right)-s\right) e^{-\beta\mathscr{H}\left(\boldsymbol{X}_i(t)\right)-d\gamma t}}{\int_{\tau^-}^{\tau^+} dt e^{-\beta_0\mathscr{H}\left(\boldsymbol{X}_i(t)\right)-d\gamma t}}, \tag{2.18}$$

where now the integration limits are fixed for all runs. In Sec. 2.5 we will give a heuristic for how long quench trajectories should be run. In practice, to generate results we run fixed length simulations using that heuristic from many initial points, then pick energy cutoffs, and then use the estimator given by Eq. 2.17, which did prove to be more accurate than Eq. 2.18. This strategy allowed us to test both estimators and works well in practice, but does require more total simulation time than if a perfect energy cutoff were known *a priori*.

Finally, we note that Ref. [65] proposes another exact estimator from the same type of trajec-

way has an error with magnitude $\mathcal{O}\left(\Delta s^2\right)$, where Δs is the width of windows.

tories,

$$\langle \phi \rangle \approx \frac{1}{N} \sum_{i=1}^{N} \int_{\tau^-}^{\tau^+} dt \frac{\phi\left(\boldsymbol{X}_i\left(t\right)\right) \rho\left(\boldsymbol{X}_i\left(t\right)\right) J\left(t\right)}{\int_{t-\tau^+}^{t-\tau^-} dt' \rho\left(\boldsymbol{X}_i\left(t'\right)\right) J\left(t'\right)}, \tag{2.19}$$

where τ^+ and τ^- are constant. We also evaluate this formula for a test case in the Supporting Information, but find that it is more difficult to use in practice for molecular systems because (**a**) it requires obtaining data from $\tau^- - \tau^+$ to $\tau^+ - \tau^-$, which is a strictly larger time window than in Eq. 2.17, and (**b**) the long reverse quench to time $\tau^- - \tau^+$ can cause the MD simulation to become unstable as the kinetic energy grows exponentially, resulting in simulations crashing. Hence we do not pursue it further in this chapter.

2.4 Implementation

We implemented quench dynamics in LAMMPS [73, 74] using the procedure described in Appendix 2.8.2, with a user-defined "fix", and run trajectories using the LAMMPS python interface. Then run many parallel trajectories in a Python framework using the parallel scripting language Parsl [79], which also interfaces with common high-performance computing queuing systems. We also perform analysis in parallel using parsl. LAMMPS source code, as well as all run and analysis scripts, are provided in a github repository for this chapter (https://github.com/hocky-research-group/quench_paper_2023).

2.5 Results

2.5.1 Computing the partition function of independent harmonic springs through quenching

To confirm the validity of the quench approach for a molecular system, as well as to check our implementation, we first start with a system for which we know the ground truth. We chose

to study a system of $\mathcal{N}$ independent harmonic springs, with a Hamiltonian defined by,

$$\mathscr{H}\left(\boldsymbol{X}\right) = \sum_{i=1}^{\mathcal{N}} \left(\frac{|\boldsymbol{P}_i|^2}{2m} + \frac{1}{2} m\omega^2 |\boldsymbol{Q}_i|^2 \right) \tag{2.20}$$

Because these springs are independent, this is equivalent to $3\mathcal{N}$ one-dimensional harmonic oscillators defined by the simple Hamiltonian,

$$\mathscr{H}\left(q,p\right) = \frac{p^2}{2m} + \frac{1}{2} m\omega^2 q^2 \tag{2.21}$$

The partition function for this system is separable such that,

$$\begin{aligned} Q_{3\mathcal{N}}\left(\beta\right) = Q_1^{3\mathcal{N}}\left(\beta\right) &= \left(\iint dqdp e^{-\frac{\beta p^2}{2m} - \frac{\beta}{2} m\omega^2 q^2} \right)^{3\mathcal{N}} \\ &= \left(\sqrt{\frac{2\pi m}{\beta}} \times \sqrt{\frac{2\pi}{\beta m\omega^2}} \right)^{3\mathcal{N}} = \left(\frac{2\pi}{\beta\omega} \right)^{3\mathcal{N}} \end{aligned} \tag{2.22}$$

We can therefore benchmark our quench approach by computing the ratio of partition functions at two different temperatures using Eq. 2.13 and compare to the exact value, which is given by $(\beta_0/\beta)^{3\mathcal{N}}$.

Using LAMMPS, we sample $\mathcal{N}$ independent harmonic springs in 3D with identical masses $m = 1.0$ and identical oscillation frequencies $\omega = \sqrt{5}$, in reduced units. We first generate 2000 starting points using Langevin dynamics [80] with friction coefficient $\gamma_{LD} = 0.01$ and time step $\Delta t = 0.001$ in reduced units. To do so, we first equilibrate the system for 10^7 steps ($\tau = 10^4$ in reduced LJ time units) at $\beta = 1$. Then we run production simulation for 2×10^7 steps and save 2000 starting points for further "quench" simulations.

In Fig. 2.1, we show the behavior of the energy of the system when running quench simulations using the EOM described by Eq. 2.8 forward and backwards in time, for several values of γ_{quench}. We observe an overall exponential decay of average total energy at small γ with respect

to $\gamma_{quench}t$, a unitless "time" that we find serves as a good progress coordinate. In contrast, when γ_{quench} is large, we observe large deviation from standard exponential decay, with a low-frequency oscillation.

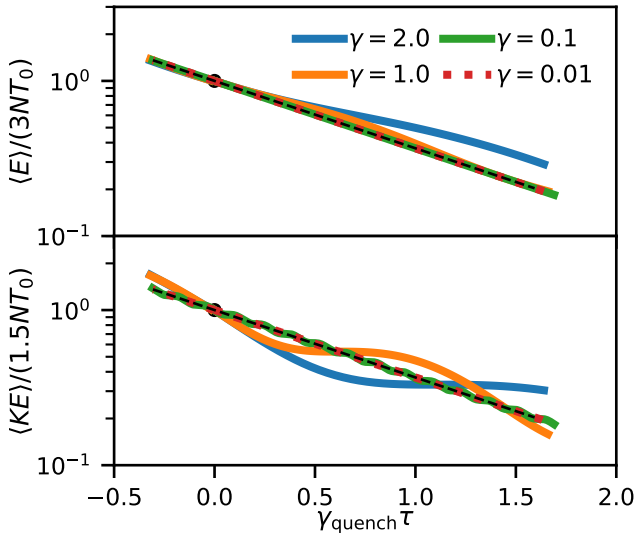

Figure 2.1: Mean total energy (top) and mean kinetic energy (bottom) with respect to reduced time ($\gamma_{quench}\tau$) for $\mathcal{N} = 1000$ springs, varying γ_{quench}, scaled by the equilibrium energy given by equipartition. Quenching is performed from 2000 starting points. Circles indicate the starting energies before quenching forwards and backwards in time. For small γ_{quench}, mean total energy and mean kinetic energy follow an exponential decay with time constant γ^{-1} (dashed line).

To understand this behavior, we can solve the EOM of a 1D system exactly given an initial condition (q_0, p_0),

$$\begin{aligned} E(t)\exp(\gamma t) = &\frac{2m^2\omega^4 q_0^2 + 2m\omega^2\gamma q_0 p_0 + 2\omega^2 p_0^2}{m\left(4\omega^2 - \gamma^2\right)} \\ &+ \frac{m^2\omega^2\gamma q_0^2 - \gamma p_0^2}{2m\sqrt{4\omega^2 - \gamma^2}}\sin\left(\sqrt{4\omega^2 - \gamma^2}\,t\right) \\ &+ \frac{m^2\omega^2\gamma^2 q_0^2 - 4m\omega^2\gamma q_0 p_0 - \gamma^2 p_0^2}{2\left(4\omega^2 - \gamma^2\right)}\cos\left(\sqrt{4\omega^2 - \gamma^2}\,t\right) \end{aligned} \tag{2.23}$$

The magnitude of the ratio between terms scales for $\gamma < \omega$ as $\mathcal{O}(1) : \mathcal{O}(\gamma/\omega) : \mathcal{O}\left(\gamma^2/\omega^2\right)$ such that the first term dominates for small γ. As γ approaches ω, then oscillations appear with a period of $\pi\left(\omega^2/\gamma^2 - 1/2\right)^{-1/2}$ when plotted against γt. Similarly, we can solve an expression for

kinetic energy, and find that in the small γ limit, the mean kinetic energy also decays exponentially with respect to time. In the quasi-static quenching limit, we can approximate all springs as independently following the same exponential decay, and therefore the sum of their energies also decays exponentially, $E(t) \approx E(0) e^{-\gamma t}$. For this situation, we can compute the partition function,

$$\begin{aligned} Q(\beta) &\propto \int_{-\infty}^{\infty} \exp\left(-\beta E_0 e^{-\gamma t} - d\gamma t\right) dt \\ &= \int_0^{\infty} \frac{1}{\gamma} \exp\left(-\beta E_0 u\right) u^{d-1} du \\ &= \frac{(d-1)!}{\gamma} (\beta E_0)^{-d}. \end{aligned} \tag{2.24}$$

This gives the correct result for the ratio of partition functions, $(\beta/\beta_0)^{-d}$. Moreover, we could calculate $Q(\beta)$ using a saddle point approach and find that the exponential term is dominated by its value when $\gamma t = \ln(\beta/\beta_0)$, using the fact that $E_0 \approx d/\beta_0$ for Harmonic oscillators.

Because the mean kinetic energy decays exponentially with respect to time starting at a value of $d/(2\beta_0)$, at this particular moment $t = \gamma^{-1} \ln(\beta/\beta_0)$ the kinetic energy obtains a value of $d/(2\beta)$ corresponding to a temperature T. In real simulations it is impossible to run to infinite times, so this harmonic model suggests that we can guess how long to run by choosing $\gamma(\tau^+ - \tau^-) > |\ln(\beta/\beta_0)|$ for a target β if we want to use the infinite time approximation as in Eq. 2.18. To confirm this for the harmonic system, we show in Fig. 2.9 that convergence of the ratio of partition functions reaches this limit where Eq. 2.18 holds once $\gamma(\tau^+ - \tau^-)$ exceeds this value ($2\ln(2) \approx 1.4$). When γ_{quench} approaches ω and the exponential decay of energy does not hold, the partition function ratio converges to an incorrect value.

Finally, we test the accuracy of our primary estimator Eq. 2.13. To do so, we take the same quenches performed for Fig. 2.1 and pick τ^+ and τ^- as the maximum final energy of the 2000 forward quenches and the minimum of the 2000 reverse quenches. In Fig. 2.2 we show that our quench estimator is highly accurate. Fig. 2.2(a) shows that the estimator gets more accurate with an increasing number of springs, despite the increasing phase space volume that must be sampled.

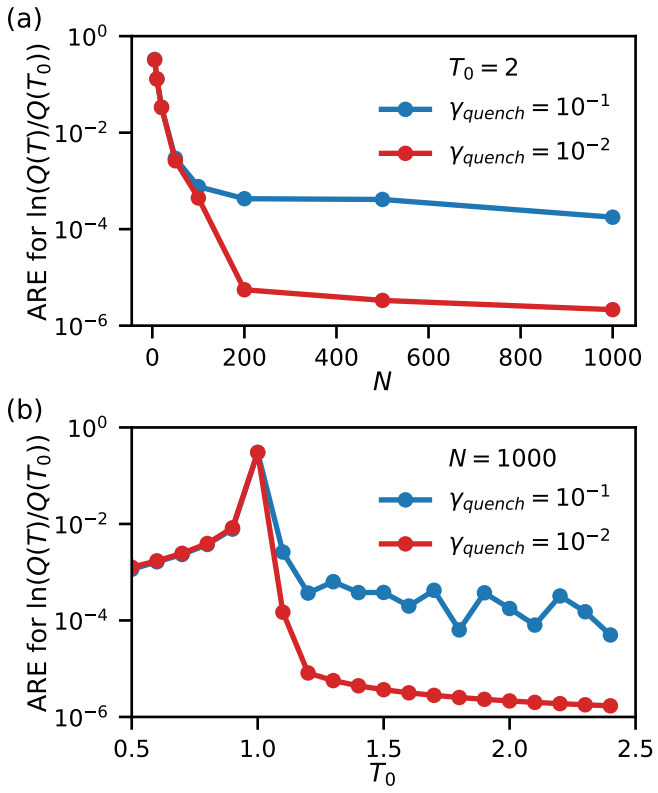

Figure 2.2: Absolute relative error in the ratio of partition functions at two different temperatures T_0 and $T = 1$ for a system of independent harmonic springs using Eq. 2.13. (a) Error computed starting from $T_0 = 2$ and varying the number of springs for two different quench rates. (b) Error computed for fixed $N = 1000$ springs when varying starting temperature at two different quench rates. Two other estimators are compared for this setup in Fig. 2.11.

Fig. 2.2(b) shows the effect of varying the initial temperature T_0, with more accurate results when T_0 is higher than T as intuitively expected. Accuracy is improved substantially by decreasing the quench rate γ, but at the expense of longer simulations.

2.5.2 Computing the free energy surface of alanine dipeptide

Having demonstrated that we are able to compute partition functions for a toy system, we are now interested in whether we can compute PMFs for a molecular system through a quenching procedure (where partition functions are not a very useful quantity on their own). We first test our approach by computing Eq. 2.17 for the test case of alanine dipeptide in vacuum, which has two major conformations (one of which has sub-populations) separated by a relatively large energy barrier, and whose configurations are well captured by considering its FES in the space of two backbone dihedral angles ϕ and ψ. This FES has been extensively characterized by a number of enhanced sampling approaches and serves as a prototypical benchmark system, although one which is quite easy to sample for some approaches such as metadynamics.

Simulations of this molecule in vacuum were run using LAMMPS simulation package and using CHARMM 27 force field without CMAP corrections [81]. Equilibrium simulations are performed using LAMMPS's Langevin dynamics thermostat with $\gamma_{\text{run}} = 0.01\text{ps}^{-1}$ and an MD timestep of $\Delta t = 1$ fs. Umbrella sampling [33] was used to obtain reference free energy surface (FES). Regarding umbrella sampling, a harmonic biased potential with spring constant 24.0 $\text{kcal}/\left(\text{mol}\cdot\text{rad}^2\right)$ was added to 400 (20×20) windows along CVs given by backbone dihedral angles (Φ, Ψ). The system was equilibrated for 400 ps at each window location, and then production runs were performed for 2 ns in each window. FES was estimated at the target temperature 300 K using WHAM to combine the data [50], and was also computed for comparison using EMUS [46]. Finally, for comparison, FESs were computed using Well-Tempered MetaDynamics (WT-MetaD) [35] in PLUMED [82–84] applied to ϕ, ψ with hills deposited every 500 steps, a hill width of 0.35 radians in each direction, a hill height of 0.286807 kcal/mol, and a bias factor of either 6 or 10 (see SI); WT-MetaD runs were performed for up to 100 ns although the FES estimate stopped changing within several nanoseconds. In "quench" simulations, the system was equilibrated for 10 ns at $T_0 = 1200$ K and 10000 starting points were drawn from a 10 ns produc-

tion simulation with frequency every 1 ps. Fixed time simulations were performed with forward quenches of length $3.4\gamma^{-1}$ and reverse quenches of length $0.6\gamma^{-1}$. For this length forward quench, the final kinetic energy predicted by our exponential decay model is equivalent to approximately $T = 110$ K, below where we want to estimate. Times τ_i^+ and τ_i^- were chosen from these data by histogramming the energies from the forward and reverse trajectories as shown in Fig. 2.12.

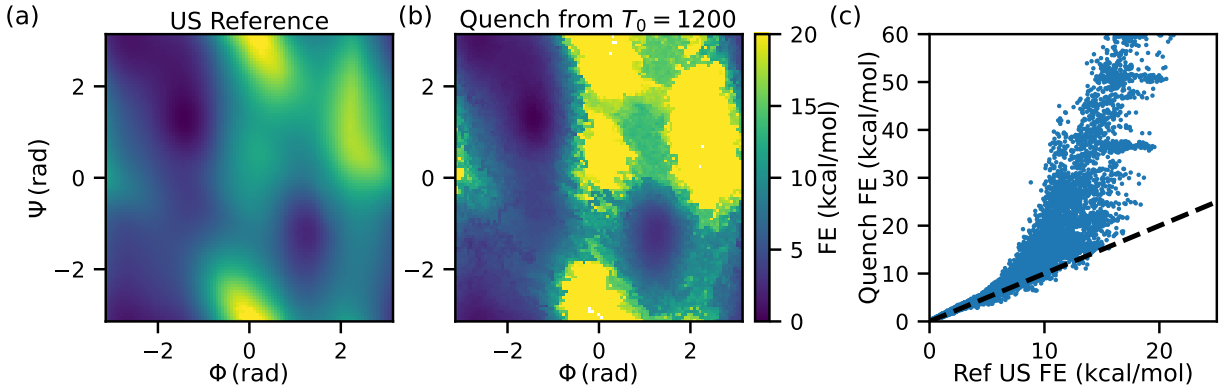

Figure 2.3: A comparison of FES between umbrella sampling and "quench" method. (a) FES computed from US at $T = 300$ K with 400 ns total sampling time. (b) FES computed from "quench" at $T = 300$ K with $T_0 = 1200$ K with $\gamma_{\text{quench}} = 1 \times 10^{-4}$ and 10^4 starting points corresponding to ≈ 410 ns of total simulation time, of which data used in computing Eq. 2.17 (including 10 ns for generating restart points) totals ≈ 226 ns. (c) Comparison of FES values on a bin-by-bin basis. While the minima are captured, the high free energy regions are not.

Fig. 2.3 shows the comparison of FES computed by the "quench" method with $\gamma_{\text{quench}} = 1 \times 10^{-4}$, for which the total amount of fixed time sampling was 400 ns, the same as that used in a reference US reference, and the time used with energy cutoffs was only 287 ns (see Tab. 2.1). While the shape of the FES was captured correctly using the "quench" method, only the lower free energy regions were captured with high fidelity. As shown in Fig. 2.3(c), quench results start to deviate from umbrella sampling results at approximately 8 kcal/mol$\approx$ $13.5k_BT$. Since this is a relatively high cutoff, it demonstrates that quench is applicable to molecular situations where the entire FES over some coordinates is not needed. Moreover, with these simulations we should be able to estimate FES at any temperature in the range 200 K to 2200 K based on the amount of forward and reverse quenching performed, which would not be available with CV

based approaches; this advantage is explored much more in the next section. We also note that this was obtained in a CV agnostic manner.

We can also perform the calculations with larger γ_{quench} by a factor of 10, yielding slightly worse but comparable results. Fig. 2.13, showing that the important regions of the landscape can be captured efficiently in tens of nanoseconds, which is similar to the convergence speed of WT-MetaD [35] (see also next section). We can moreover perform this calculation with a 10 or 100 times slower quench, resulting in much more accurate results, but with a relatively small improvement compared to the amount of additional sampling (Fig. 2.14, 2.15). Next, we show that if we wish to resolve the high free energy regions in detail, it is possible to combine quenching and CV based approaches.

2.5.3 Combining quenching and umbrella sampling

Because quenching does a good job capturing the low free energy regions, we predicted that it could be combined with US to produce an efficient sampling approach. By nature of introducing a harmonic potential, we obtain by construction a more convex landscape, albeit one that can still be hard to sample due to slow orthogonal degrees of freedom. To directly compute the FES from our quench data, we also derived a modified version of the WHAM equations following the derivation of Ref. [50] (See Appendix 2.8.4). Surprisingly, it turns out that we can use exactly the same WHAM equations when $\beta = \beta_0$ by realizing that Jacobian $J(t)$ is same among umbrella windows and the ratio of any two umbrella windows are fixed over time. When we do not start from the target temperature, we can estimate the biased densities in from each window and combine them with the usual WHAM equations [85]. In Sec. 2.8.10 we also show that we can apply the Eigenvector Method for Umbrella Sampling [46] (EMUS) with our quench data. EMUS is a meshless estimator like MBAR and other post-WHAM approaches [45, 46, 86]. In other words, it uses the true bias potentials instead of approximate ones to estimate the weight of a sample generated in one umbrella versus that in another. EMUS is more expensive to apply than WHAM and we did not find the results to be significantly different, hence we demonstrate that it works in the Supporting Information but include results from WHAM in the main text.

We first tested combining these methods by taking starting points generated by the equilibrium US procedure at $T = 300$ K, and quench in the presence of the same harmonic bias with $\gamma_{\text{quench}} = 0.001$ for a forward time of $1\gamma^{-1}$ and a reverse time of $1\gamma^{-1}$, with the idea that we can use our estimator to compute unbiased weighted histograms in each position and then combine the results with any standard free energy approach. We note that this quench time can be thought of as approximately increasing and decreasing the temperature by a factor of e, giving access to temperatures between approximately 110 and 815 K. For each window, we select an upper and lower bound to fix τ_i^+ and τ_i^- as in the previous section.

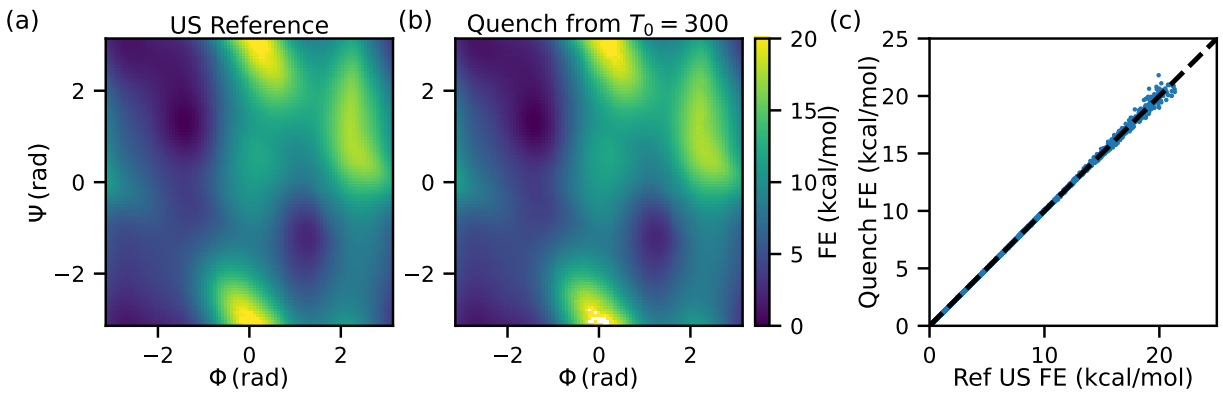

Figure 2.4: A comparison of FES between US and "quench" combined with US. (a) FES computed from US at $T = 300$ K with 400 ns total sampling time after start point generation. (b) FES computed from by quench + US sampling with 400 ns total sampling time including start point generation. Data used in computing Eq. 2.17 totals ≈ 287.6 ns. (c) Comparison of FES values on a bin-by-bin basis. The quench+US landscape agrees almost exactly with the US one.

Fig. 2.4 shows the comparison of FES between US and quench+US. The resulting FES is almost identical between our reference result and our newly computed surface up to the maximum range accessed from US. We also show that this is the case when the quench FES is computed using EMUS, when comparing to a reference FES computed by EMUS or by WHAM in Fig. 2.16 and 2.17, respectively.

Although so far we have reported results for a total equivalent amount of sampling time of 400 ns (with 287.6 ns being used in the estimator), the results of quench+US can be obtained much more quickly than that. In Fig. 2.5 we show that the full FES converges much more quickly than that. In Fig. 2.5(b) we highlight the case of 60 ns (43.2 ns used for the estimator) at which point there is virtually no error and only some very high energy regions are not fully sampled. Fig. 2.5(c) shows that the error is already minimal by $\sim$ 10 ns and stops decreasing by $\sim$ 200 ns. Although the deviation between US and quench+US does not converge to zero, we also show in this plot that this difference is much smaller than that obtained when using a different method (WT-MetaD) or even a different method of estimating the FES from the same reference data (EMUS, open red circles). Thus we do not consider an error of 0.1 kcal/mol over the entire surface to be significant.

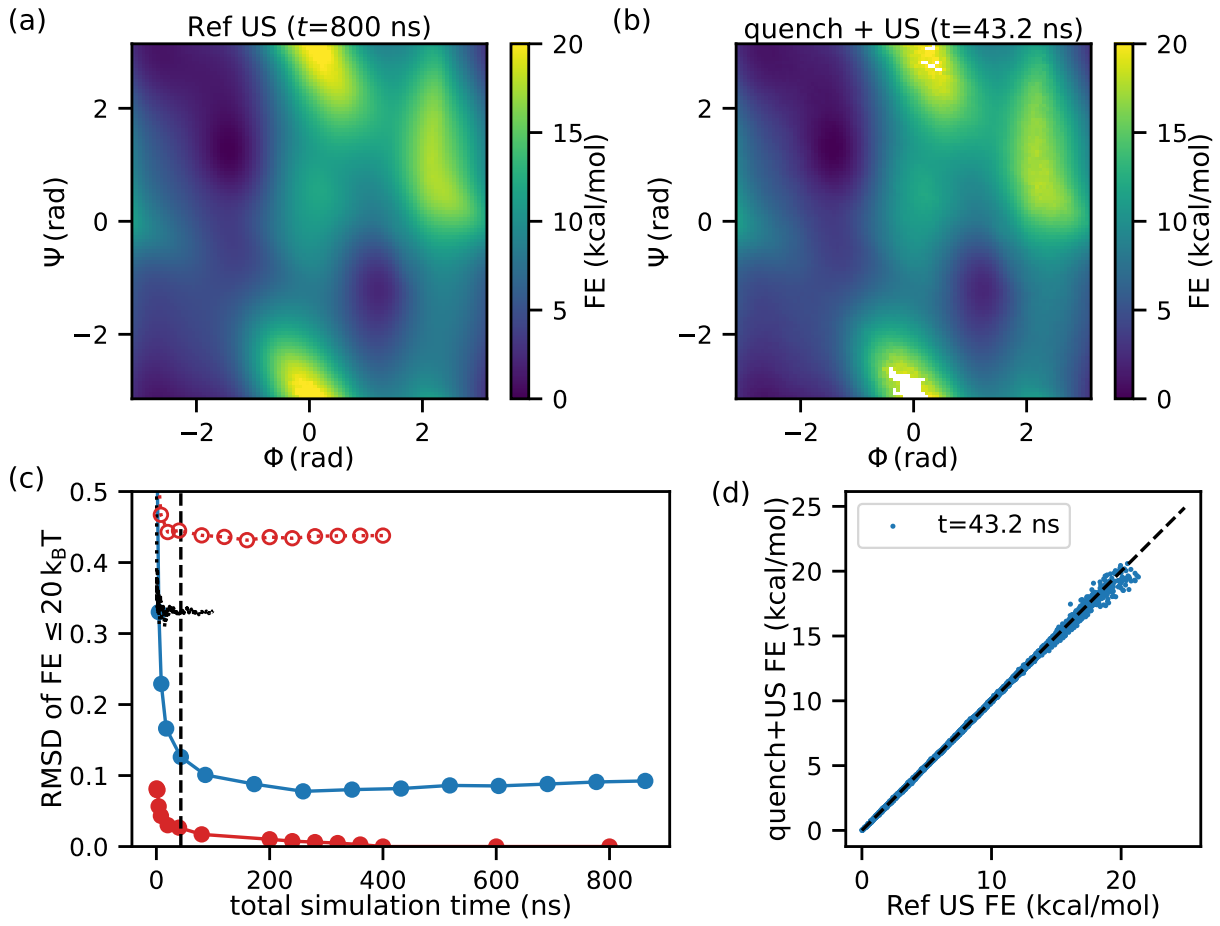

Figure 2.5: (a) Reference US FES with 800 ns of simulation. (b) FES computed from quench + US using 50 quenches per window corresponding to 60 ns of total simulation data, with $\sim$ 41 ns used to compute Eq. 2.17. (c) RMSD of free energy for bins with FE $\leq$ 20 k_BT (in kcal/mol) compared to reference umbrella sampling as a function of total sampling time, which is adjusted by changing the number of starting points used. Black dashed line shows comparison with WT-MetaD using bias factor 6 (see Fig. 2.18), and open red circles show error when computing FES on reference data using EMUS rather than WHAM (see Fig. 2.17). (d) Comparison of FES values in (a) and (b) on a bin-by-bin basis.

Next, we show that quenching helps US when using a bad CV. This is demonstrated by considering the case of US only along ψ, which does not distinguish the positive and negative ϕ basins well. Two sets of simulations using the same US parameters were run, biasing at 20 windows Ψ using the same spring constant of 24.0 kcal/$\left(\text{mol} \cdot \text{rad}^2\right)$. The reference US simulation used 2 ns per window, resulting in 40 total ns of simulation. For quenching, we use 666 starting points separated by 1 ps in each window, with $\gamma_{\text{quench}} = 0.001$ for $\gamma\tau^+ = 1, \gamma\tau^- = -1$ corresponding to an

equivalent amount of simulation time and 28.5 ns used in the estimator. The FES computed from US and WHAM (Fig. 2.6 a) shows that little sampling is achieved. When adding quenching, the heating phase allows the system to overcome some hidden energy barriers that are not captured by bad CV Ψ, resulting in a surface which captures all minima relatively well but does not resolve the barrier between the basins correctly.

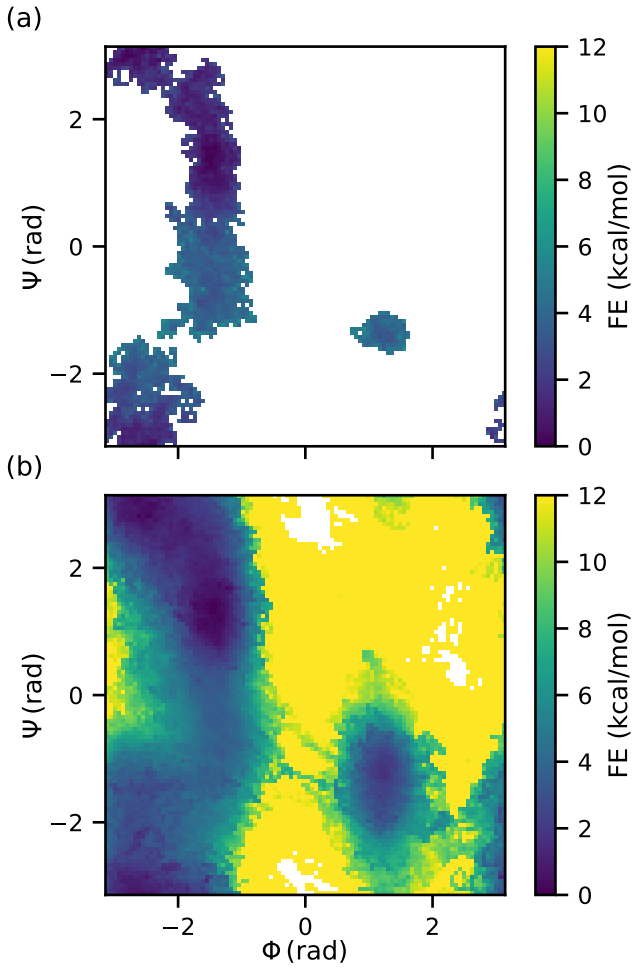

Figure 2.6: FES computed using only one bad CV. (a) FES computed from US on Ψ with 40 ns total sampling. (b) FES computed from quench+US from $T_0 = 1200$ on Ψ with approximately 40 ns total sampling (28.5 ns used in computing Eq. 2.17), showing much more exploration.

Last, we again emphasize that quenching in principle allows estimating FES at various target temperatures using the same simulation data. To quantify this, we tested whether we could ob-

tain the FES at a range of temperatures, quenching from above, below, and in the middle. Using starting points drawn from $T_0 = 75$, 300, and 1200 K, and three quench simulations, we demonstrate that the FES can be obtained for a wide range of temperatures above and below 300 K. In Fig. 2.7 we show the bin-by-bin comparison with US for target temperatures $T = 200$, 300, and 400 K. In all but one case, the results are quite robust across the whole FES. In contrast, US reweighted to other temperatures using the WHAM equations performs poorly in all six cases tested despite using twice the amount of total sampling data. The one case shown where quench+US fails is sampling initial points from $T_0 = 300$ K and estimating at $T = 200$ K, which may be due to insufficient sampling in the chosen initial points at $T_0 = 300$ K that was not evident when estimating at higher temperatures.

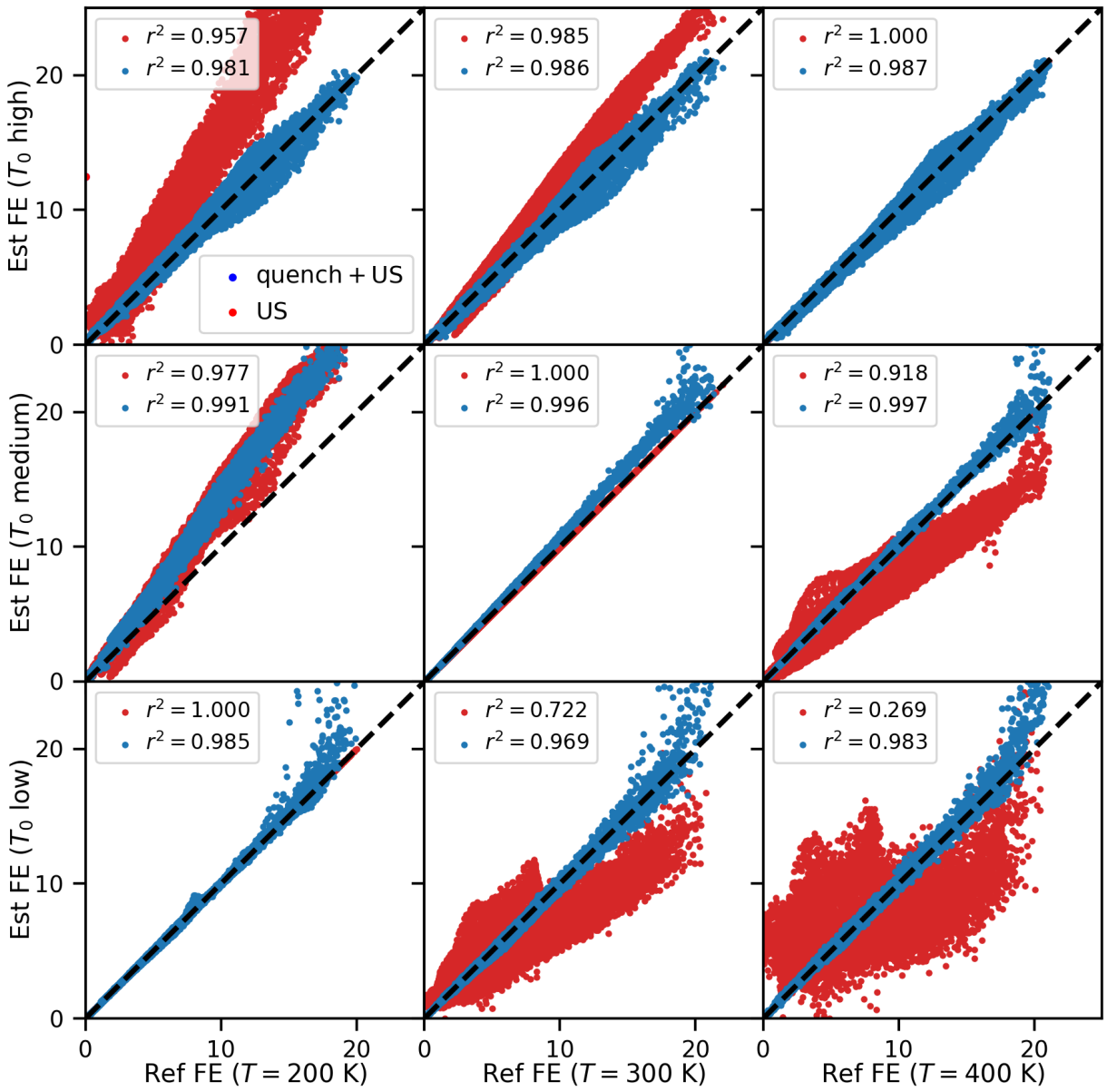

Figure 2.7: Comparison of the quality of FES computed by sampling at T_0 and estimating at T. In each case, the x-axis shows the FES computed by 800 ns of US at a reference temperature of either $T = 200$, 300, or 400 K. The y-axis shows the FES computed at T when starting at a low, medium, or high T_0. For US (red circles), these are T_0 =200, 300, 400 K and for quench+US (blue circles) these are 75, 300, and 1200 K. US alone fails at extrapolating even by 100 K (33%) while quench+US is much more robust. For US, total simulation time is 800 ns and for quench+US, $\gamma_{\text{quench}} = 0.001$ and total sampling time is 800 ns (see Tab. 2.1). To be consistent, the FES for quench+US is computed in all cases by estimating the unbiased density in each case and combining by WHAM, even though we could use our exact WHAM equation for the $T = T_0 = 300$ case.

2.6 Preliminary extension to solvated systems

Most MD simulations are run in solution which adds thousands of additional DOF from solvent molecules. We were concerned that quenching could present difficulties in this case both (a) due to numerical issues arising from the extensivity of the terms in the exponential within our FES estimation formula, and (b) due to the problem of super-heating the solution. To mitigate these issues, we propose a "solute-quench" approach (s-quench) in the spirit of solute tempering [42]. Here we investigate a scheme where the solvent degrees of freedom are evolved by Newtonian dynamics and the solute degrees of freedom follow quench dynamics. Mathematically, this is expressed as,

$$\begin{cases} \dot{\boldsymbol{Q}}_{\text{solute}} = \mathbf{M}_{\text{solute}}^{-1} \boldsymbol{P}_{\text{solute}} \\ \dot{\boldsymbol{P}}_{\text{solute}} = -\boldsymbol{\nabla}_{\text{solute}} U\left(\boldsymbol{Q}_{\text{all}}\right) - \gamma \boldsymbol{P}_{\text{solute}} \\ \dot{\boldsymbol{Q}}_{\text{solvent}} = \mathbf{M}_{\text{solvent}}^{-1} \boldsymbol{P}_{\text{solvent}} \\ \dot{\boldsymbol{P}}_{\text{solvent}} = -\boldsymbol{\nabla}_{\text{solvent}} U\left(\boldsymbol{Q}_{\text{all}}\right) \end{cases} \tag{2.25}$$

For these equations, the Jacobian is still trivial to derive,

$$J\left(t\right) = \exp\left(-d_{\text{solute}} \gamma t\right) \tag{2.26}$$

It is possible to implement this algorithm in LAMMPS with same "fix" by simply choosing two groups of atoms to integrate separately by two different integrators. We tested s-quench on alanine dipeptide in water. Using CHARMM-GUI [87–89] we generated a LAMMPS input file for alanine in a box of dimension 28Å per side with 752 water molecules. As in vacuum, the system is equilibrated for 200 ps in each window, and then the rest of the US protocol is the same. Here, the forcefield used is CHARMM36m. We show in Fig. 2.8 that this approach works and s-quench+US gives identical results to standard US in this case. Interestingly, due to the small size of the solute relative to the solvent, energy fluctuations make the histograms of minimum

and maximum energy overlap. Hence for this case we rely on the infinite time approximation of Eq. 2.18.

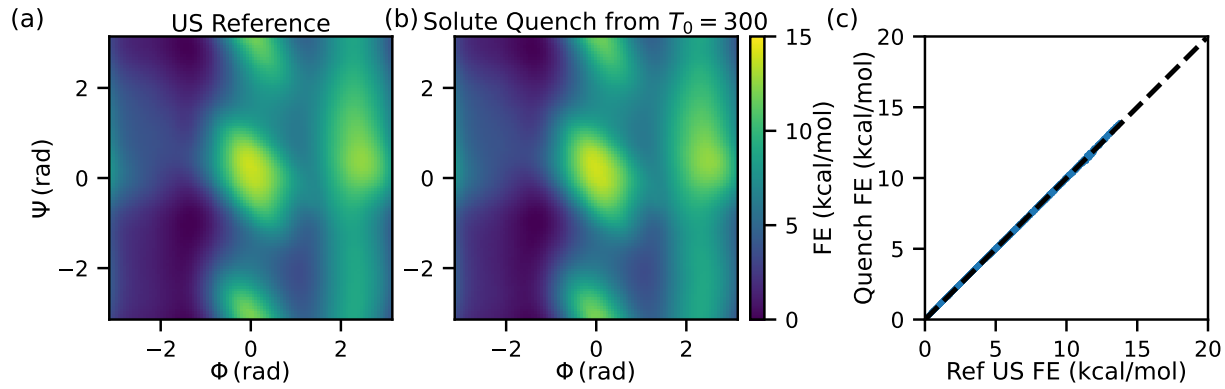

Figure 2.8: A comparison of FES of alanine dipeptide in water between US and solute quench+US. (a) FES computed from umbrella sampling at T =300 K using 800 ns of total data. (b) FES computed from s-quench+US with T_0 = 300 K using 400 ns of total data. (c) Bin-by-bin comparison of FES between these two cases shows exact agreement.

2.7 Conclusions

In the system of independent harmonic springs, a nearly exponential decay of the mean total energy and the mean kinetic energy has been observed in the limit of small γ_{quench}. With this observation, we can use $E(t) = E(0)\, e^{-\gamma t}$ to predict the results of Eqn. 2.10 or Eqn. 2.14. We found that these results give a good heuristic for how long to run quench simulations in practice.

Quenching alone gave moderate performance for alanine in vacuum if the desired quantity is a fast estimation of the entire FES at a single temperature, and would not be the method of choice at least for this simple problem where e.g. metadynamics is extremely efficient as demonstrated in many papers. On the other hand, quenching does give access to the FES at a range of temperatures in a single shot, and so it may be efficient in cases where that is desired. Moreover, we demonstrated that quench combined with US is accurate and efficient across a range of temperatures. Quench+US can also alleviate issues in US due to hidden slow unbiased CVs, since reverse

quenching heats up the system and high free energy regions are more likely to be sampled.

Here, we demonstrated that the WHAM equations also apply to this method when starting temperature T_0 is equal to target temperature T, and showed that it is possible to estimate the FES at other temperatures using a density-based WHAM approach. It is known that WHAM is not the most optimal method of reweighting data, and other approaches could be used to combine quench+US data; e.g. we tried combining US + quench data using EMUS [46], but this was more computationally demanding and resulted in larger errors due to numerics.

Furthermore, we proposed a solute tempering version of quenching so that it can be applied to solution systems where there are many DOF due to the presence of solvent molecules. We showed that a perfect FES can also be obtained by combining this approach with US. As with solute tempering, the major advantage of estimating the FES at other values of T is no longer applicable, and so it remains to be tested whether this approach is more efficient than the equilibrium approach of US combined with solute tempering for a more difficult system.

Given our results, we feel that estimators based on nonequilibrium trajectories can offer an alternative approach worth considering. These approaches may prove to be well suited to certain classes of problems that we have not yet tested, such as computing basin volumes or phase equilibria of simple systems, and we plan to do so going forward.

2.8 Supplementary data

2.8.1 Derivation of the Jacobian $J(t)$

The Jacobian $J(t)$ can be understood as a time-dependent factor that describes how phase space volume changes over time because of the non-equilibrium process. We can compute the change in phase space volume analogous to how it is done for the derivation of Liouville's theorem [19]. First, let us consider an infinitesimal phase space volume at time t, defined by $d+1$ arbitrarily closely spaced points $\boldsymbol{X}(t) = (X_1, ..., X_d)$, $\left\{\boldsymbol{X}_i'(t) = \boldsymbol{X}(t) + \delta\boldsymbol{X}_i(t)\right\}_{i=1,...,d}$. An infinitesimal phase space volume at time t is defined as follows,

$$\delta V(t) = \det\left(\delta\boldsymbol{X}_1(t), ..., \delta\boldsymbol{X}_d(t)\right),$$

where $\delta\boldsymbol{X}_i(t) = \boldsymbol{X}_i'(t) - \boldsymbol{X}(t)$.

After infinitesimal time δt,

$$\begin{aligned}
\boldsymbol{X}_i(t+\delta t) &= \boldsymbol{X}_i(t) + \boldsymbol{b}\left(\boldsymbol{X}(t)\right)\,\delta t + \mathcal{O}\left(\delta t^2\right) \\
\boldsymbol{X}_i'(t+\delta t) &= \boldsymbol{X}_i'(t) + \boldsymbol{b}\left(\boldsymbol{X}_i'(t)\right)\,\delta t + \mathcal{O}\left(\delta t^2\right) \\
&= \boldsymbol{X}_i(t) + \delta\boldsymbol{X}_i(t) + \left[\boldsymbol{b}\left(\boldsymbol{X}(t)\right) + \boldsymbol{\nabla}\boldsymbol{b}\left(\boldsymbol{X}(t)\right)\cdot\delta\boldsymbol{X}_i(t) + \mathcal{O}\left(\left|\delta\boldsymbol{X}_i\right|^2(t)\right)\right]\delta t + \mathcal{O}\left(\delta t^2\right) \\
\delta\boldsymbol{X}_i(t+\delta t) &= \boldsymbol{X}_i'(t+\delta t) - \boldsymbol{X}_i(t+\delta t) \\
&\approx \delta\boldsymbol{X}_i(t) + \left[\boldsymbol{\nabla}\boldsymbol{b}\left(\boldsymbol{X}(t)\right)\cdot\delta\boldsymbol{X}_i(t)\right]\delta t + \mathcal{O}\left(\delta t^2\right) \\
&= \left(\mathbf{I} + \delta t\boldsymbol{\nabla}\boldsymbol{b}\left(\boldsymbol{X}(t)\right)\right)\cdot\delta\boldsymbol{X}_i(t) + \mathcal{O}\left(\delta t^2\right)
\end{aligned}$$

$$\mathbf{I} + \delta t\boldsymbol{\nabla}\boldsymbol{b}\left(\boldsymbol{X}(t)\right) = \begin{pmatrix} 1 + \delta t\frac{\partial b_1(\boldsymbol{X}(t))}{\partial X_1} & \cdots & \delta t\frac{\partial b_1(\boldsymbol{X}(t))}{\partial X_d} \\ \vdots & \ddots & \vdots \\ \delta t\frac{\partial b_d(\boldsymbol{X}(t))}{\partial X_1} & \cdots & 1 + \delta t\frac{\partial b_d(\boldsymbol{X}(t))}{\partial X_d} \end{pmatrix}$$

$$\begin{aligned}
\delta V\left(t+\delta t\right) &= \det\left(\delta \boldsymbol{X}_1\left(t+\delta t\right), ..., \delta \boldsymbol{X}_d\left(t+\delta t\right)\right) \\
&\approx \det\left(\left(\mathbf{I}+\delta t \boldsymbol{\nabla} \boldsymbol{b}\left(\boldsymbol{X}\left(t\right)\right)\right) \cdot \left(\delta \boldsymbol{X}_1\left(t\right), \ldots, \delta \boldsymbol{X}_d\left(t\right)\right)\right) \\
&= \det\left(\mathbf{I}+\delta t \boldsymbol{\nabla} \boldsymbol{b}\left(\boldsymbol{X}\left(t\right)\right)\right) \cdot \det\left(\delta \boldsymbol{X}_1\left(t\right), \ldots, \delta \boldsymbol{X}_d\left(t\right)\right) \\
&= \left(1+\delta t \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(t\right)\right)+\mathcal{O}\left(\delta t^2\right)\right) \delta V\left(t\right)
\end{aligned}$$

The last equation gives at lowest order,

$$\frac{1}{\delta V\left(t\right)} \frac{d}{dt} \delta V\left(t\right) = \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(t\right)\right)$$

and by integrating this equation gives

$$\begin{aligned}
\delta V\left(t\right) &= \delta V\left(0\right) \times \exp\left(\int_0^t \boldsymbol{\nabla} \cdot \boldsymbol{b}\left(\boldsymbol{X}\left(s\right)\right) \, ds\right) \\
&= \delta V\left(0\right) \times J\left(t\right)
\end{aligned}$$

2.8.2 Time evolution of quench equations of motion

Since the functional form of quench is similar to Langevin Dynamics, we use the so-called BAOAB scheme [75] to get an accurate numerical update of coordinates and momenta. We split quench equation of motion into three parts.

- (B) is $\dot{\boldsymbol{P}}\left(t\right) = -\boldsymbol{\nabla} U\left(\boldsymbol{Q}\left(t\right)\right)$
- (A) is $\dot{\boldsymbol{Q}}\left(t\right) = \mathbf{M}^{-1} \boldsymbol{P}\left(t\right)$
- (O) is $\dot{\boldsymbol{P}}\left(t\right) = -\gamma \boldsymbol{P}\left(t\right)$

By iterating these parts in the order BAOA for discrete time steps proportional to δt, time is advanced. We can solve the (O) step analytically to advance time more accurately, while the

other two steps are advanced to first order in δt. This results in a scheme:

$$\begin{cases} (\mathrm{B}) : \boldsymbol{P} \leftarrow \boldsymbol{P} - \boldsymbol{\nabla} U(\boldsymbol{Q}) \cdot \delta t \\ (\mathrm{A}) : \boldsymbol{Q} \leftarrow \boldsymbol{Q} + \mathbf{M}^{-1} \boldsymbol{P} \cdot \dfrac{\delta t}{2} \\ (\mathrm{O}) : \boldsymbol{P} \leftarrow \exp(-\gamma \delta t) \boldsymbol{P} \\ (\mathrm{A}) : \boldsymbol{Q} \leftarrow \boldsymbol{Q} + \mathbf{M}^{-1} \boldsymbol{P} \cdot \dfrac{\delta t}{2} \end{cases}$$

This is implemented in a LAMMPS 'fix' `quench_exponential`, implemented by the code "fix_quench_exponential.cpp" and "fix_quench_exponential.h" available in the github repository for this chapter.

2.8.3 Numerical error in free energy calculation

We compute free energy surfaces on a discrete grid, resulting in a discretization error. First, let us consider free energy along 1D CV $s \in \mathbb{R}$, but this can be generalized to higher dimensional CVs in a similar manner. We assume the true free energy $F(s)$ is differentiable.

$$F(s) = -\frac{1}{\beta} \ln \left(\frac{\int_{\mathbb{R}^{2d}} \delta(S(\boldsymbol{x}) - s) \exp(-\beta \mathscr{H}(\boldsymbol{x})) \, d\boldsymbol{x}}{\int_{\mathbb{R}^{2d}} \exp(-\beta \mathscr{H}(\boldsymbol{x})) \, d\boldsymbol{x}} \right) \tag{2.27}$$

In practice, this delta function is replaced by a window function which is 1 within a range of s and zero outside.

Suppose $\bar{F}(s_0)$ is the average free energy we should get at the specific site s_0 over an interval

with length Δs.

$$\begin{aligned}
\bar{F}(s_0)\,\Delta s &= \int_{s_0-\Delta s/2}^{s_0+\Delta s/2} F(s)\; ds \\
&= \int_{-\Delta s/2}^{\Delta s/2} F(s_0+s)\; ds \\
&= \int_{-\Delta s/2}^{\Delta s/2} F(s_0) + F'(s_0)\, s + \frac{1}{2}F''(s_0)\, s^2 + \mathcal{O}\left(s^3\right)\; ds \\
&= F(s_0)\,\Delta s + \frac{1}{24}F''(s_0)\,\Delta s^3 + \mathcal{O}\left(\Delta s^5\right)
\end{aligned} \tag{2.28}$$

To apply this averaging idea, we can apply it to the probability distribution rather than the free energy itself, resulting in:

$$\begin{aligned}
\exp\left(-\beta\tilde{F}(s_0)\right)\Delta s &= \int_{s_0-\Delta s/2}^{s_0+\Delta s/2} \exp\left(-\beta F(s)\right) ds \\
&= \int_{-\Delta s/2}^{\Delta s/2} \exp\left(-\beta F(s_0+s)\right) ds \\
&= \int_{-\Delta s/2}^{\Delta s/2} \exp\left(-\beta\left[F(s_0) + F'(s_0)\, s + \frac{1}{2}F''(s_0)\, s^2 + \mathcal{O}\left(s^3\right)\right]\right) ds \\
&\approx \exp\left(-\beta F(s_0)\right) \int_{-\Delta s/2}^{\Delta s/2} 1 - \beta F'(s_0)\, s + \frac{1}{2}\left(\beta^2 F'^2(s_0) - \beta F''(s_0)\right) s^2 + \mathcal{O}\left(s^3\right)\; ds \\
&= \exp\left(-\beta F(s_0)\right)\left[\Delta s + \frac{1}{24}\left(\beta^2 F'^2(s_0) - \beta F''(s_0)\right)\Delta s^3 + \mathcal{O}\left(\Delta s^5\right)\right]
\end{aligned} \tag{2.29}$$

From Eq. 2.28 and Eq. 2.29, we find that the true free energy, the mean free energy over an interval, and the free energy computed from the block functions are approximately equal within first order accuracy and the error has a magnitude of $\mathcal{O}\left(\Delta s^2\right)$.

2.8.4 Derivation of WHAM equations

WHAM [44] (Weighted Histogram Analysis Method) is a widely-used technique to reweight data from different windows in umbrella sampling simulations. However, WHAM is valid only

for an equilibrium process, so here we derive a nonequilibrium version of WHAM that applies in this case. We follow the derivation in Ref. [50], and modify some parts to match the "quenching" case.

In the case of regular umbrella sampling simulations, let ρ°_{kl} be the unbiased probability density at window (ϕ_k, ψ_l) which we wish to determine. To compute this, we apply bias potentials $\omega_{ij,kl}$ at a window (ϕ_k, ψ_l) with an additional harmonic potential centered at site $\left(\phi_i, \psi_j\right)$ defined by:

$$\omega_{ij,kl} = \frac{1}{2}\kappa \left[(\phi_i - \phi_k)^2 + \left(\psi_j - \psi_l\right)^2 \right] \tag{2.30}$$

Note that because the bias is applied to dihedral angles, dihedral differences are computed taking into account periodicity of 2π radians. Under the influence of this potential, we measure $n_{ij,kl}$ the 'counts' (number of sampled data points) in the window centered at (ϕ_k, ψ_l).

Since the ratio of probability densities between two windows is fixed, the overall probability density $\rho_{ij,kl}$ at window (ϕ_k, ψ_l) with biased potential at site $\left(\phi_i, \psi_j\right)$ is the linear combination of biased probability densities:

$$\rho_{ij,kl} = c_{ij}\rho^{\circ}_{kl} \exp\left(-\beta_0 \omega_{ij,kl}\right), \tag{2.31}$$

where c_{ij} is the normalization factor at $t = 0$:

$$c_{ij}^{-1} = \sum_{k,l} \rho^{\circ}_{kl} \exp\left(-\beta_0 \omega_{ij,kl}\right) \tag{2.32}$$

The overall probability of getting the sampled data is proportional to the product of these biased probability densities:

$$\begin{aligned} \mathbb{P} &\propto \prod_{i,j} \prod_{k,l} \left(\rho_{ij,kl}\right)^{n_{ij,kl}} \\ &= \prod_{i,j} \prod_{k,l} \left(c_{ij}\rho^{\circ}_{kl} \exp\left(-\beta_0 \omega_{ij,kl}\right)\right)^{n_{ij,kl}} \end{aligned} \tag{2.33}$$

We can estimate the true probability density by maximizing the log-likelihood,

$$\begin{aligned}\frac{\partial \ln \mathbb{P}}{\partial \rho_{kl}^{\circ}} &= \sum_{i,j} n_{ij,kl} \frac{1}{\rho_{kl}^{\circ}} + \sum_{i,j} \sum_{k,l} n_{ij,kl} \frac{1}{c_{ij}} \frac{\partial c_{ij}}{\partial \rho_{kl}^{\circ}} \\ &= \sum_{i,j} n_{ij,kl} \frac{1}{\rho_{kl}^{\circ}} - \sum_{i,j} \left(\sum_{k,l} n_{ij,kl} \right) c_{ij} \exp\left(-\beta_0 \omega_{ij,kl}\right) \\ &= 0\end{aligned} \tag{2.34}$$

Summarizing this equation and normalization condition gives WHAM equations for regular umbrella sampling simulations:

$$\begin{cases} \rho_{kl}^{\circ} = \dfrac{\sum_{i,j} n_{ij,kl}}{\sum_{i,j} \left(\sum_{k,l} n_{ij,kl} \right) c_{ij} \exp\left(-\beta_0 \omega_{ij,kl}\right)} \\ c_{ij}^{-1} = \displaystyle\sum_{k,l} \rho_{kl}^{\circ} \exp\left(-\beta_0 \omega_{ij,kl}\right) \end{cases} \tag{2.35}$$

Similarly, in quenching, let $\rho_{ij,kl,t}$ be probability density at window (ϕ_k, ψ_l) and at time t with biased potential centered at $\left(\phi_i \psi_j\right)$:

$$\rho_{ij,kl,t} = c_{ij} \rho_{kl}^{\circ} \exp\left(-\beta_0 \omega_{ij,kl}\right) \exp\left(-d\gamma t\right), \tag{2.36}$$

where c_{ij} is the normalization factor:

$$c_{ij}^{-1} = \sum_{k,l} \rho_{kl}^{\circ} \exp\left(-\beta_0 \omega_{ij,kl}\right) \tag{2.37}$$

We eventually get exactly the same equation as Eq. 2.35 except for the expression of $n_{ij,kl}$:

$$n_{ij,kl} = \sum_{t} n_{ij,kl,t} \tag{2.38}$$

Note that ρ_{kl}° is unbiased probability density at starting temperature β_0. If we would like to

estimate the unbiased probability ρ°_{kl} at the target temperature β while the simulations are run at a different starting temperature β_0, then there are two ways to compute $n_{ij,kl}$.

In the first approach, we can estimate the density of states $n(\phi_k, \psi_l, E)$ instead with the following relations:

$$\begin{aligned} \exp\left(-\beta F\left(\phi_k, \psi_l\right)\right) &\approx \sum_E n\left(\phi_k, \psi_l, E\right) \exp\left(-\beta E\right) \Delta E \\ &\approx \sum_E n\left(\phi_k, \psi_l, E\right) \exp\left(-\beta E\right) E \Delta \ln E \end{aligned} \tag{2.39}$$

Here the conversion to use $\Delta \ln(E)$ is shown because, in practice, the exponential decay/increase in energy from quench dynamics leads to a very wide range of energy values, and hence it is more computationally convenient to histogram the log of the energy.

$$\begin{cases} n\left(\phi_k, \psi_l, E\right) = \dfrac{\sum_{i,j} n_{ij,kl,E}}{\sum_{i,j}\left(\sum_{k,l,E} n_{ij,kl,E}\right) c_{ij} \exp\left(-\beta_0\left(E + \omega_{ij,kl}\right)\right)} \\ c_{ij}^{-1} = \sum_{k,l,E} n\left(\phi_k, \psi_l, E\right) \exp\left(-\beta_0\left(E + \omega_{ij,kl}\right)\right) \end{cases} \tag{2.40}$$

In the second, we compute "effective counts". We use Eq. 2.17 to estimate the biased probability $\rho_{ij,kl}$, which is normalized counts $n_{ij,kl}$ up to a constant. We can feed these effective counts to the WHAM equations (Eq. 2.35) to estimate $\rho_{ij,kl}$.

Although the first way is mathematically more rigorous, the second method is more convenient and cheaper to implement, and hence we did not use the first approach in this chapter.

2.8.5 Amount of simulation time used in each example

A table listing the amount of effort used for each simulation is given in Tab. 2.1.

Table 2.1: Table of computational cost for original data of each example given in the main text.

System	Figure	Method	Cost/start	Starts	Windows
Harmonic	1,2	Quench	$10 + 2/\gamma_{\text{quench}}$	2000	-
Ala	3a,4a,7	2dUS	1 ns	1	400
Ala	3b	Quench	$1+4\gamma^{-1}$ =41 ps	10^4	-
Ala	4b	Quench+2dUS	3 ps	300	400
Ala	5a,5c	2dUS	2 ns	1	400
Ala	5b	Quench+2dUS	3 ps	50	400
Ala	6a	1dUS	2 ns	1	20
Ala	6b	Quench+1dUS	3 ps	666	20
Ala	7	Quench+2dUS			
		$T_0 = 300$	3 ps	666	400
		$T_0 = 75$	4 ps	500	400
		$T_0 = 1200$	4 ps	500	400
Ala+H_2O	8a	2dUS	2 ns	1	400
Ala+H_2O	8b	S-Quench+2dUS	3 ps	333	400

2.8.6 Convergence of approximation in Eq. 2.18

Fig. 2.9 shows the convergence of the "infinite time" limit of our quench estimator is assessed for a system of harmonic springs in Fig. 2.9 as described in the main text. We also assess the accuracy considering a range of starting temperatures and target temperatures, simulating for a range of times. To make the comparisons fair, we adjust the forward and backward time automatically by taking $\gamma\tau^+ = \gamma\tau_0^+ + \ln(T_0/T)$ and $\gamma\tau^- = \gamma\tau_0^- + \ln(T_0/T)$ so that total simulation time is fixed and the overall temperature range is similar. We show in Fig. 2.10 that accurate results are obtained once total simulation time exceeds our heuristic value of $\gamma_{\text{quench}}(\tau^+ - \tau^-) = \gamma_{\text{quench}}\tau = 2|\ln(T_0/T)|$.

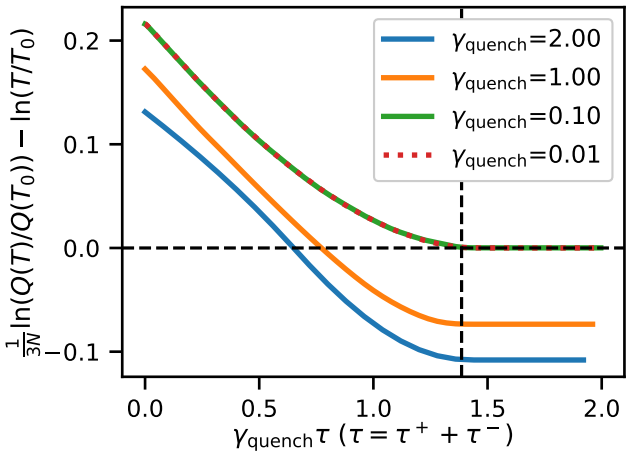

Figure 2.9: Error in the log of the ratio of partition functions at two different temperatures as a function of total quenching time for $\mathcal{N} = 1000$ 3D harmonic oscillators. Quench is performed from 2000 initial samples obtained at $T_0 = 2.0$, estimating the partition function at $T = 1$. For small γ_{quench}, the partition function converges exactly to the theoretical value (horizontal dashed line shows zero error) at $\gamma_{\text{quench}}\tau = 2\ln(T/T_0)$ (vertical dashed line). When the quench is heavily damped, the ratio does not converge to the correct value.

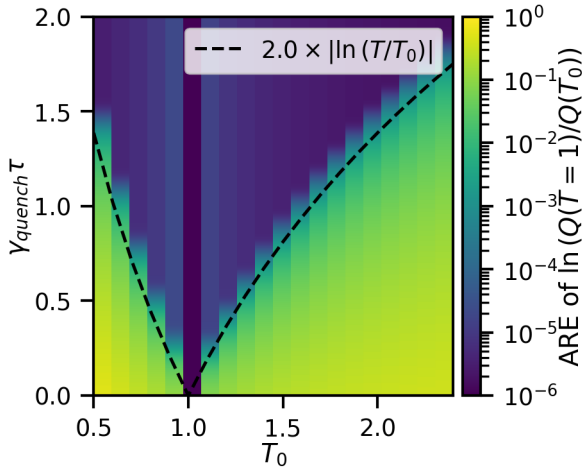

Figure 2.10: Absolute relative error of the log of the ratio of the partition function at $T = 1$ with varying T_0, using $\gamma_{quench} = 0.01$. As T_0 goes away from 1, more sampling time $\gamma_{\text{quench}}\tau$ is required. Total simulation time $\gamma_{\text{quench}}\tau$ indicates quenching forward for $\gamma_{\text{quench}}\tau/2 + \ln(T_0/T)$ and backwards for $-\gamma_{\text{quench}}\tau/2 + \ln(T_0/T)$. Dashed line shows the shape of $2\ln(\beta/\beta_0)$, which is the time of convergence shown in Fig. 2.9 .

2.8.7 Comparison of accuracy of the estimator in Eq. 2.13 with the infinite time approximation and Cao and Vanden-Eijnden variants for harmonic springs

In Fig. 2.11 we compare the accuracy of estimating the ratio of partition functions for $N = 1000$ harmonic springs with quench in two ways, (**1**) the infinite time limit using fixed total sampling time $\tau^+ = 2\gamma^{-1}, \tau^- = -2\gamma^{-1}$ and (**2**) the finite time version of Eq. 2.19 from Ref. [65] for many different combinations of τ^+ and τ^-. The infinite limit gives equivalent accuracy, while the estimator from Ref. [65] can give higher accuracy results for specially chosen quench times, but is less accurate for typical choices, using total simulation time $4\ \gamma^{-1}$.

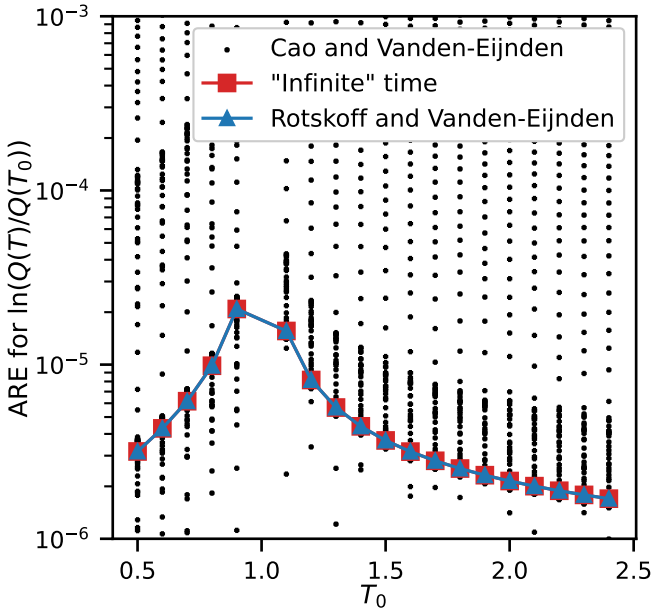

Figure 2.11: Absolute relative error of the log of the partition function ratio for $N = 1000$ and $\gamma = 0.01$ as in Fig. 2.2

2.8.8 Energy cutoffs for alanine quench

Fig. 2.12 shows the energies sampled from quench dynamics for ADP.

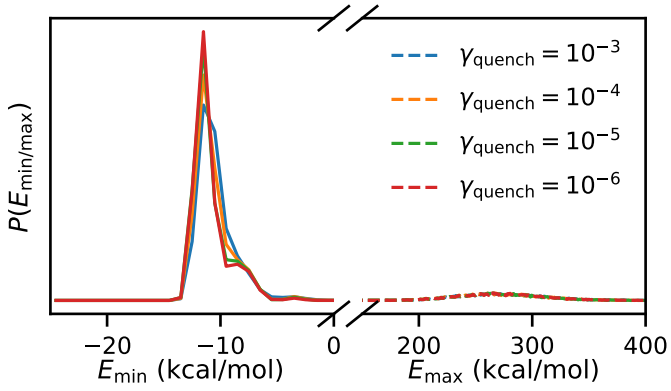

Figure 2.12: Energy histograms for fixed length forward and reverse quenches for alanine dipeptide. The minimum of the upper energies and the maximum of the lower energies were used as cutoffs for Fig. 2.3. Here we started with $T_0 = 1200$ K using $\gamma_{\text{quench}} = 1 \times 10^{-4}$ and 10^4 starting points corresponding to ≈ 251 ns of total simulation time. We chose $T_{\text{min}} = 200$ K as a target lower quench temperature such that, $\gamma\tau^+ = \ln(1200/200) = \ln(6) \approx 1.8$, and $\gamma\tau^- = -2 + \ln(1200/300) = -2 + \ln(4) \approx -0.6$.

2.8.9 Quench alone at different rates for alanine dipeptide

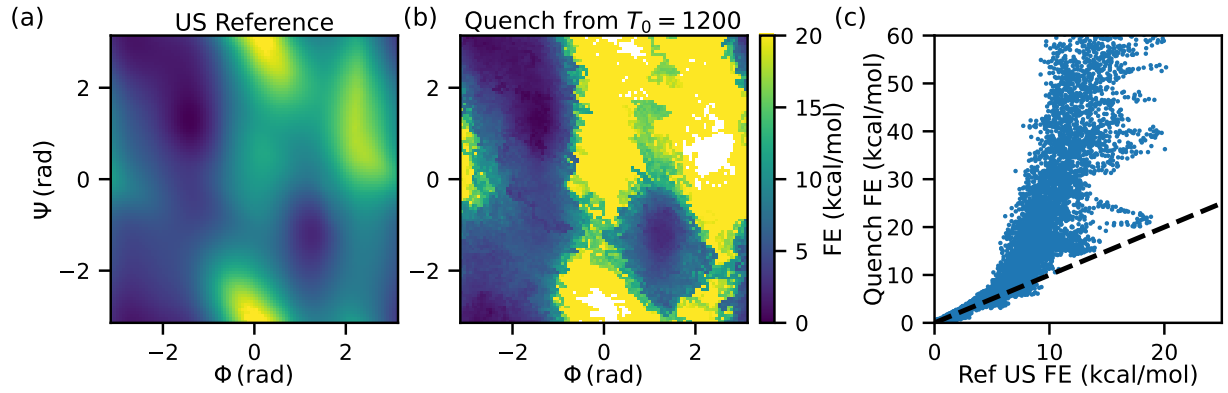

Figure 2.13: This shows the same procedure as in Fig. 2.3, but with $\gamma_{\text{quench}} = 0.001$ (10 times faster). Here, the total amount of sampling time in US is 400 ns and in quench is $\approx$ 50 ns. Data used in computing Eq. 2.17 including generating restart points totals ≈ 33.5 ns.

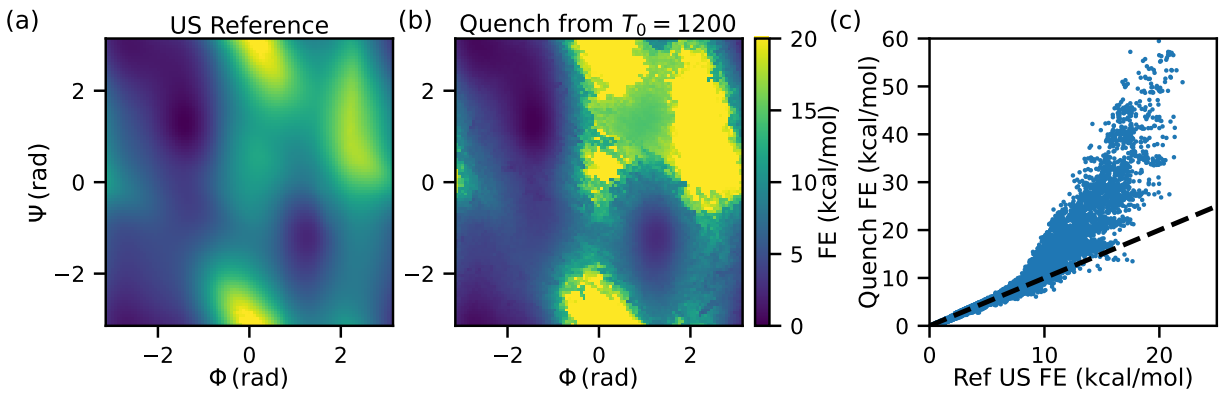

Figure 2.14: This shows the same procedure as in Fig. 2.3, but with $\gamma_{\text{quench}} = 1 \times 10^{-5}$ (10 times slower). Here, the total amount of sampling time in US is 400 ns and in quench is $\approx$ 4010 ns. Data used in computing Eq. 2.17 including generating restart points totals $\approx$ 2529 ns.

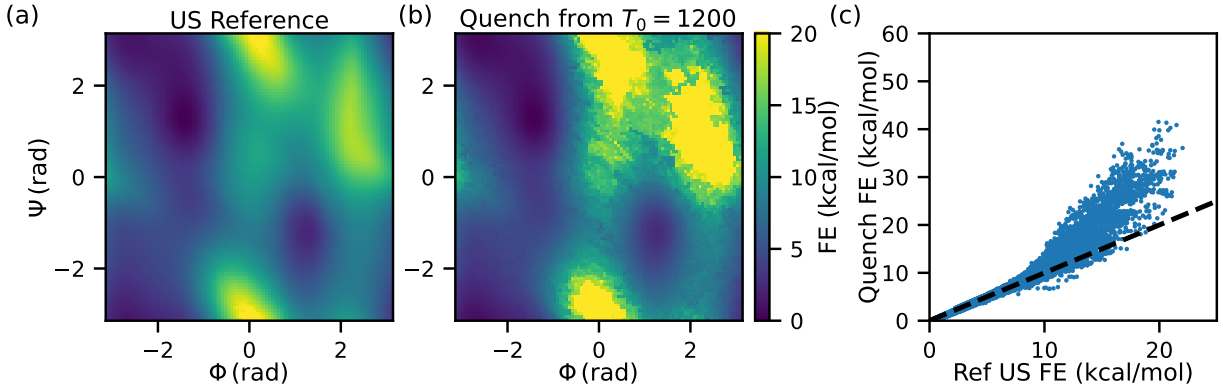

Figure 2.15: This shows the same procedure as in Fig. 2.3, but with $\gamma_{\text{quench}} = 1 \times 10^{-6}$ (100 times slower). Here, the total amount of sampling time in US is 400 ns and in quench is $\approx$ 40010 ns. Data used in computing Eq. 2.17 including generating restart points totals $\approx$ 23997 ns.

2.8.10 Evaluation of FES computed by Quench using EMUS instead of WHAM for alanine dipeptide

Here we show how EMUS [46] works in a normal umbrella sampling simulation and then derive a "quench" version of EMUS using Eq. 2.10. In the EMUS algorithm, we first estimate three

quantities by sample means:

$$\begin{cases} \left\langle \dfrac{\phi}{\sum_k \psi_k} \right\rangle_i \equiv \langle \phi^* \rangle_i \approx \dfrac{1}{N_i} \displaystyle\sum_{t=0}^{N_i-1} \dfrac{\phi\left(\boldsymbol{X}_i\left(t\right)\right)}{\sum_k \psi_k\left(\boldsymbol{X}_i\left(t\right)\right)} \\ \left\langle \dfrac{1}{\sum_k \psi_k} \right\rangle_i \equiv \langle 1^* \rangle_i \approx \dfrac{1}{N_i} \displaystyle\sum_{t=0}^{N_i-1} \dfrac{1}{\sum_k \psi_k\left(\boldsymbol{X}_i\left(t\right)\right)} \\ F_{ij} = \langle F_j \rangle_i \approx \dfrac{1}{N_i} \displaystyle\sum_{t=0}^{N_i-1} \dfrac{\psi_j\left(\boldsymbol{X}_i\left(t\right)\right)}{\sum_k \psi_k\left(\boldsymbol{X}_i\left(t\right)\right)} \end{cases}, \tag{2.41}$$

where ϕ is an observable of interest, $\psi_k\left(\boldsymbol{X}_i\left(t\right)\right) = \frac{1}{2}\kappa\left(S\left(\boldsymbol{X}_i\left(t\right)\right) - s_k\right)^2$ is the exact biased potential given sample data $\boldsymbol{X}_i\left(t\right)$ from the ith biased simulation as if it were sampled from the kth biased simulation. $\mathbf{F}$ is the overlap matrix and F_{ij} is the element of $\mathbf{F}$ in ith row and jth column. With these definitions, we can compute expectations of ϕ by the following expression:

$$\langle \phi \rangle = \frac{\sum_{i=1}^{L} z_i \langle \phi^* \rangle_i}{\sum_{i=1}^{L} z_i \langle 1^* \rangle_i}, \tag{2.42}$$

where z_i solves

$$z_j = \sum_{i=1}^{L} z_i F_{ij}. \tag{2.43}$$

and z is uniquely determined since

$$\sum_{i=1}^{L} z_i = 1. \tag{2.44}$$

In the case of "quench" and umbrella sampling simulation, we use Eq. 2.10 as a bridge to connect an equilibrium average and the data sampled out of equilibrium. We first estimate three

quantities as follows

$$\begin{cases} \langle \phi^* \rangle_i \approx \dfrac{1}{N} \displaystyle\sum_{k=1}^{N} \dfrac{\int \frac{\phi(\boldsymbol{X}_k(t))}{\sum_{l=1}^{N_w} \psi_l(\boldsymbol{X}_k(t))} e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt}{\int e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt} \\ \langle 1^* \rangle_i \approx \dfrac{1}{N} \displaystyle\sum_{k=1}^{N} \dfrac{\int \frac{1}{\sum_{l=1}^{N_w} \psi_l(\boldsymbol{X}_k(t))} e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt}{\int e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt} \\ \langle F_j \rangle_i \approx \dfrac{1}{N} \displaystyle\sum_{k=1}^{N} \dfrac{\int \frac{\psi_j(\boldsymbol{X}_k(t))}{\sum_{l=1}^{N_w} \psi_l(\boldsymbol{X}_k(t))} e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt}{\int e^{-\beta_0 \mathscr{H}_i(\boldsymbol{X}_k(t)) - d\gamma t}\, dt} \end{cases}, \tag{2.45}$$

Here, the integrals are computed from τ_k^- to τ_k^+ as in the main text.

After computing these quantities, we find the left eigenvector z_i that satisfies Eq. 2.43 and use Eq. 2.42 to estimate an arbitrary observable from our quench data. An FES is obtained if we choose $\phi = \delta(S(\boldsymbol{X}) - s)$. Note that a numerical error is introduced since we have used Eq. 2.10 as a bridge, although no numerical error is introduced when making grids as in the case of WHAM. Below, we demonstrate that EMUS does work for the quench results, as compared to an equilibrium EMUS calculation and as compared to WHAM. However, EMUS is much more computationally expensive so we have used WHAM to reweight data from "quench" and umbrella sampling simulations in practice.

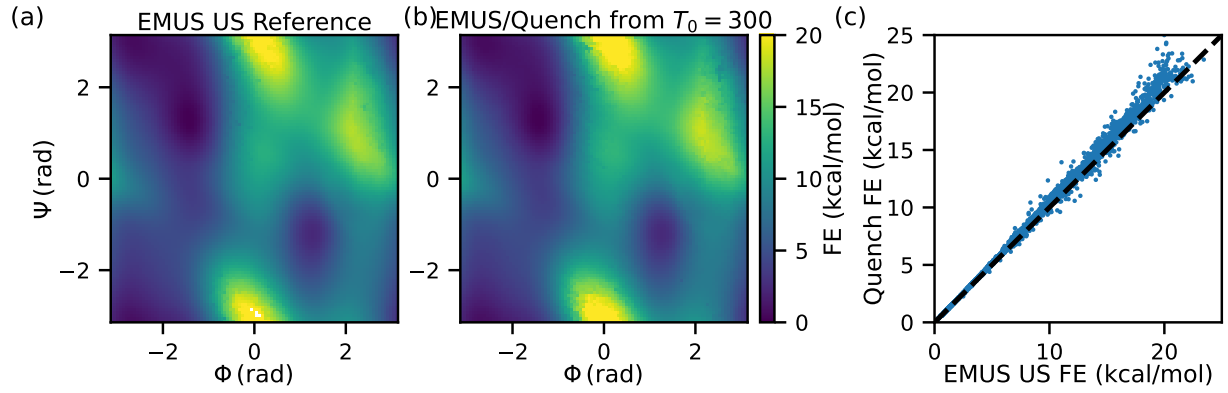

Figure 2.16: A comparison of FES between EMUS and "quench" version of EMUS. (a) FES computed from US at $T = 300$K with 800 ns total sampling time. (b) FES computed from "quench" version of EMUS derived in Sec. 2.8.10 with $\gamma_{\text{quench}} = 0.001$ and 863.1 ns total simulation time.

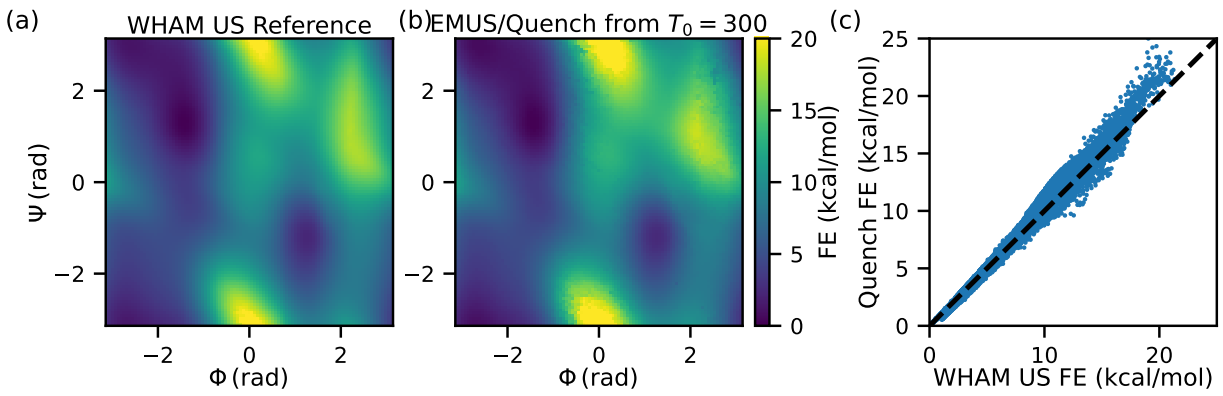

Figure 2.17: Same as Fig. 2.16, but comparing Quench+US/EMUS to the FES computed by WHAM as in the main text.

2.8.11 Comparison of FES computed by US+WHAM with EMUS and Metadynamics for alanine dipeptide

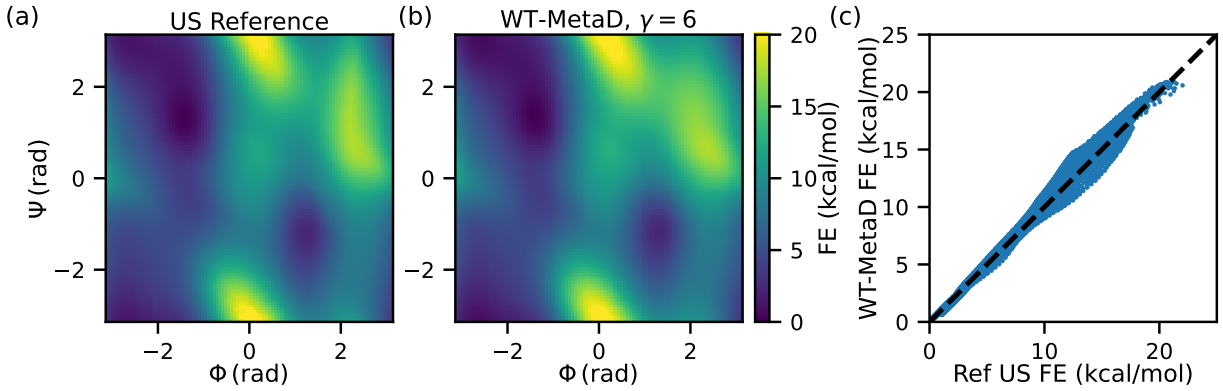

Figure 2.18: Sames as Fig. 2.3, but panel (b) is computed with WT-MetaD with a bias factor of 6, as described in the main text.

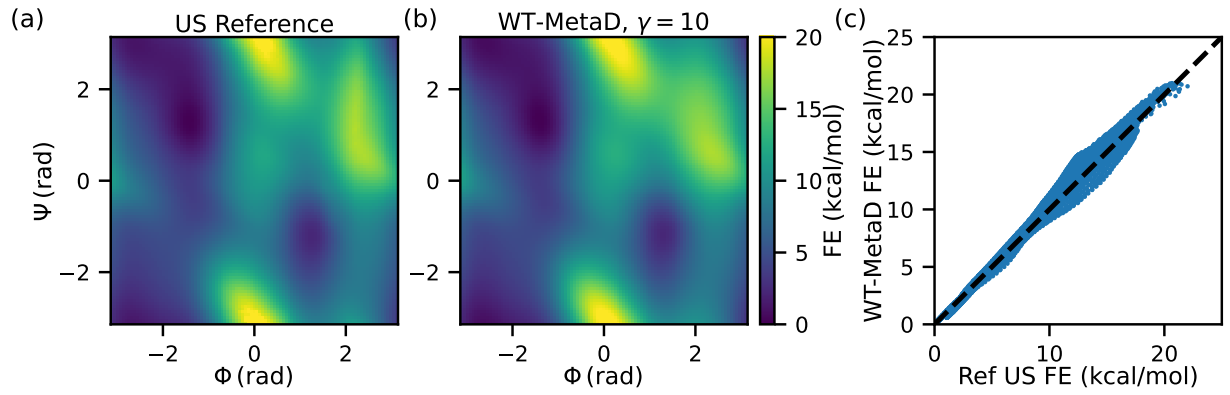

Figure 2.19: Sames as Fig. 2.3, but panel (b) is computed with WT-MetaD with a bias factor of 10, as described in the main text.

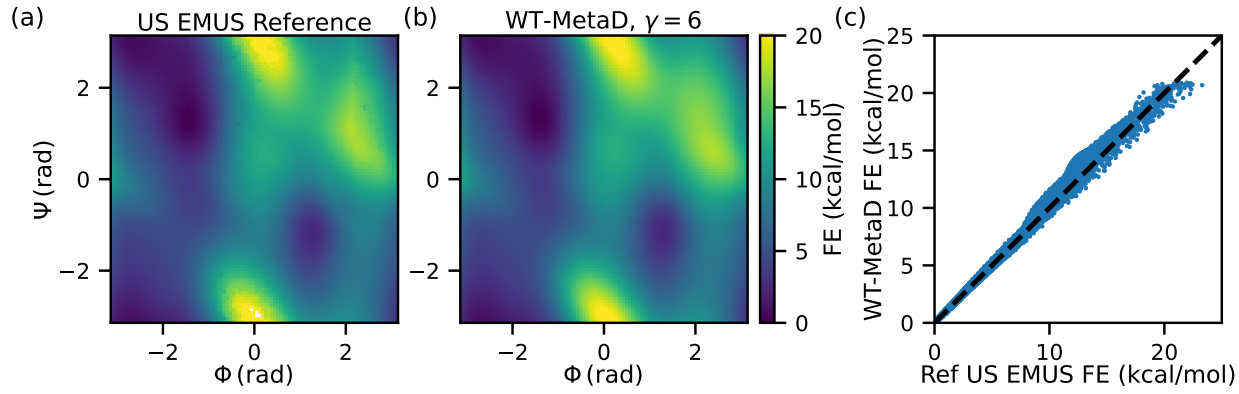

Figure 2.20: Sames as Fig. 2.18, but panel (a) is computed EMUS rather than WHAM.

3 | Optimal unbiased NEIS estimator with given invertible dynamics

This chapter presents an extension to the theory reported in the previous chapter.

3.1 Abstract

An unbiased non-equilibrium importance sampling (NEIS) estimator within an upper boundary and a lower boundary was introduced in the pioneering work from Rotskoff and Vanden-Eijnden [60] and then this theory was further expanded upon by Cao and Vanden-Eijnden [65]. In their work, they categorized these unbiased NEIS estimators into two groups. One is *infinite time* non-equilibrium importance sampling (ITNEIS) estimator where the NEIS estimator within two boundaries can be considered as a special case of ITNEIS estimator and the other is *finite time* non-equilibrium importance sampling (FTNEIS) estimator which is more feasible in real molecular simulations. In this chapter, we extend these ideas and generalize these NEIS estimators with fixed invertible dynamics. Firstly, we show that these generalized NEIS estimators are unbiased. Then we reveal the hidden nature of the unbiased property of NEIS estimators and derive the correct way to compute variance of NEIS estimators. Finally, we prove that there exists a universal optimal unbiased NEIS estimator with given invertible dynamics.

3.2 Introduction

Given a general invertible equation of motion defined on Γ as shown in Eq. 3.1, let $\boldsymbol{X}(t, \boldsymbol{x})$ denote the coordinate starting at $\boldsymbol{x} \in \Gamma$ and following the equation of motion after time t. There is a corresponding Jacobian defined as in Eq. 3.2 that represents how the volume of phase space changes over time through the invertible equation of motion.

$$\frac{d}{dt}\boldsymbol{X}(t, \boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{X}(t, \boldsymbol{x})) \tag{3.1}$$

$$J(t, \boldsymbol{x}) = \exp\left(\int_0^t \boldsymbol{\nabla} \cdot \boldsymbol{b}(\boldsymbol{X}(s, \boldsymbol{x}))\, ds\right) \tag{3.2}$$

There are several properties that are of critical importance.

1. $\boldsymbol{X}(s, \boldsymbol{X}(t, \boldsymbol{x})) = \boldsymbol{X}(s + t, \boldsymbol{x})$
2. $J(s, \boldsymbol{X}(t, \boldsymbol{x}))\, J(t, \boldsymbol{x}) = J(s + t, \boldsymbol{x})$
3. $\tilde{\boldsymbol{x}} \equiv \boldsymbol{X}(t, \boldsymbol{x}) \Rightarrow d\tilde{\boldsymbol{x}} = J(t, \boldsymbol{x})\, d\boldsymbol{x}$

The first property is the definition of invertible equation of motion. The second property shows that the ratio of the final volume of phase space through an invertible equation of motion starting at initial coordinate $\boldsymbol{x}$ after time $s+t$ to the initial volume of phase space is the product of the ratio of the final volume of phase space starting at an intermediate coordinate $\boldsymbol{X}(t, \boldsymbol{x})$ to the intermediate volume of phase space and the ratio of the intermediate volume of phase space after time t to the initial volume of phase space. The third property shows how an infinitesimal volume changes over time through an invertible equation of motion and tells us how to make a change of variable from a mathematical perspective, which is of critical importance in mathematical derivation and proof.

From the work of Ref. [60], an unbiased ITNEIS estimator (Eq. 3.3) was proposed,

$$\phi_{\text{neq}}^{\text{ITNEIS}}(\boldsymbol{x}) \equiv \frac{\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x}))\,\rho_1(\boldsymbol{X}(t,\boldsymbol{x}))\,J(t,\boldsymbol{x})\,dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x}))\,J(s,\boldsymbol{x})\,ds} \tag{3.3}$$

Here, computing an observable follows initiate-then-propagate algorithm. First, several starting points are drawn from starting density ρ_0. Then for each starting point $\boldsymbol{x}$, an invertible possibly non-equilibrium equation of motion is performed both forward in time and backward in time and a trajectory $\boldsymbol{X}(t,\boldsymbol{x})$ is generated from a starting point $\boldsymbol{x}$. Finally, for each starting point $\boldsymbol{x}$, an estimate is calculated by approximating Eq. 3.3 through an average over N separate trajectories generated from starting points $\{\boldsymbol{x}_i\}_{i=1}^{N}$:

$$\langle\phi\rangle_1 \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x}_i))\,\rho_1(\boldsymbol{X}(t,\boldsymbol{x}_i))\,J(t,\boldsymbol{x}_i)\,dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x}_i))\,J(s,\boldsymbol{x}_i)\,ds}. \tag{3.4}$$

Note that the stage of generating trajectories and the stage of calculating estimates can both be done in parallel and considerable time is saved thanks to high performance computers (HPC).

We can show that the estimator is unbiased at any target density ρ_1 (Eq. 3.5),

$$\begin{aligned}
\left\langle\phi_{\text{neq}}^{\text{ITNEIS}}\right\rangle_0 &\equiv \int_\Gamma \rho_0(\boldsymbol{x})\,\phi_{\text{neq}}(\boldsymbol{x})\,d\boldsymbol{x} \\
&= \int_\Gamma \rho_0(\boldsymbol{x}) \frac{\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x}))\,\rho_1(\boldsymbol{X}(t,\boldsymbol{x}))\,J(t,\boldsymbol{x})\,dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x}))\,J(s,\boldsymbol{x})\,ds}\,d\boldsymbol{x} \\
&= \int_\Gamma \int_{-\infty}^{+\infty} \frac{\rho_0(\boldsymbol{X}(-t,\tilde{\boldsymbol{x}}))\,\phi(\tilde{\boldsymbol{x}})\,\rho_1(\tilde{\boldsymbol{x}})}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{X}(-t,\tilde{\boldsymbol{x}})))\,J(s,\boldsymbol{X}(-t,\tilde{\boldsymbol{x}}))\,ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi(\tilde{\boldsymbol{x}})\,\rho_1(\tilde{\boldsymbol{x}}) \int_{-\infty}^{+\infty} \frac{\rho_0(\boldsymbol{X}(-t,\tilde{\boldsymbol{x}}))\,J(-t,\tilde{\boldsymbol{x}})}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s-t,\tilde{\boldsymbol{x}}))\,J(s-t,\tilde{\boldsymbol{x}})\,ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi(\tilde{\boldsymbol{x}})\,\rho_1(\tilde{\boldsymbol{x}}) \int_{-\infty}^{+\infty} \frac{\rho_0(\boldsymbol{X}(-t,\tilde{\boldsymbol{x}}))\,J(-t,\tilde{\boldsymbol{x}})}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(-s,\tilde{\boldsymbol{x}}))\,J(-s,\tilde{\boldsymbol{x}})\,ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi(\tilde{\boldsymbol{x}})\,\rho_1(\tilde{\boldsymbol{x}})\,d\tilde{\boldsymbol{x}} \equiv \langle\phi\rangle_1
\end{aligned} \tag{3.5}$$

By Jensen's inequality [63], it has a smaller variance than the vanilla one when estimating an observable ϕ at starting density ρ_0. In our work (Ref. [66]) reported in the previous chapter, we show how long one should extend simulations to achieve convergence of the ITNEIS estimator. Since we are using finite-time simulations with an infinite time estimator, an error is inevitably introduced and makes the estimator biased, although this numerical error is negligible.

In the work from Ref. [65], an unbiased FTNEIS estimator (Eq. 3.6) was proposed.

$$\phi_{\text{neq}}^{\text{FTNEIS}} \equiv \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\,dt \tag{3.6}$$

Eq. 3.7 shows that the estimator is unbiased at any target density ρ_1.

$$\begin{aligned}
\left\langle\phi_{\text{neq}}^{\text{FTNEIS}}\right\rangle_0 &\equiv \int_\Gamma \rho_0\left(\boldsymbol{x}\right)\int_{\tau^-}^{\tau^+}\frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)ds}\,dt\,d\boldsymbol{x} \\
&= \int_\Gamma\int_{\tau^-}^{\tau^+}\frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)\phi\left(\tilde{\boldsymbol{x}}\right)\rho_1\left(\tilde{\boldsymbol{x}}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)\right)J\left(s,\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma\phi\left(\tilde{\boldsymbol{x}}\right)\rho_1\left(\tilde{\boldsymbol{x}}\right)\int_{\tau^-}^{\tau^+}\frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)J\left(-t,\tilde{\boldsymbol{x}}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(s-t,\tilde{\boldsymbol{x}}\right)\right)J\left(s-t,\tilde{\boldsymbol{x}}\right)ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma\phi\left(\tilde{\boldsymbol{x}}\right)\rho_1\left(\tilde{\boldsymbol{x}}\right)\int_{\tau^-}^{\tau^+}\frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)J\left(-t,\tilde{\boldsymbol{x}}\right)}{\int_{\tau^-}^{\tau^+}\rho_0\left(\boldsymbol{X}\left(-s,\tilde{\boldsymbol{x}}\right)\right)J\left(-s,\tilde{\boldsymbol{x}}\right)ds}\,dt\,d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma\phi\left(\tilde{\boldsymbol{x}}\right)\rho_1\left(\tilde{\boldsymbol{x}}\right)\,d\tilde{\boldsymbol{x}} \equiv \left\langle\phi\right\rangle_1
\end{aligned} \tag{3.7}$$

An optimal flow $\boldsymbol{b}^{\text{opt}}$ condition (Eq. 3.8) of ITNEIS estimator that has the smallest variance is also derived and discussed in that work.

$$\int_{-\infty}^{+\infty}\rho_0\left(\boldsymbol{x}\right)J^{\text{opt}}\left(t,\boldsymbol{x}\right)\,d\boldsymbol{x} = \int_{-\infty}^{+\infty}\rho_1\left(\boldsymbol{x}\right)J^{\text{opt}}\left(t,\boldsymbol{x}\right)\,d\boldsymbol{x} \tag{3.8}$$

This optimal flow condition can be understood as the requirement of implementation of Jensen's inequality [63] when the target density ρ_1 is different from starting density ρ_0. Regarding FTNEIS

estimators, if total simulation step is fixed ($\tau^+ - \tau^- = \tau$ is a constant), and a simulation with a time range $[-2\tau, 2\tau]$ is required by definition, then for any $\tau^- \in [-\tau, 0]$ and $\tau^+ = \tau^- + \tau$, the FTNEIS estimator is unbiased and there is no clear criterion to choose the optimal τ^-, τ^+. Moreover, the functional form of the optimal flow condition is too complicated to be applicable with chosen τ^-, τ^+.

3.3 Generalized NEIS estimator

In practice, a specific equation of motion called "quench" (Eq. 3.9) is used. There are several benefits of this equation of motion.

$$\begin{cases} \dot{\boldsymbol{Q}} = \boldsymbol{M}^{-1}\boldsymbol{P} \\ \dot{\boldsymbol{P}} = -\boldsymbol{\nabla} U\left(\boldsymbol{Q}\right) - \gamma \boldsymbol{P} \end{cases} \tag{3.9}$$

- It is dissipative. In other word, low-energy region is sampled by performing it forward in time and high-energy region is sampled by performing it backward in time. High energy barriers are partially overcome during backward simulations.

$$\dot{E} = \boldsymbol{\nabla} U \cdot \dot{\boldsymbol{Q}} + \boldsymbol{M}^{-1} \cdot \dot{\boldsymbol{P}} = -\gamma \boldsymbol{P} \cdot \boldsymbol{P} \leqslant 0$$

- The functional form of the corresponding Jacobian is simple and it is independent from starting point $\boldsymbol{x}$. Here d is the degree of freedom (DOF) of system.

$$J\left(t, \boldsymbol{x}\right) = \exp\left(-d\gamma t\right)$$

Unlike in Ref. [65], we would like to find a way to optimize unbiased NEIS estimators with given dissipative flow since dissipative flow has so many benefits. As discussed in our work

(Ref. [66]) and the appendix of the previous chapter, there are several unbiased FTNEIS estimators besides the unbiased ITNEIS estimator with given dynamics. For example, a FTNEIS estimator defined in Eq. 3.10,

$$\phi_{\text{neq}} \equiv \frac{1}{\tau^+ - \tau^-} \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\rho_0\left(\boldsymbol{x}\right)} dt \tag{3.10}$$

is unbiased as shown in Eq. 3.11.

$$\begin{aligned} \left\langle \phi_{\text{neq}} \right\rangle_0 &\equiv \int_\Gamma \rho_0\left(\boldsymbol{x}\right) \frac{1}{\tau^+ - \tau^-} \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\rho_0\left(\boldsymbol{x}\right)} dt\, d\boldsymbol{x} \\ &= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^-}^{\tau^+} \frac{1}{\tau^+ - \tau^-} dt\, d\tilde{\boldsymbol{x}} \\ &= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right)\, d\tilde{\boldsymbol{x}} \equiv \left\langle \phi \right\rangle_1 \end{aligned} \tag{3.11}$$

Hence the following question arises naturally: *Is there a generalized class of unbiased NEIS estimators so that we can optimize them to get the optimal unbiased NEIS estimator with the smallest variance?*

A class of generalized NEIS estimators characterized by a virtual Jacobian $\tilde{J}$ satisfying the property $\tilde{J}\left(s, \boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \tilde{J}\left(t, \boldsymbol{x}\right) = \tilde{J}\left(s + t, \boldsymbol{x}\right)$ is defined by Eq. 3.12.

$$\phi_{\text{neq}} \equiv \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t, \boldsymbol{x}\right)\right) J\left(t, \boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-} \rho_0\left(\boldsymbol{X}\left(s, \boldsymbol{x}\right)\right) \tilde{J}\left(s, \boldsymbol{x}\right)\, ds} dt \tag{3.12}$$

We can prove that this class of generalized NEIS estimators is unbiased (Eq. 3.13),

$$
\begin{aligned}
\left\langle \phi_{\text{neq}} \right\rangle_0 &\equiv \int_\Gamma \rho_0\left(\boldsymbol{x}\right) \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-} \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \tilde{J}\left(s,\boldsymbol{x}\right) ds} \, dt \, d\boldsymbol{x} \\
&= \int_\Gamma \int_{\tau^-}^{\tau^+} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right)}{\int_{t-\tau^+}^{t-\tau^-} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right)\right) \tilde{J}\left(s,\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) ds} \, dt \, d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^-}^{\tau^+} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(-t,\tilde{\boldsymbol{x}}\right)}{\int_{t-\tau^+}^{t-\tau^-} \rho_0\left(\boldsymbol{X}\left(s-t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(s-t,\tilde{\boldsymbol{x}}\right) ds} \, dt \, d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^-}^{\tau^+} \frac{\rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(-t,\tilde{\boldsymbol{x}}\right)}{\int_{\tau^-}^{\tau^+} \rho_0\left(\boldsymbol{X}\left(-s,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(-s,\tilde{\boldsymbol{x}}\right) ds} \, dt \, d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \rho_1\left(\tilde{\boldsymbol{x}}\right) \, d\tilde{\boldsymbol{x}} \equiv \left\langle \phi \right\rangle_1
\end{aligned}
\tag{3.13}
$$

A $\tilde{J}$ satisfying such property can be defined by Eq. 3.14, where f is an arbitrary scalar function.

$$
\tilde{J}\left(t,\boldsymbol{x}\right) \equiv \exp\left(\int_0^t f\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \, ds \right) \tag{3.14}
$$

3.4 Physical interpretation of $\tilde{J}$

To understand the physical interpretation of the virtual Jacobian $\tilde{J}$, we derive the general NEIS estimator in another way following the idea from the work [60]. In their work, an unbiased NEIS

estimator is derived as follows,

$$
\begin{aligned}
\langle\phi\rangle_{\mathrm{ne}} &\equiv \int_{\Gamma} \phi\left(\boldsymbol{x}\right) \rho_{\mathrm{ne}}\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&\propto \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\, dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \phi\left(\tilde{\boldsymbol{x}}\right) \int_{-\tau^{+}(\tilde{\boldsymbol{x}})}^{-\tau^{-}(\tilde{\boldsymbol{x}})} \rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) J\left(-t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
&= \int_{\Gamma} \phi\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^{-}(\tilde{\boldsymbol{x}})}^{\tau^{+}(\tilde{\boldsymbol{x}})} \rho_0\left(\boldsymbol{X}\left(t,\tilde{\boldsymbol{x}}\right)\right) J\left(t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
\Rightarrow \rho_{\mathrm{ne}}\left(\boldsymbol{x}\right) &\propto \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt \\
\langle\phi\rangle_0 &= \left\langle \frac{\phi\rho_0}{\rho_{\mathrm{ne}}} \right\rangle_{\mathrm{ne}} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)}{\int_{\tau^{-}(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^{+}(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\right) J\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^{+}(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s+t,\boldsymbol{x}\right)\right) J\left(s+t,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \left\langle \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} dt \right\rangle_0
\end{aligned}
$$

In the first two lines of the proof, a non-equilibrium density ρ_{ne} is expressed by both definition and an average through trajectories starting from points generated by initial density. Each sample point in each trajectory has the same weight 1. One can set $\phi \equiv 1$ to get the normalization factor of the non-equilibrium density ρ_{ne}.

If we do not assume that all weights are 1s and give a path-dependent weight for each sample point in each trajectory $\frac{J(t,\boldsymbol{x})}{\tilde{J}(t,\boldsymbol{x})} \equiv \exp\left(\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) - f\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)\, ds\right)$, then following the

same idea, we can derive a general non-equilibrium density ρ_{ne} as follows.

$$
\begin{aligned}
\langle\phi\rangle_{\mathrm{ne}} &\equiv \int_{\Gamma} \phi\left(\boldsymbol{x}\right) \rho_{\mathrm{ne}}\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&\propto \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \frac{J\left(t,\boldsymbol{x}\right)}{\tilde{J}\left(t,\boldsymbol{x}\right)} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \phi\left(\tilde{\boldsymbol{x}}\right) \int_{-\tau^{+}(\tilde{\boldsymbol{x}})}^{-\tau^{-}(\tilde{\boldsymbol{x}})} \rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(-t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
&= \int_{\Gamma} \phi\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^{-}(\tilde{\boldsymbol{x}})}^{\tau^{+}(\tilde{\boldsymbol{x}})} \rho_0\left(\boldsymbol{X}\left(t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
\Rightarrow \rho_{\mathrm{ne}}\left(\boldsymbol{x}\right) &\propto \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \tilde{J}\left(t,\boldsymbol{x}\right)\, dt \\
\langle\phi\rangle_0 &= \left\langle \frac{\phi\rho_0}{\rho_{\mathrm{ne}}} \right\rangle_{\mathrm{ne}} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \frac{J(t,\boldsymbol{x})}{\tilde{J}(t,\boldsymbol{x})}}{\int_{\tau^{-}(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^{+}(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\right) \tilde{J}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^{+}(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s+t,\boldsymbol{x}\right)\right) \tilde{J}\left(s+t,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_{\Gamma} \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \tilde{J}\left(s,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \left\langle \int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^{-}(\boldsymbol{x})}^{\tau^{+}(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \tilde{J}\left(s,\boldsymbol{x}\right)\, ds} dt \right\rangle_0
\end{aligned}
$$

Besides, this procedure can be extended to the derivation of ITNEIS estimators of any target

density ρ_1 as well as the derivation of FTNEIS estimators of any target density ρ_1 as follows,

$$
\begin{aligned}
\langle \phi \rangle_1 &= \left\langle \frac{\phi \rho_1}{\rho_{\text{ne}}} \right\rangle_{\text{ne}} \\
&= \int_\Gamma \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \frac{J(t,\boldsymbol{x})}{\tilde{J}(t,\boldsymbol{x})}}{\int_{\tau^-(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^+(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\right) \tilde{J}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^-(\boldsymbol{X}(t,\boldsymbol{x}))}^{\tau^+(\boldsymbol{X}(t,\boldsymbol{x}))} \rho_0\left(\boldsymbol{X}\left(s+t,\boldsymbol{x}\right)\right) \tilde{J}\left(s+t,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \tilde{J}\left(s,\boldsymbol{x}\right)\, ds} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \left\langle \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)}{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \tilde{J}\left(s,\boldsymbol{x}\right)\, ds} dt \right\rangle_0
\end{aligned}
$$

$$
\begin{aligned}
\langle \phi \rangle_{\text{ne}} &\equiv \int_\Gamma \phi\left(\boldsymbol{x}\right) \rho_{\text{ne}}\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&\propto \int_\Gamma \int_{\tau^-}^{\tau^+} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \frac{J\left(t,\boldsymbol{x}\right)}{\tilde{J}\left(t,\boldsymbol{x}\right)} dt \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \int_{\tau^-}^{\tau^+} \rho_0\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(-t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi\left(\tilde{\boldsymbol{x}}\right) \int_{-\tau^+}^{-\tau^-} \rho_0\left(\boldsymbol{X}\left(t,\tilde{\boldsymbol{x}}\right)\right) \tilde{J}\left(t,\tilde{\boldsymbol{x}}\right)\, dt\, d\tilde{\boldsymbol{x}} \\
\Rightarrow \rho_{\text{ne}}\left(\boldsymbol{x}\right) &\propto \int_{-\tau^+}^{-\tau^-} \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \tilde{J}\left(t,\boldsymbol{x}\right)\, dt
\end{aligned}
$$

$$
\begin{aligned}
\langle\phi\rangle_1 &= \left\langle \frac{\phi\rho_1}{\rho_{\mathrm{ne}}} \right\rangle_{\mathrm{ne}} \\
&= \int_\Gamma \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\frac{J(t,\boldsymbol{x})}{\tilde{J}(t,\boldsymbol{x})}}{\int_{-\tau^+}^{-\tau^-}\rho_0\left(\boldsymbol{X}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\right)\tilde{J}\left(s,\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right)\,ds} dt\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} \\
&= \int_\Gamma \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{-\tau^+}^{-\tau^-}\rho_0\left(\boldsymbol{X}\left(s+t,\boldsymbol{x}\right)\right)\tilde{J}\left(s+t,\boldsymbol{x}\right)\,ds} dt\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} \\
&= \int_\Gamma \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)\tilde{J}\left(s,\boldsymbol{x}\right)\,ds} dt\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} \\
&= \left\langle \int_{\tau^-}^{\tau^+} \frac{\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)}{\int_{t-\tau^+}^{t-\tau^-}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)\tilde{J}\left(s,\boldsymbol{x}\right)\,ds} dt \right\rangle_0
\end{aligned}
$$

3.5 Performance of NEIS estimators

Since the generalized NEIS estimators are unbiased by Eq. 3.13 and the above derivation, the performance of NEIS estimators is measured by the variance of NEIS estimators. In this section, we show that achieving good performance (low variance) of NEIS estimators is not as simple as we may have thought.

At a first glance, one may think that the variance of NEIS estimators should be computed as follows,

$$
\begin{aligned}
V\left[\phi_{\mathrm{neq}}\right] &= \left\langle \left|\phi_{\mathrm{neq}}\right|^2 \right\rangle_0 - \left\langle \phi_{\mathrm{neq}} \right\rangle_0^2 = \left\langle \left|\phi_{\mathrm{neq}}\right|^2 \right\rangle_0 - \langle\phi\rangle_1^2 \\
&= \int_\Gamma \phi_{\mathrm{neq}}^2\left(\boldsymbol{x}\right)\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} - \langle\phi\rangle_1^2
\end{aligned}
$$

Since $\langle\phi\rangle_1$ is a constant, to optimize $\tilde{J}$, we just need to optimize the first term of the above expression. However, one may notice that we can choose $\tilde{J}$ as large as possible to make the variance as small as possible and this is obviously not true in practice. Therefore we should find another reasonable way to compute the variance of NEIS estimators.

We start from ITNEIS estimators and then extend the idea to FTNEIS estimators. Regarding NEIS estimators, since we are considering trajectories with deterministic dynamics, the whole

space of trajectories can be decomposed into those for a particular set of starting points $\boldsymbol{x}$. If there exists a time t, such that $\boldsymbol{X}(t, \boldsymbol{x}) = \tilde{\boldsymbol{x}}$, then $\tilde{\boldsymbol{x}}$ belongs to the trajectory characterized by starting point $\boldsymbol{x}$. Between two trajectories, either there is no intersection point or every point in one trajectory is in the other trajectory and vice versa. Therefore $\langle \phi \rangle_1$ is made up of contributions of mean value from each trajectory, and by independence of each trajectory, the contribution of variance of the NEIS estimator can be optimized respectively.

In the case of ITNEIS estimators, with given trajectory, we cannot distinguish the starting point among sample points along the trajectory generated by the starting point following a chosen equation of motion or flow. In other words, any sample point along the trajectory can be the generator of the trajectory and we get exactly the same trajectories from these sample points. Hence the contribution of a specific trajectory can be computed as follows,

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} \phi_{\mathrm{neq}}\left(\boldsymbol{X}(t, \boldsymbol{x})\right) \rho_0\left(\boldsymbol{X}(t, \boldsymbol{x})\right) J(t, \boldsymbol{x})\, dt \\
&= \phi_{\mathrm{neq}}(\boldsymbol{x}) \int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}(t, \boldsymbol{x})\right) \tilde{J}(t, \boldsymbol{x})\, dt \\
&= \int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}(t, \boldsymbol{x})\right) \rho_1\left(\boldsymbol{X}(t, \boldsymbol{x})\right) J(t, \boldsymbol{x})\, dt
\end{aligned}
$$

Although sample points are evenly sampled in terms of time, they are not evenly sampled regarding phase space. Hence the $J(t, \boldsymbol{x})$ term in the line integral serves as "width" between two adjacent sample points $\boldsymbol{X}(t, \boldsymbol{x})$ and $\boldsymbol{X}(t + \delta t, \boldsymbol{x})$. A similar idea was introduced by Tuckerman and coworkers [61, 62], where the Jacobian was understood as the ratio of the determinant of some metric tensor which describes the geometry of the phase space. In practice, one would expect sample points distribute exponentially along the trajectory when "quench" dynamics is performed.

The first line of the above equality computes the contribution of the infinitely long trajectory generated by the starting point $\boldsymbol{x}$ from a non-equilibrium perspective: sum of the product of the probability of each sample point along the trajectory as if it were chosen as the starting point

with starting density $\rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)$ with the non-equilibrium estimate at each starting point $\phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)$.

The middle line holds true by the fact that $\phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)=\phi_{\text{neq}}\left(\boldsymbol{x}\right)\tilde{J}\left(t,\boldsymbol{x}\right)$, which represents some kind of detailed balance along the infinitely long trajectory generated by the starting point $\boldsymbol{x}$.

The last line computes the contribution of the trajectory from an equilibrium perspective: sum of the product of the probability of each sample point along the trajectory as if it were drawn as the starting point with target density $\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)$ with the corresponding equilibrium estimate at each starting point $\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)$. We have verified that the contributions of the infinitely long trajectory generated by the starting point $\boldsymbol{x}$ from these two perspectives are the same.

Then the unbiased property of ITNEIS estimators can be understood in two ways as shown in Eq. 3.15. In the first line, the mean value is computed in a conventional way: sum of the non-equilibrium estimate at each starting point $\boldsymbol{x}$ with the probability density $\rho_0\left(\boldsymbol{x}\right)$. In the second line, the mean value is computed in another way in terms of "path integral": sum of the non-equilibrium estimate at each point along the infinitely long trajectory generated by the starting point $\boldsymbol{x}$ (the former part of the expression) with the probability density of the starting point $\rho_0\left(\boldsymbol{x}\right)$ (the latter part of the expression).

$$\begin{aligned}\left\langle\phi\right\rangle_1&=\int_{-\infty}^{+\infty}\phi_{\text{neq}}\left(\boldsymbol{x}\right)\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x}\\&=\int_{\Gamma}\frac{\int_{-\infty}^{+\infty}\phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt}{\int_{-\infty}^{+\infty}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x}\end{aligned}\tag{3.15}$$

Therefore, the variance of the ITNEIS estimator can be computed in the two corresponding ways up to a constant $\left\langle\phi\right\rangle_1^2$ as shown in Eq. 3.16. In the first line, the variance of the ITNEIS estimator $\text{Var}^{(1)}$ is computed in a conventional way: sum of the error function $V_1\left(\boldsymbol{x}\right)$ at each starting point $\boldsymbol{x}$ with the probability density $\rho_0\left(\boldsymbol{x}\right)$. In the second line, the variance of the ITNEIS

estimator $\mathrm{Var}^{(2)}$ is computed in another way: sum of the error function $V_2(\boldsymbol{x})$ at each starting point $\boldsymbol{x}$ (the former part of the expression) with the probability density $\rho_0(\boldsymbol{x})$ (the latter part of the expression). The physical interpretation of the error function V_2 is the mean variance along the infinitely long trajectory generated by the starting point $\boldsymbol{x}$.

$$\begin{aligned}
\mathrm{Var}^{(1)}\left[\phi_{\mathrm{neq}}\right] &= \int_{\Gamma} \left|\phi_{\mathrm{neq}}(\boldsymbol{x})\right|^2 \rho_0(\boldsymbol{x})\, d\boldsymbol{x} \equiv \langle V_1 \rangle_0 \\
\mathrm{Var}^{(2)}\left[\phi_{\mathrm{neq}}\right] &= \int_{\Gamma} \frac{\int_{-\infty}^{+\infty} \left|\phi_{\mathrm{neq}}(\boldsymbol{X}(t,\boldsymbol{x}))\right|^2 \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds} \rho_0(\boldsymbol{x})\, d\boldsymbol{x} \equiv \langle V_2 \rangle_0
\end{aligned} \tag{3.16}$$

3.6 Numerical results

To verify our new theory of computing variance of the generalized NEIS estimators, we test our theory by comparing consistency of accuracy of estimation of ratio of partition functions (Eq. 3.17) at two different temperatures (starting temperature $T_0 = 2.0$ and target temperature $T = 1.0$) in reduced unit and the value of variance calculated by Eq. 3.16.

$$\frac{Q(\beta)}{Q(\beta_0)} = \int_{\Gamma} \rho_0(\boldsymbol{x}) \frac{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \exp\left(-\beta \mathscr{H}(\boldsymbol{X}(t,\boldsymbol{x})) - d\gamma t\right) dt}{\int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} \exp\left(-\beta_0 \mathscr{H}(\boldsymbol{X}(s,\boldsymbol{x}))\right) \tilde{J}(s,\boldsymbol{x})\, ds} d\boldsymbol{x} \tag{3.17}$$

Using LAMMPS [73, 74], we sample $N = 1000$ independent harmonic springs in 3D with identical masses $m = 1.0$ and identical oscillation frequencies $\omega = \sqrt{5}$, in reduced units. We first generate 2000 starting points using Langevin dynamics [80] with friction coefficient $\gamma_{LD} = 0.01$ and time step $\Delta t = 0.001$ in reduced units. To do so, we first equilibrate the system for 10^7 steps ($\tau = 10^4$ in reduced LJ time units) at $T = 2.0, \beta_0 = 0.5$. Then we run production simulation for 2×10^7 steps and save 2000 starting points for further "quench" dynamics simulations with forward unitless time $\gamma\tau^+ = 2.0 + \ln(T_0/T)$ and backward unitless time $\gamma\tau^- = 2.0 - \ln(T_0/T)$ with $\gamma_{\mathrm{quench}} = 0.01$ as suggested from our work in the previous chapter [66]. Simulations un-

der "quench" dynamics are done in parallel by python module Parsl [79]. A lower total energy boundary is set as the maximal value of final total energies $\{\mathscr{H}(\boldsymbol{X}(\tau^+, \boldsymbol{x}_i))\}_{i=1,\ldots,N}$ in forward simulations, and an upper total energy boundary is set as the minimal value of final total energies $\{\mathscr{H}(\boldsymbol{X}(\tau^-, \boldsymbol{x}_i))\}_{i=1,\ldots,N}$ in backward simulations. $\tau^-(\boldsymbol{x})$ is the first time when the trajectory starting from $\boldsymbol{x}$ touches upper boundary in the forward simulation and $\tau^+(\boldsymbol{x})$ is the first time when the trajectory starting from $\boldsymbol{x}$ touches lower boundary in the backward simulation.

In Fig. 3.1, black line shows how accuracy of various generalized unbiased ITNEIS estimators characterized by $\tilde{J} = J^\alpha$ changes with α. Accuracy is measured by absolute relative error of $\ln(Q(\beta)/Q(\beta_0))$. Two red lines show how variances computed by two different ways of generalized unbiased ITNEIS estimators change with α. Red dashed line uses the conventional way (V_1 in Eq. 3.16) to compute the variance of the generalized unbiased ITNEIS estimators while red solid line uses the new way (V_2 in Eq. 3.16) to compute the variance of the generalized unbiased ITNEIS estimators.

Maximal accuracy with minimal absolute relative error of the generalized unbiased ITNEIS estimators is achieved around $\alpha = 1.0$, which suggests that the original unbiased ITNEIS estimator (Eq. 3.3) could be the optimal unbiased NEIS estimator with given dynamics. The conventional variance V_1 reaches maximum around $\alpha = 1.0$, which implies that the original unbiased ITNEIS estimator is very likely to be the worst unbiased ITNEIS estimators among all these generalized unbiased ITNEIS estimators characterized by $\tilde{J} = J^\alpha$, while the new variance V_2 reaches its minimum value around $\alpha = 1.0$, which confirms our observation of accuracy of the generalized unbiased ITNEIS estimators.

Therefore, the new V_2 is more reasonable as a measurement of the variance of the unbiased generalized ITNEIS estimators that correctly measures the performance of the generalized unbiased ITNEIS estimators. It also suggests that the unbiased property of the generalized ITNEIS estimators comes from the unbiased contribution of the infinitely long trajectory generated by

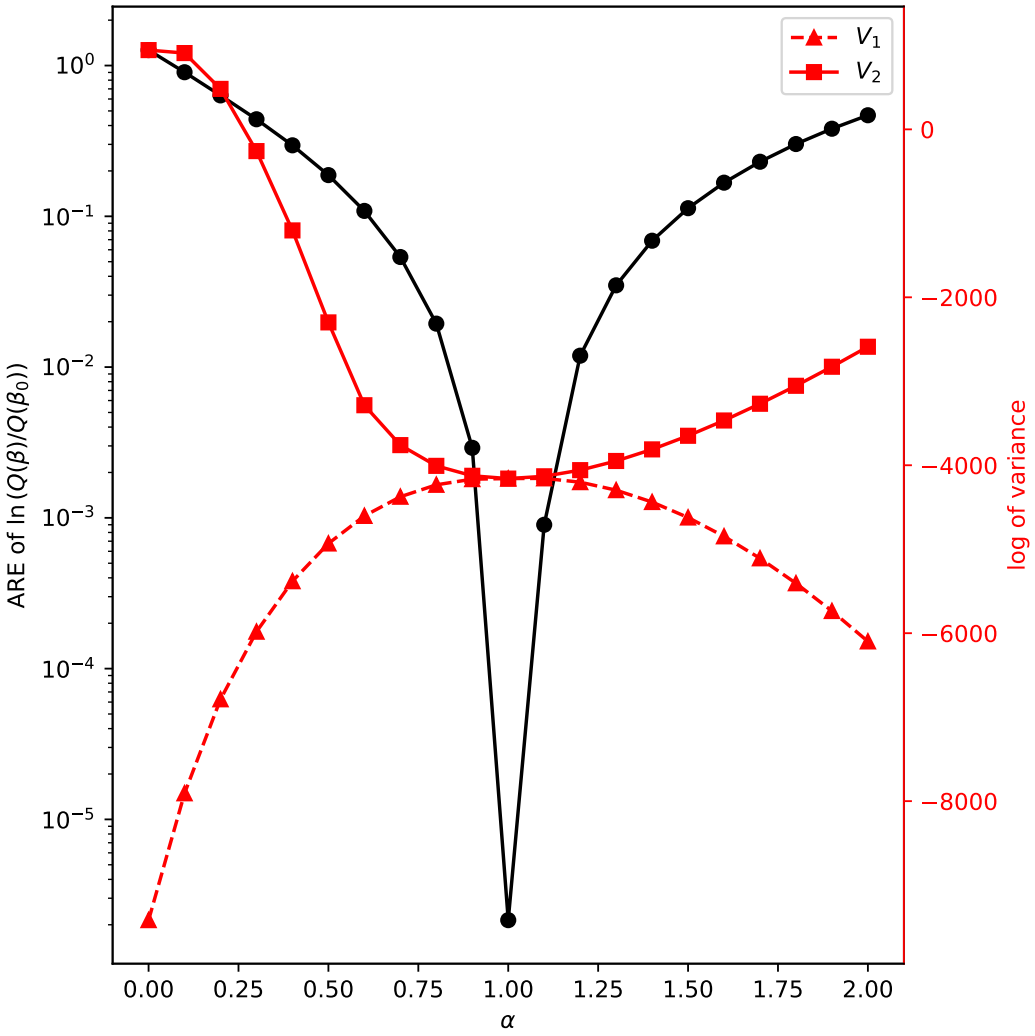

Figure 3.1: Comparison of trend of accuracy (black) of various generalized unbiased ITNEIS estimators characterized by $\tilde{J} = J^{\alpha}$ and trend of value of the corresponding two variances (red) of the generalized unbiased ITNEIS estimators.

any starting point as follows,

$$\langle \phi \rangle_1 = \int_{\Gamma} \frac{\int_{-\infty}^{+\infty} \phi_{\text{neq}} \left(\boldsymbol{X} \left(t, \boldsymbol{x} \right) \right) \rho_0 \left(\boldsymbol{X} \left(t, \boldsymbol{x} \right) \right) J \left(t, \boldsymbol{x} \right) \, dt}{\int_{-\infty}^{+\infty} \rho_0 \left(\boldsymbol{X} \left(s, \boldsymbol{x} \right) \right) J \left(s, \boldsymbol{x} \right) \, ds} \rho_0 \left(\boldsymbol{x} \right) \, d\boldsymbol{x}.$$

3.7 Optimal NEIS estimator

In order to optimize the second measurement V_2 of the variance of the ITNEIS estimator in Eq. 3.16, the expression of ϕ_{neq} is plugged in and the value of $V(\boldsymbol{x})$ at each starting point $\boldsymbol{x}$ is optimized respectively with respect to the virtual Jacobian $\tilde{J}$ as follows,

$$\begin{aligned} V(\boldsymbol{x}) &\equiv \frac{\int_{-\infty}^{+\infty} \left|\phi_{\text{neq}}(\boldsymbol{X}(t,\boldsymbol{x}))\right|^2 \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds} \\ &= \left|\phi_{\text{neq}}(\boldsymbol{x})\right|^2 \frac{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) \frac{\tilde{J}^2(t,\boldsymbol{x})}{J(t,\boldsymbol{x})}\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds} \end{aligned}.$$

$$\phi_{\text{neq}}^{\text{opt}}(\boldsymbol{x}) \equiv \frac{\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x})) \rho_1(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds} \tag{3.18}$$

The minimum is achieved when $\tilde{J}(t,\boldsymbol{x}) = J(t,\boldsymbol{x})$ and we get an optimal ITNEIS estimator $\phi_{\text{neq}}^{\text{opt}}$ defined by Eq. 3.18. This result suggests that the original ITNEIS estimator we have used in the previous chapter is the optimal one with given invertible trajectories. Besides, the measurement of the variance of the optimal ITNEIS estimator V_2 in terms of "path integral" happens to be the same as the conventional measurement of the variance of the optimal ITNEIS estimator V_1.

In the case of FTNEIS estimators, we can still find the same equality along the given infinitely long trajectory. Firstly, we define a FTNEIS estimator with fixed τ^-, τ^+ as follows,

$$\phi_{\text{neq}}^{\tau^-,\tau^+}(\boldsymbol{x}) \equiv \int_{\tau^-}^{\tau^+} \frac{\phi(\boldsymbol{X}(t,\boldsymbol{x})) \rho_1(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})}{\int_{t-\tau^+}^{t-\tau^-} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) \tilde{J}(s,\boldsymbol{x})\, ds}\, dt$$

For simplicity, let $\boldsymbol{X}_i$ denote $\boldsymbol{X}(i\Delta t,\boldsymbol{x})$ and J_i denote $J(i\Delta t,\boldsymbol{x})$. We also let $\tau^- = -\Delta t, \tau^+ = 0$,

and the following result can be extended to any choice of τ^-, τ^+.

$$
\begin{aligned}
&\phi_{\text{neq}}\left(\boldsymbol{X}_0\right)\rho_0\left(\boldsymbol{X}_0\right)J_0 \\
&= \frac{\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0}{\rho_0\left(\boldsymbol{X}_{-1}\right)\tilde{J}_{-1}+\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0}\phi\left(\boldsymbol{X}_{-1}\right)\rho_1\left(\boldsymbol{X}_{-1}\right)J_{-1} + \frac{\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0}{\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0+\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1}\phi\left(\boldsymbol{X}_0\right)\rho_1\left(\boldsymbol{X}_0\right)J_0 \\
&\phi_{\text{neq}}\left(\boldsymbol{X}_1\right)\rho_0\left(\boldsymbol{X}_1\right)J_1 \\
&= \frac{\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1}{\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0+\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1}\phi\left(\boldsymbol{X}_0\right)\rho_1\left(\boldsymbol{X}_0\right)J_0 + \frac{\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1}{\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1+\rho_0\left(\boldsymbol{X}_2\right)\tilde{J}_2}\phi\left(\boldsymbol{X}_1\right)\rho_1\left(\boldsymbol{X}_1\right)J_1 \\
&\Rightarrow \phi_{\text{neq}}\left(\boldsymbol{X}_0\right)\rho_0\left(\boldsymbol{X}_0\right)J_0 + \phi_{\text{neq}}\left(\boldsymbol{X}_1\right)\rho_0\left(\boldsymbol{X}_1\right)J_1 \\
&= \frac{\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0}{\rho_0\left(\boldsymbol{X}_{-1}\right)\tilde{J}_{-1}+\rho_0\left(\boldsymbol{X}_0\right)\tilde{J}_0}\phi\left(\boldsymbol{X}_{-1}\right)\rho_1\left(\boldsymbol{X}_{-1}\right)J_{-1} \\
&+ \phi\left(\boldsymbol{X}_0\right)\rho_1\left(\boldsymbol{X}_0\right)J_0 \\
&+ \frac{\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1}{\rho_0\left(\boldsymbol{X}_1\right)\tilde{J}_1+\rho_0\left(\boldsymbol{X}_2\right)\tilde{J}_2}\phi\left(\boldsymbol{X}_1\right)\rho_1\left(\boldsymbol{X}_1\right)J_1
\end{aligned}
$$

$$
\Rightarrow \sum_{i=-\infty}^{+\infty}\phi_{\text{neq}}\left(\boldsymbol{X}_i\right)\rho_0\left(\boldsymbol{X}_i\right)J_i = \sum_{i=-\infty}^{+\infty}\phi\left(\boldsymbol{X}_i\right)\rho_1\left(\boldsymbol{X}_i\right)J_i
$$

$$
\Rightarrow \int_{-\infty}^{+\infty}\phi_{\text{neq}}^{\tau^-,\tau^+}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt = \int_{-\infty}^{+\infty}\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt
$$

By the equality shown in the last line, we can understand the unbiased property of generalized FTNEIS estimators exactly the same way as we have already done in the case of generalized ITNEIS estimators as follows,

$$
\begin{aligned}
&\int_\Gamma \frac{\int_{-\infty}^{+\infty}\phi_{\text{neq}}^{\tau^-,\tau^+}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt}{\int_{-\infty}^{+\infty}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} \\
&= \int_\Gamma \frac{\int_{-\infty}^{+\infty}\phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt}{\int_{-\infty}^{+\infty}\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} \\
&= \int_\Gamma \phi_{\text{neq}}^{\text{ITNEIS}}\left(\boldsymbol{x}\right)\rho_0\left(\boldsymbol{x}\right)\,d\boldsymbol{x} = \left\langle\phi\right\rangle_1
\end{aligned}
$$

Note that the equality holds when τ^-, τ^+ are fixed within each trajectory and it does not re-

quire all trajectories share the same τ^-, τ^+. The corresponding measurement $V(\boldsymbol{x})$ of the variance of the generalized FTNEIS estimator $\phi_{\text{neq}}^{\tau^-,\tau^+}$ is shown as follows,

$$
\begin{aligned}
V(\boldsymbol{x}) &\equiv \frac{\int_{-\infty}^{+\infty} \left|\phi_{\text{neq}}^{\tau^-,\tau^+}(\boldsymbol{X}(t,\boldsymbol{x}))\right|^2 \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds} \\
&\geqslant \left(\frac{\int_{-\infty}^{+\infty} \phi_{\text{neq}}^{\tau^-,\tau^+}(\boldsymbol{X}(t,\boldsymbol{x})) \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds}\right)^2 \\
&= \left(\frac{\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x})) \rho_1(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt}{\int_{-\infty}^{+\infty} \rho_0(\boldsymbol{X}(s,\boldsymbol{x})) J(s,\boldsymbol{x})\, ds}\right)^2 \\
&= \left|\phi_{\text{neq}}^{\text{opt}}(\boldsymbol{x})\right|^2 \equiv V^{\text{opt}}(\boldsymbol{x})
\end{aligned}
$$

This suggests that whatever $\tau^-, \tau^+, \tilde{J}$ are chosen in the expression of generalized FTNEIS estimator $\phi_{\text{neq}}^{\tau^-,\tau^+}$, the performance of generalized FTNEIS estimators is always no better than the optimal ITNEIS estimator $\phi_{\text{neq}}^{\text{opt}}$ defined in Eq. 3.18. Therefore, the optimal ITNEIS estimator $\phi_{\text{neq}}^{\text{opt}}$ is actually the optimal NEIS estimator among the class of unbiased generalized NEIS estimators with given dynamics.

The unbiased ITNEIS estimator defined by Eq. 3.3 does not happen to be the optimal unbiased NEIS estimators but destined to be the optimal unbiased NEIS estimator. Firstly, we show that unbiased property of NEIS estimators is equivalent to the fact that the contribution of the infinitely long trajectory generated by any starting point is unbiased:

$$
\langle\phi\rangle_1 = \left\langle\phi_{\text{neq}}\right\rangle_0 \Leftrightarrow
$$

$$
\int_{-\infty}^{+\infty} \phi(\boldsymbol{X}(t,\boldsymbol{x})) \rho_1(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt = \int_{-\infty}^{+\infty} \phi_{\text{neq}}(\boldsymbol{X}(t,\boldsymbol{x})) \rho_0(\boldsymbol{X}(t,\boldsymbol{x})) J(t,\boldsymbol{x})\, dt
$$

The forward direction ($\Rightarrow$) is obvious since the generalized NEIS estimators should be unbi-

ased in every trajectory starting at $\boldsymbol{x}$. The backward direction ($\Leftarrow$) is proved as follows,

$$
\begin{aligned}
\langle \phi \rangle_1 &\equiv \int_\Gamma \phi\left(\boldsymbol{x}\right) \rho_1\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \phi\left(\boldsymbol{x}\right) \rho_1\left(\boldsymbol{x}\right) \frac{\int_{-\infty}^{+\infty} \phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds}\, d\boldsymbol{x} \\
&= \int_\Gamma \phi_{\text{neq}}\left(\tilde{\boldsymbol{x}}\right) \rho_0\left(\tilde{\boldsymbol{x}}\right) \int_{-\infty}^{+\infty} \frac{\phi\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) \rho_1\left(\boldsymbol{X}\left(-t,\tilde{\boldsymbol{x}}\right)\right) J\left(-t,\tilde{\boldsymbol{x}}\right)}{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(s-t,\tilde{\boldsymbol{x}}\right)\right) \rho_1\left(\boldsymbol{X}\left(s-t,\tilde{\boldsymbol{x}}\right)\right) J\left(s-t,\tilde{\boldsymbol{x}}\right)\, ds}\, dt\, d\tilde{\boldsymbol{x}} \\
&= \int_\Gamma \phi_{\text{neq}}\left(\tilde{\boldsymbol{x}}\right) \rho_0\left(\tilde{\boldsymbol{x}}\right)\, d\tilde{\boldsymbol{x}} \equiv \left\langle \phi_{\text{neq}} \right\rangle_0
\end{aligned}
$$

With this knowledge, the unbiased property of generalized NEIS estimators and the corresponding variance of these unbiased NEIS estimators can be expressed as follows,

$$
\begin{aligned}
\langle \phi \rangle_1 &\equiv \int_\Gamma \phi\left(\boldsymbol{x}\right) \rho_1\left(\boldsymbol{x}\right)\, d\boldsymbol{x} = \int_\Gamma \phi_{\text{neq}}^{\text{opt}}\left(\boldsymbol{x}\right) \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \frac{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \frac{\int_{-\infty}^{+\infty} \phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x}
\end{aligned}
$$

$$
\begin{aligned}
\text{Var}\left[\phi_{\text{neq}}\right] &\equiv \int_\Gamma \frac{\int_{-\infty}^{+\infty} \left|\phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\right|^2 \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&\geqslant \int_\Gamma \left(\frac{\int_{-\infty}^{+\infty} \phi_{\text{neq}}\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_0\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \right)^2 \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \left(\frac{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) \rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right) J\left(t,\boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right) J\left(s,\boldsymbol{x}\right)\, ds} \right)^2 \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \\
&= \int_\Gamma \left|\phi_{\text{neq}}^{\text{opt}}\left(\boldsymbol{x}\right)\right|^2 \rho_0\left(\boldsymbol{x}\right)\, d\boldsymbol{x} \equiv \text{Var}\left[\phi_{\text{neq}}^{\text{opt}}\right]
\end{aligned}
$$

The equality is achieved when the optimal unbiased NEIS estimator gives a constant non-equilibrium estimate at any point along the given trajectory with given dynamics. Hence the

expression of the optimal unbiased NEIS estimator can be easily derived as follows,

$$\begin{aligned}\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt &= \int_{-\infty}^{+\infty} \phi_{\text{neq}}^{\text{opt}}\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)\rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,dt \\ &= \phi_{\text{neq}}^{\text{opt}}\left(\boldsymbol{x}\right)\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,dt\end{aligned}$$

$$\Rightarrow \phi_{\text{neq}}^{\text{opt}}\left(\boldsymbol{x}\right) = \frac{\int_{-\infty}^{+\infty} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt}{\int_{-\infty}^{+\infty} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds}$$

In practice, it is not possible to get sample points at time $t = \pm\infty$, so it is not a bad idea to use a truncated version (Eq. 3.19) as long as both the numerator and the denominator converge within adequately small tolerance and this can be achieved by a dissipative "quench" dynamics from our work [66] in the previous chapter, although it is not rigorously unbiased.

$$\phi_{\text{neq}}^{\text{opt-like}}\left(\boldsymbol{x}\right) \equiv \frac{\int_{\tau^-}^{\tau^+} \phi\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)\rho_1\left(\boldsymbol{X}\left(t,\boldsymbol{x}\right)\right)J\left(t,\boldsymbol{x}\right)\,dt}{\int_{\tau^-}^{\tau^+} \rho_0\left(\boldsymbol{X}\left(s,\boldsymbol{x}\right)\right)J\left(s,\boldsymbol{x}\right)\,ds} \tag{3.19}$$

3.8 Conclusions

Inspired by the idea of Ref. [65], we generalized unbiased NEIS estimators characterized by a virtual Jacobian $\tilde{J}$ with given dynamics. In the course of study of the generalized NEIS estimators, we realized that the traditional way of computing variance of generalized NEIS estimator was incorrect and we have proposed the correct way of calculating variance of the generalized NEIS estimators in the second line of Eq. 3.16. With the correct way of computing variance of the generalized NEIS estimators, we have proved that there exists an optimal NEIS estimator $\phi_{\text{neq}}^{\text{opt}}$ (Eq. 3.18) among the class of unbiased generalized NEIS estimators with given dynamics, which is exactly the same as what we used in the previous chapter. Moreover, with further understanding of the unbiased property of the generalized NEIS estimators as well as the correct way of computing variance of the generalized NEIS estimators, this unbiased NEIS estimator is destined to

be the optimal one by Jensen's inequality [63]. This is a bittersweet story. The happy side is that there is no need to optimize FTNEIS estimators with given dynamics while the sad side is that we do not obtain a better unbiased NEIS estimator than what we currently use and the only way to improve the performance of the unbiased NEIS estimators is to find better dynamics or flow, which is not a trivial or easy task. In the end, our theory provides a suggestion on when to terminate simulation under "quench" dynamics and a corresponding good NEIS estimator defined in Eq. 3.19.

4 | Computing basin volume by density propagation method

This chapter is an extension of the Ch. 5 of Ref. [71], and has been performed in close discussion with Prof. Stefano Martiniani and Dr. Mathias Casiulis.

4.1 Abstract

Computation of basin volume can be used as a direct computation of configurational entropy by so-called "mean basin volume" (MBV) method. A random walk with several starting points is used to compute basin volume. For each starting point, a process of energy minimization using gradient descent dynamics is used to determine which basin the point belongs to. Data generated via such a process is mostly wasted since only the final point of the trajectory of gradient descent dynamics is used to determine the basin. With the idea of density propagation, all sample points along the trajectory of gradient descent dynamics can be used to determine the basin center as well as to compute basin volume. By drawing starting points from the surface of a hyperball centered at basin center with proper radius, all the starting points are accepted inside the basin volume and hence efficiency is further improved. We show that the method works well in high dimensional gaussian system. In hypercube system where there are many spikes, an "umbrella sampling"-like dynamics can be used to improve the performance and achieve a decent accuracy.

4.2 Introduction

Entropy [17, 19, 90, 91], defined as a measure of the number of accessible microstates, is the key quantity to understand the mechanism of emergent behavior such as phase transitions. Computational cost of direct sampling of microstates is prohibitively expensive. A promising method, the so-called "mean basin volume" (MBV) method [92–95], gives a direct computation of the configurational entropy of a system. The idea of the MBV method is as follows. The whole space can be considered as a composition of basins of attraction. Here, a basin of attraction is a collection of points that eventually fall into the same point (considered as the center of basin) under gradient descent. With this simple relationship, the number of basin of attraction Ω can be expressed by mean basin volume $\langle v \rangle$ with accessible volume $\mathcal{V}$ known apriori,

$$\mathcal{V} = \sum_{i=1}^{\Omega} v_i = \Omega \langle v \rangle . \tag{4.1}$$

Then the remaining challenge is how to make an estimate of the volume of basin of attraction efficiently in high dimensions since for each sample point, a gradient descent dynamics is performed to verify which basin of attraction the starting point belongs to, and almost all the points along the trajectory of gradient descent dynamics are wasted. Here we use so-called density propagation method, which make use of novel mathematical results, so that all the points along the trajectory of gradient descent dynamics (to verify the starting point belongs to the basin of attraction) as well as the trajectory of gradient ascent dynamics (to explore as large space of the basin of attraction as possible) contribute to the estimate of volume of the basin of attraction.

4.3 Density propagation method

4.3.1 Starting from a known volume

Consider a d-dimensional system with a potential energy $U(\boldsymbol{x})$ as a scalar function of coordinates $\boldsymbol{x} \in \mathbb{R}^d$. A basin of attraction is characterized by a local minimum $\boldsymbol{c}$ of the potential energy $U(\boldsymbol{x})$. A point $\boldsymbol{x}$ is assigned to the basin if the point $\boldsymbol{x}$ converges to the local minimum $\boldsymbol{c}$ of the potential energy $U(\boldsymbol{x})$ following gradient descent dynamics (Eq. 4.2), where $\boldsymbol{X}(t, \boldsymbol{x}) \in \mathbb{R}^d$ is the coordinate after time t following an invertible dynamics starting at $\boldsymbol{x} \in \mathbb{R}^d$. Note that $\boldsymbol{X}(t, \boldsymbol{x}) \to \boldsymbol{c}$ as $t \to +\infty$.

$$\frac{d}{dt}\boldsymbol{X}(t, \boldsymbol{x}) = -\boldsymbol{\nabla} U(\boldsymbol{X}(t, \boldsymbol{x})) \tag{4.2}$$

Following Ch. 3, the corresponding Jacobian of the gradient descent dynamics is defined by Eq. 4.3, where Tr is the trace of a matrix and $\mathbf{H}$ is the hessian matrix,

$$J(t, \boldsymbol{x}) = \exp\left(\int_0^t -\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} U(\boldsymbol{X}(t, \boldsymbol{x}))\, ds\right) = \exp\left(\int_0^t -\operatorname{Tr}\left(\mathbf{H}\left\{U(\boldsymbol{X}(t, \boldsymbol{x}))\right\}\right)\, ds\right) \tag{4.3}$$

Let Γ denote all points $\boldsymbol{x}$ that belongs to the basin characterized by the local minimum $\boldsymbol{c}$. In that case, the volume of Γ is given by $V_\Gamma = \int_{\mathbb{R}^d} 1_\Gamma(\boldsymbol{x})\, d\boldsymbol{x}$.

A gradient ascent dynamics, which can be understood as performing gradient descent dynamics backward in time, is used to determine the boundary of the basin. Also, note that $\boldsymbol{X}(t, \boldsymbol{x}) \to \boldsymbol{y} \in \partial\Gamma$ as $t \to -\infty$.

In practice, it is not possible to run simulations infinitely long. Therefore, we enforce that a simulation is terminated at a specific step if it meets some finite-time convergence criterion such as $|\boldsymbol{\nabla} U(\boldsymbol{x})| \leqslant \epsilon$, where ϵ is a predefined tolerance. The local minimum $\boldsymbol{c}$ of the potential energy $U(\boldsymbol{x})$ is then estimated by the mean of all final coordinates.

A reference region $\gamma \subset \Gamma$ with a volume V_γ that contains the local minimum $\boldsymbol{c}$ is introduced.

Here we choose γ as a d-dimensional hyperball centered at $\boldsymbol{c}$ since the volume V_γ and the surface area $S_{\partial\gamma}$ of the reference region γ can be calculated analytically [96–98]. Since we do not know the local minimum $\boldsymbol{c}$ exactly from simulations, the radius of the hyperball cannot be set too small. On the other hand, the radius of hyperball cannot be set too large, since not all points in the hyperball belong to the basin of attraction if the local minimum lies near the boundary of the basin of attraction. If a starting point is drawn in the region $\Gamma - \gamma = \{\boldsymbol{x} | \boldsymbol{x} \in \Gamma, \boldsymbol{x} \notin \gamma\}$, then there exists and only exists a positive finite time $\tau^+(\boldsymbol{x})$, such that $\boldsymbol{X}(\tau^+(\boldsymbol{x}), \boldsymbol{x}) \in \partial\gamma$ since every point in the basin Γ inevitably falls into the basin center (local minimum $\boldsymbol{c}$). The non-equilibrium density $\rho(\boldsymbol{x}, t)$ is governed by the mass conservation law (Eq. 4.4) with initial condition defined by Eq. 4.5.

$$\frac{\partial}{\partial t}\rho(\boldsymbol{x}, t) = \boldsymbol{\nabla} \cdot (\rho(\boldsymbol{x}, t)\, \boldsymbol{\nabla} U(\boldsymbol{x})) \tag{4.4}$$

$$\rho_0(\boldsymbol{x}) \equiv \rho(\boldsymbol{x}, 0) = \begin{cases} \dfrac{1}{V_{\Gamma-\gamma}} & \boldsymbol{x} \in \Gamma - \gamma \\ 0 & \boldsymbol{x} \in \gamma \end{cases} \tag{4.5}$$

Now consider the total flux through the reference region over time $\Phi_{\partial\gamma}$ (Eq. 4.6), where $\hat{\boldsymbol{n}}(\boldsymbol{x})$ is the unit vector normal to the reference surface $\partial\gamma$ at $\boldsymbol{x} \in \partial\gamma$. Since starting points are drawn in the region $\Gamma - \gamma$, only forward simulation with positive times contributes to the total flux.

$$\Phi_{\partial\gamma} \equiv \oint_{\partial\gamma} \int_0^{+\infty} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot (\rho(\boldsymbol{x}, t)\, \boldsymbol{\nabla} U(\boldsymbol{x}))\, dt\, d\boldsymbol{x} \tag{4.6}$$

On the other hand, all the change of mass comes from the flux through the reference surface and eventually all the mass should be inside the reference region γ, which gives us the following

result by the divergence theorem [99, 100],

$$\begin{aligned}
1 &= \int_{\gamma} \rho\left(\boldsymbol{x}, +\infty\right) - \rho\left(\boldsymbol{x}, 0\right)\, d\boldsymbol{x} \\
&= \int_{\gamma} \int_{0}^{+\infty} \frac{\partial}{\partial t} \rho\left(\boldsymbol{x}, t\right)\, dt\, d\boldsymbol{x} \\
&= \int_{\gamma} \int_{0}^{+\infty} \boldsymbol{\nabla} \cdot \left(\rho\left(\boldsymbol{x}, t\right) \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right)\, dt\, d\boldsymbol{x} \\
&= \oint_{\partial\gamma} \int_{0}^{+\infty} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \left(\rho\left(\boldsymbol{x}, t\right) \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right)\, dt\, d\boldsymbol{x} \equiv \Phi_{\partial\gamma}
\end{aligned}$$

Moreover, the mass density at position $\boldsymbol{x}$ and time t is generated by the initial density ρ_0 in the region $\Gamma - \gamma$ being transported by the gradient descent dynamics, which gives

$$\rho\left(\boldsymbol{x}, t\right) = \int_{\Gamma-\gamma} \rho_0\left(\boldsymbol{y}\right) \delta\left(\boldsymbol{X}\left(t, \boldsymbol{y}\right) - \boldsymbol{x}\right)\, d\boldsymbol{y}$$

Summarizing these results, we have the key equality (Eq. 4.7) in this chapter.

$$\begin{aligned}
1 &= \oint_{\partial\gamma} \int_{0}^{+\infty} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \left(\rho\left(\boldsymbol{x}, t\right) \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right)\, dt\, d\boldsymbol{x} \\
&= \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} \int_{\Gamma-\gamma} \rho_0\left(\boldsymbol{y}\right) \delta\left(\boldsymbol{X}\left(t, \boldsymbol{y}\right) - \boldsymbol{x}\right)\, d\boldsymbol{y}\, dt\, d\boldsymbol{x} \\
&= \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} \int_{\Gamma-\gamma} \rho_0\left(\boldsymbol{X}\left(-t, \tilde{\boldsymbol{y}}\right)\right) \delta\left(\tilde{\boldsymbol{y}} - \boldsymbol{x}\right) J\left(-t, \tilde{\boldsymbol{y}}\right)\, d\tilde{\boldsymbol{y}}\, dt\, d\boldsymbol{x} \\
&= \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} \rho_0\left(\boldsymbol{X}\left(-t, \boldsymbol{x}\right)\right) J\left(-t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x}
\end{aligned} \tag{4.7}$$

Note that from the second line to the third line in Eq. 4.7, we have made a change of variable $\tilde{\boldsymbol{y}} \equiv \boldsymbol{X}\left(t, \boldsymbol{y}\right)$ and by Property 3 of invertible dynamics, we get the result of the third line. The last line comes from one of Dirac delta function [101] properties.

We can define a surface density $\nu\left(\boldsymbol{x}\right)$ on the reference surface $\partial\gamma$ by Eq. 4.8.

$$\nu\left(\boldsymbol{x}\right) \equiv \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} \rho_0\left(\boldsymbol{X}\left(-t, \boldsymbol{x}\right)\right) J\left(-t, \boldsymbol{x}\right)\, dt \tag{4.8}$$

By the last line of Eq. 4.7, this surface density $\nu(\boldsymbol{x})$ is normalized. The physical interpretation of $\nu(\boldsymbol{x})$ is the contribution of mass density from a backward trajectory (gradient ascent dynamics), which is the same as counting all the mass drawn under some initial density ρ_0 in the region $\Gamma - \gamma$ and eventually stopping at the point $\boldsymbol{x}$ on the reference surface $\partial\gamma$ from a forward trajectory (gradient descent dynamics). This explains where the term "density propagation" comes from.

Following this idea, the mean value of a test function $f(\boldsymbol{x})$ on the reference surface $\partial\gamma$ with probability $\nu(\boldsymbol{x})$ can be expressed by the sum of mass contribution of all trajectories that stops at the point $\boldsymbol{x}$ on the reference surface $\partial\gamma$, as shown in Eq. 4.9.

$$\oint_{\partial\gamma} f(\boldsymbol{x})\,\nu(\boldsymbol{x})\,d\boldsymbol{x} = \int_{\Gamma-\gamma} f\left(\boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right)\right)\rho_0(\boldsymbol{x})\,d\boldsymbol{x}, \boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right) \in \partial\gamma \tag{4.9}$$

In other words, we have established a connection between the average on the reference surface $\langle f\rangle_{\nu} = \langle f\left(\boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right)\right)\rangle_0$, and its value in the volume.

With this connection, volume of the region $\Gamma - \gamma$ can be expressed by the surface of the reference region γ by Eq. 4.10, where $\tilde{\boldsymbol{x}} \equiv \boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right) \in \partial\gamma$.

$$\begin{aligned}
S_{\partial\gamma} &= \oint_{\partial\gamma} 1\,d\boldsymbol{x} = \oint_{\partial\gamma} \frac{\nu(\boldsymbol{x})}{\nu(\boldsymbol{x})}\,d\boldsymbol{x} \\
&\equiv \left\langle \frac{1}{\nu(\boldsymbol{x})} \right\rangle_{\nu} = \left\langle \frac{1}{\nu\left(\boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right)\right)} \right\rangle_0 \\
&\equiv \int_{\Gamma-\gamma} \frac{\rho_0(\boldsymbol{x})}{\nu\left(\boldsymbol{X}\left(\tau^{+}(\boldsymbol{x}),\boldsymbol{x}\right)\right)}\,d\boldsymbol{x} \\
&= \int_{\Gamma-\gamma} \frac{\rho_0(\boldsymbol{x})}{\hat{\boldsymbol{n}}(\tilde{\boldsymbol{x}})\cdot\nabla U(\tilde{\boldsymbol{x}})\int_0^{+\infty}\rho_0\left(\boldsymbol{X}(-t,\tilde{\boldsymbol{x}})\right)J(-t,\tilde{\boldsymbol{x}})\,dt}\,d\boldsymbol{x} \\
&\equiv V_{\Gamma-\gamma}\left\langle \frac{1}{\hat{\boldsymbol{n}}(\tilde{\boldsymbol{x}})\cdot\nabla U(\tilde{\boldsymbol{x}})\int_0^{+\infty}J(-t,\tilde{\boldsymbol{x}})\,dt} \right\rangle_0
\end{aligned} \tag{4.10}$$

Summarizing the above results, the volume of the basin can be expressed as Eq. 4.11, where

$\tilde{\boldsymbol{x}} \equiv \boldsymbol{X}\left(\tau^{+}\left(\boldsymbol{x}\right), \boldsymbol{x}\right) \in \partial\gamma$.

$$
\begin{aligned}
V_{\Gamma} &= V_{\gamma} + V_{\Gamma-\gamma} \\
&= V_{\gamma} + \frac{S_{\partial\gamma}}{\left\langle \left(\hat{\boldsymbol{n}}\left(\tilde{\boldsymbol{x}}\right) \cdot \boldsymbol{\nabla} U\left(\tilde{\boldsymbol{x}}\right) \int_{0}^{+\infty} J\left(-t, \tilde{\boldsymbol{x}}\right)\, dt \right)^{-1} \right\rangle_{0}}
\end{aligned}
\tag{4.11}
$$

Simulation are performed using the so-called "initiate-then-propagate" algorithm. Several starting points are drawn uniformly in space. Note that some of the starting points may not be used to compute basin volume V_{Γ} if they do not belong to basin Γ. Then for each starting point, we evolve the starting point forward in time under gradient descent dynamics and backward in time under gradient ascent dynamics until the simulation meets the termination condition, such as $|\boldsymbol{\nabla} U\left(\boldsymbol{x}\right)| \leqslant \epsilon$ with $\epsilon \ll 1$.

If we know the exact location of basin center, we can use that in our estimates; otherwise, the location of $\boldsymbol{c}$ is evaluated as the mean position of the final positions in the forward simulations. Note that $\boldsymbol{c}$ is generally not the center of the basin of attraction. The accuracy of the basin center $\boldsymbol{c}$ does not affect our results as long as we make sure that all the forward simulations of the starting points that belong to the basin Γ eventually fall inside the reference hyperball centered at the chosen $\boldsymbol{c}$. Eventually with these trajectories generated by valid starting points that belong to the basin Γ, the basin volume V_{Γ} is computed by Eq. 4.11.

4.3.2 Starting from a known surface

Some starting points may be wasted if they do not belong to the basin of interest Γ. One way to improve the efficiency of the sampling is to draw starting points on the surface $\partial\gamma$ of the reference hyperball γ with the knowledge of $\boldsymbol{c}$, either using the final position of one forward trajectory under gradient descent dynamics or knowing the local minimum apriori.

In order to establish a connection between the volume of interest V_{Γ} and the reference surface

$S_{\partial\gamma}$, we generalize the key equality in this chapter (Eq. 4.7) and obtain three equalities shown in Eq. 4.12, where $J(t,\boldsymbol{x}) \equiv \exp\left(\int_0^t -\boldsymbol{\nabla}\cdot\boldsymbol{\nabla}U(\boldsymbol{X}(s,\boldsymbol{x}))\,ds\right)$ is the corresponding Jacobian of gradient descent dynamics.

$$\begin{cases} V_{\Gamma-\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x})\cdot\boldsymbol{\nabla}U(\boldsymbol{x}) \int_{-\infty}^{0} J(t,\boldsymbol{x})\,dt\,d\boldsymbol{x} \\ V_{\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x})\cdot\boldsymbol{\nabla}U(\boldsymbol{x}) \int_{0}^{+\infty} J(t,\boldsymbol{x})\,dt\,d\boldsymbol{x} \\ V_{\Gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x})\cdot\boldsymbol{\nabla}U(\boldsymbol{x}) \int_{-\infty}^{+\infty} J(t,\boldsymbol{x})\,dt\,d\boldsymbol{x} \end{cases} \tag{4.12}$$

Note that the first equality is exactly the same as Eq. 4.7 after changing variable t to $-t$ and plugging in the expression of $\rho_0(\boldsymbol{x})$ defined by Eq. 4.5. This equality can be understood in another way; the latter factor $\int_{-\infty}^{0} J(t,\boldsymbol{x})\,dt\,d\boldsymbol{x}$ is the total volume contribution along the trajectory generated by the point $\boldsymbol{x}$, propagated from the starting volume $d\boldsymbol{x}$ and the former factor $\hat{\boldsymbol{n}}(\boldsymbol{x})\cdot\boldsymbol{\nabla}U(\boldsymbol{x})$ is the projective volume contribution that passes through the reference surface $\partial\gamma$.

The second equality can be derived following exactly the same procedure as for Eq. 4.7 with starting points uniformly drawn in the reference volume γ.

Summing the first and the second equalities gives the third equality.

When starting points are drawn from the reference surface $\partial\gamma$ under an arbitrary distribution μ, then three volumes $V_{\Gamma-\gamma}, V_{\Gamma}, V_{\gamma}$ can be estimated by Eq. 4.13, where $J(t,\boldsymbol{x})$ is the corresponding Jacobian of gradient descent dynamics.

$$
\begin{cases}
V_{\Gamma-\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad = \oint_{\partial\gamma} \frac{\mu\left(\boldsymbol{x}\right)}{\mu\left(\boldsymbol{x}\right)} \left(\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad \equiv \left\langle \frac{\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{x}\right)\, dt}{\mu\left(\boldsymbol{x}\right)} \right\rangle_{\mu} \\
V_{\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad = \oint_{\partial\gamma} \frac{\mu\left(\boldsymbol{x}\right)}{\mu\left(\boldsymbol{x}\right)} \left(\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right) \int_{0}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad \equiv \left\langle \frac{\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt}{\mu\left(\boldsymbol{x}\right)} \right\rangle_{\mu} \\
V_{\Gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad = \oint_{\partial\gamma} \frac{\mu\left(\boldsymbol{x}\right)}{\mu\left(\boldsymbol{x}\right)} \left(\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right)\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt\, d\boldsymbol{x} \\
\quad \equiv \left\langle \frac{\hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt}{\mu\left(\boldsymbol{x}\right)} \right\rangle_{\mu}
\end{cases}
\tag{4.13}
$$

A trivial case is that starting points are drawn uniformly from the reference surface $\partial\gamma$ with $\mu_0\left(\boldsymbol{x}\right) = \frac{1}{S_{\partial\gamma}}$, then three volumes can be estimated by Eq. 4.14, where $J\left(t, \boldsymbol{x}\right)$ is the corresponding Jacobian of gradient descent dynamics.

$$
\begin{cases}
V_{\Gamma-\gamma} = S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0} \\
V_{\gamma} = S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0} \\
V_{\Gamma} = S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0}
\end{cases}
\tag{4.14}
$$

Then there are three ways to estimate volume of the basin of interest V_{Γ}, as shown in Eq. 4.15, where V_{γ} is the analytical expression of the volume of the hyperball and $S_{\partial\gamma}$ is the analytical

expression of the surface of the hyperball.

$$\begin{cases} V_{\Gamma}^{(1)} = V_{\gamma} + S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0} \\ V_{\Gamma}^{(2)} = S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0} \\ V_{\Gamma}^{(3)} = V_{\gamma} \dfrac{\left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0}}{\left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{x}\right) \int_{0}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0}} \end{cases} \tag{4.15}$$

Another special case is that starting points $\boldsymbol{y}$ are drawn by finding the intersection points between trajectories generated by points $\boldsymbol{x}$ drawn uniformly in the space and the reference surface. The surface density of such starting points $\boldsymbol{y} \equiv \boldsymbol{X}\left(\tau\left(\boldsymbol{x}\right), \boldsymbol{x}\right)$ have been derived in Eq. 4.8. Then the first estimator of volume $V_{\Gamma-\gamma}$ in Eq. 4.13 is reduced to Eq. 4.16.

$$\begin{aligned} V_{\Gamma-\gamma} &= \left\langle \frac{\hat{\boldsymbol{n}}\left(\boldsymbol{y}\right) \cdot \boldsymbol{\nabla} U\left(\boldsymbol{y}\right) \int_{-\infty}^{0} J\left(t, \boldsymbol{y}\right)\, dt}{\nu\left(\boldsymbol{y}\right)} \right\rangle_{\nu} \\ &= V_{\Gamma} \left\langle \frac{\int_{-\infty}^{0} J\left(t, \boldsymbol{y}\right)\, dt}{\int_{-\infty}^{+\infty} J\left(t, \boldsymbol{y}\right)\, dt} \right\rangle_{\nu} \\ &= V_{\Gamma} \left\langle \frac{\int_{-\infty}^{\tau(\boldsymbol{x})} J\left(t, \boldsymbol{x}\right)\, dt}{\int_{-\infty}^{+\infty} J\left(t, \boldsymbol{x}\right)\, dt} \right\rangle_{0} \end{aligned} \tag{4.16}$$

Note that from the second line in Eq. 4.16 to the third line in Eq. 4.16, the procedure of drawing starting points $\boldsymbol{y} \equiv \boldsymbol{X}\left(\tau\left(\boldsymbol{x}\right), \boldsymbol{x}\right)$ with surface density ν is exactly the same as that of drawing starting points $\boldsymbol{x}$ uniformly in the basin of attract V_{Γ} with probability density $\rho_0\left(\boldsymbol{x}\right) = \frac{1}{V_{\Gamma}}$. Besides, the result of the ratio in the second line is the same as that in the third line since the common factor from changing variables cancels out. The significance of Eq. 4.16 is that we show that the unbiased ITNEIS estimator of basin volume used in the previous chapters can be considered as a special case of estimators using density propagation method at points on the reference surface.

4.4 Numerical results

There are two main sources of error when using the density propagation method to estimate volume of the basin of attraction. One is the sampling error, which can be reduced by drawing more starting points. Another is the numerical error comes from the numerical integration part of solving equation of motion, which can be reduced by choosing smaller numerical integration step Δt. To control the error in numerical integration part, the integration time step Δt at position $\boldsymbol{x}$ can be set adaptively by:

$$\Delta t = \frac{\Theta}{|\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} U(\boldsymbol{x})|}, \tag{4.17}$$

where Θ is a small constant, serving as the tolerance in the computation of the corresponding Jacobian $J(t, \boldsymbol{x}) \equiv \int_0^t -\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} U(\boldsymbol{x})\, ds$. In that case, performing this integral through a discrete summation over time integrals gives contributions of approximately equal magnitude Θ for every generated point.

In practice, it is not possible to get a sample point exactly on the surface $\partial\gamma$ of the reference hyperball. A most likely scenario is that we have $|\boldsymbol{X}(t_i, \boldsymbol{x}) - \boldsymbol{c}| > r, |\boldsymbol{X}(t_{i+1}, \boldsymbol{x}) - \boldsymbol{c}| < r$, where r is the radius of the hyperball. In this case, we propose a simple linear interpolation method (Eq. 4.18) to find the interpolated sample point $\boldsymbol{x}_{\partial\gamma}$ on the reference surface $\partial\gamma$ between two sample points $\boldsymbol{x}_0 \equiv \boldsymbol{X}(t_0, \boldsymbol{x}), \boldsymbol{x}_1 \equiv \boldsymbol{X}(t_1, \boldsymbol{x})$ at time t_0, t_1 and the corresponding time $t_{\partial\gamma}$ by linear algebra [102], where $r_0 = |\boldsymbol{x}_0 - \boldsymbol{c}|, r_1 = |\boldsymbol{x}_1 - \boldsymbol{c}|, d = |\boldsymbol{x}_1 - \boldsymbol{x}_0|$.

$$\begin{cases} s = \dfrac{r_0^2 - r_1^2 + d^2 \pm \sqrt{\left(r_0^2 - r_1^2 + d^2\right)^2 - 4d^2\left(r_0^2 - r^2\right)}}{2d^2} \\ \boldsymbol{x}_{\partial\gamma} = \boldsymbol{x}_0 + s(\boldsymbol{x}_1 - \boldsymbol{x}_0) \\ t_{\partial\gamma} = t_0 + s(t_1 - t_0) \end{cases} \tag{4.18}$$

Note that there are two possible choices of s in Eq. 4.18, and s such that $\left(t_{\partial\gamma} - t_0\right) \cdot \left(t_{\partial\gamma} - t_1\right) \leqslant 0$

is used when applied for our numerical results because the linear interpolated time $t_{\partial\gamma}$ should be within the range of t_0 and t_1. The following potentials are implemented by CVODE [103–106], an open-source software library that uses variable-order, variable-step multistep methods to solve stiff and nonstiff ordinary differential equation (ODE) systems.

4.4.1 Gaussian potential

The method is tested for a d-dimensional Gaussian potential, defined in Eq. 4.19, where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the minimum of the basin of attraction.

$$U(\boldsymbol{x}) = -\exp\left(-\frac{|\boldsymbol{x}-\boldsymbol{\mu}|^2}{2}\right), \boldsymbol{x} \in \mathbb{R}^d \tag{4.19}$$

Here we choose $\boldsymbol{\mu} = \mathbf{0}$ for simplicity. A boundary $|\boldsymbol{x}-\boldsymbol{\mu}| = 4$ is set manually when computing the basin volume since there is no upper bound of the potential energy and the basin volume is infinity without the boundary. The ideal basin volume is the analytical expression of the hyperball with radius 4 in d dimensions, and this value is used to calculate the absolute relative error $\eta = \left|\frac{\hat{V}-V}{V}\right|$, where $\hat{V}$ is the estimated basin volume and V is the ideal value of the basin volume.

Fig. 4.1 shows the performance (measured by absolute relative error η) of estimators in Eq. 4.11 and Eq. 4.15 with respect to dimension D with choices of various radius r of the reference hyperball and tolerance Θ (Eq. 4.17). Estimations are made from trajectories generated by 2000 starting points.

The performance of these estimators are controlled by tolerance Θ and error has the same magnitude as that of Θ. Hence estimators are expected to have better performance with smaller tolerance Θ at the cost of more expensive computation cost. The performance of these estimators are also influenced by the radius r of the reference hyperball slightly.

We observe that the performance of the estimator in Eq. 4.11 is slightly better with larger radius r while the performance of three estimators in Eq. 4.15 is better with smaller radius r.

Results of three estimators in Eq. 4.15 are hardly different and we will use the first estimator $V_{\Gamma}^{(1)}$ next to estimate the volume of the basin of attraction with starting points drawn from the reference surface. The performance of estimators in Eq. 4.15 is slightly better than that of estimator in Eq. 4.11.

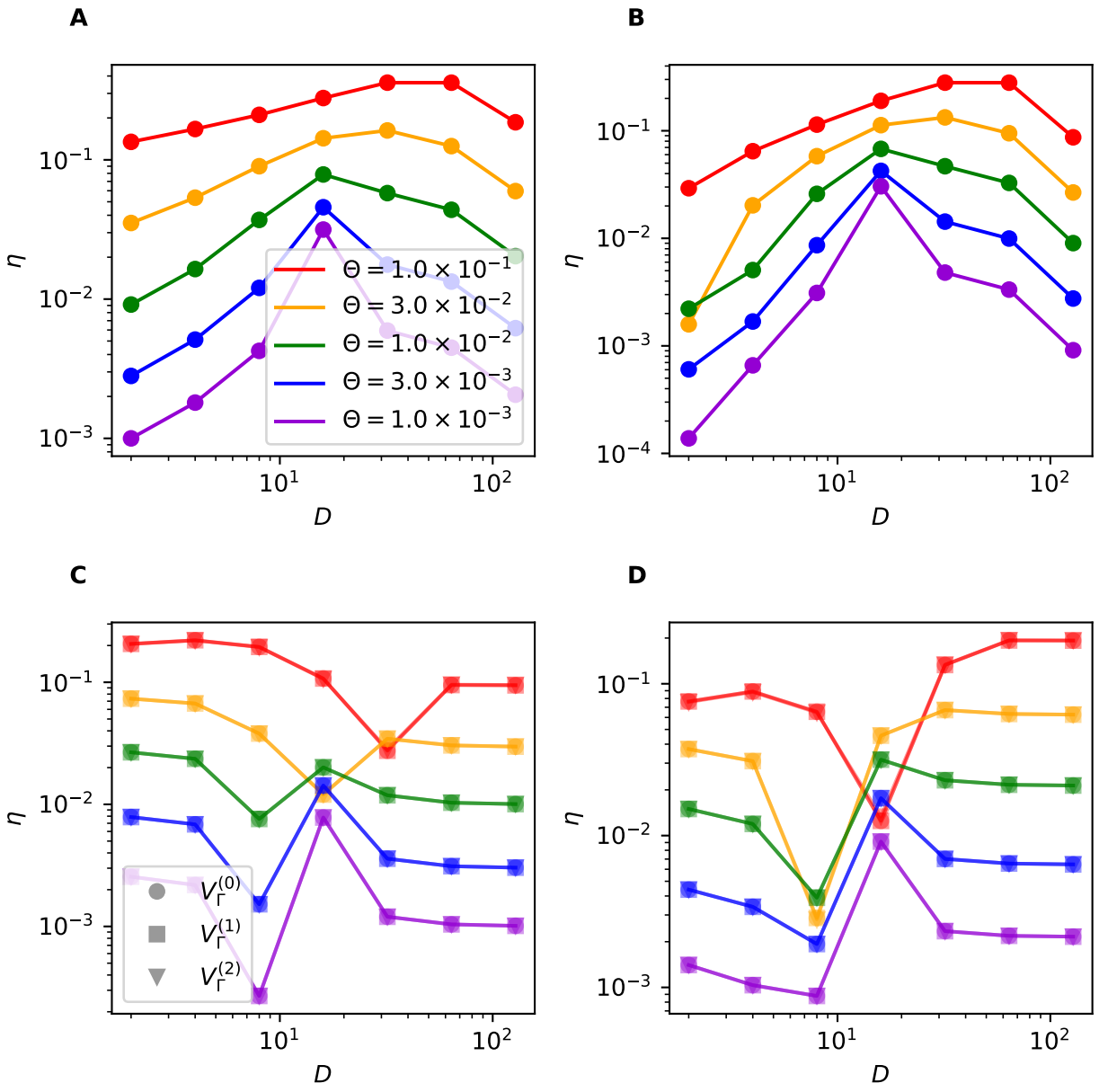

Figure 4.1: Absolute relative error (η) of estimators for varying dimension (D) computed using a range of Θ as described in the text. Estimators in Eq. 4.11 (**A**, **B**) and Eq. 4.15 (**C**, **D**) are tested with five Θs, from $\Theta = 0.1$ (red) to $\Theta = 0.001$ (violet), in the rainbow color order. Three estimators in Eq. 4.15 (**C**, **D**) are represented as circle, square and triangle. The radius of the reference hyperball is set as 0.01 (**A**, **C**) and 0.1 (**B**, **D**) respectively. Each estimator uses trajectories generated from 2000 starting points.

4.4.2 "PowSumCos" potential

The method is tested in a periodic potential with period 1, defined in Eq. 4.20.

$$U(x_1, \ldots, x_d) = \left(1 + d + \sum_{i=1}^{d} \cos(2\pi x_i)\right)^{\frac{1}{2}} \tag{4.20}$$

Therefore, the basin of attraction is a hypercube with length 1 in d dimensions, and the corresponding ideal basin volume is $1^d = 1$. Fig. 4.2 **A** shows performance (measured by absolute relative error η) of the first estimator in Eq. 4.15 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly on the reference surface. Fig. 4.2 **B** shows the performance of the estimator in Eq. 4.11 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly in the space $\Gamma - \gamma$.

Both estimators fail when $D \geqslant 16$ but they fail in two different manners. The estimator using starting points on the reference surface (Eq. 4.15) gives an estimate of 0 while the estimator using starting points in the basin (Eq. 4.11) gives a very large value. Moreover, the estimator in Eq. 4.11 works slightly better while the performance of the estimators in Eq. 4.15 is more stable.

Fig. 4.2 **C** shows the performance of the first estimator in Eq. 4.15 with respect to dimension D and number of starting points on the reference surface N_T with radius $r = 0.01$ and tolerance $\Theta = 0.01$. Using more starting points does not solve the issue of poor estimates. Fig. 4.2 **D** shows some selective sample trajectories on top of the plot of potential energy of the "PowSumCos" potential in 2D. Almost all trajectories terminate at the corner of the square ("spikes" of the hypercube when the dimension is large), although starting points are drawn uniformly on the reference surface. Furthermore, some rare trajectories that terminate at points other than corners of the square contribute several magnitude greater than those common trajectories, which could be the reason why these estimators fail in high dimensions when using starting points from the

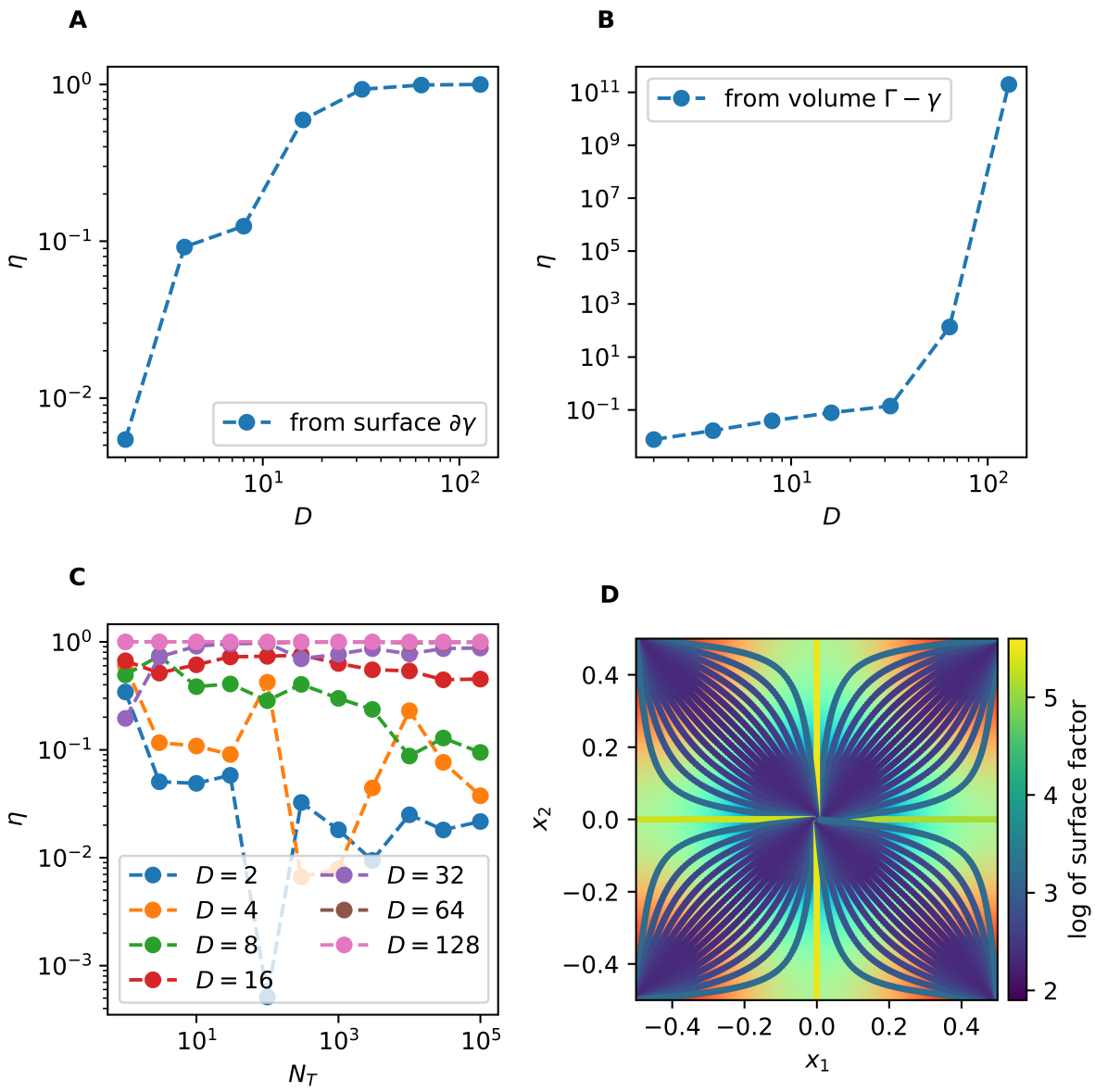

Figure 4.2: **A** Absolute relative error η of the first estimator in Eq. 4.15 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly on the reference surface. **B** Absolute relative error η of the estimator in Eq. 4.11 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly in the space $\Gamma - \gamma$. **C** Absolute relative error η of the first estimator in Eq. 4.15 with respect to dimension D and number of starting points N_T drawn uniformly on the reference surface with tolerance $\Theta = 0.01$ and radius of the reference hyperball $r = 0.01$. **D** Selective trajectories generated by starting points on the reference surface with radius $r = 0.01$ under gradient ascent dynamics on top of potential energy plot of the "PowSumCos" potential in 2D. Surface factor is the term in the bracket of Eq. 4.11.

reference surface. On the other hand, since little space is mapped to almost all the reference surface, most starting points from the basin of attraction belong to the trajectory with dominant contribution. These two facts explain why estimators using starting points from the reference surface underestimate the basin volume while estimators using starting points from the basin of attraction overestimate the basin volume.

Note that in equality (Eq. 4.7), it is true for any starting probability density in space $\rho_0(\boldsymbol{x})$ and any dynamics $\boldsymbol{b}(\boldsymbol{x})$. Hence a generalized version of the key equality is obtained, as shown in Eq. 4.21. The physical interpretation of Eq. 4.21 is that the overall projective "mass" contribution of the trajectories generated from starting points on the reference surface should be the "mass" of the space $V_{\Gamma-\gamma}$, which is 1. However, there is no guarantee that trajectories are terminated at the boundary of the basin of attraction with an arbitrary dynamics $\boldsymbol{b}(\boldsymbol{x})$. A termination condition relying on the original potential $U(\boldsymbol{x})$, $\exists i, \nabla_i U(\boldsymbol{X}(t,\boldsymbol{x})) \cdot \nabla_i U(\boldsymbol{X}(t+\Delta t,\boldsymbol{x})) \leqslant 0$ is used to terminate trajectories properly.

$$1 = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot \boldsymbol{b}(\boldsymbol{x}) \int_0^{\tau(\boldsymbol{x})} \rho_0(\boldsymbol{X}(-t,\boldsymbol{x})) J(-t,\boldsymbol{x})\, dt\, d\boldsymbol{x}, \tag{4.21}$$

where $J(t,\boldsymbol{x}) \equiv \exp\left(\int_0^t \boldsymbol{\nabla}\cdot\boldsymbol{b}(\boldsymbol{X}(s,\boldsymbol{x}))\, ds\right)$ is the corresponding Jacobian of the dynamics $\boldsymbol{b}(\boldsymbol{x})$ and $\tau(\boldsymbol{x})$ is the stopping time that satisfies $\boldsymbol{X}(-\tau(\boldsymbol{x}),\boldsymbol{x}) \in \partial\Gamma$. Then three estimators of volume in Eq. 4.12 becomes

$$\begin{cases} V_{\Gamma-\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot \boldsymbol{b}(\boldsymbol{x}) \int_{\tau^-(\boldsymbol{x})}^{0} J(t,\boldsymbol{x})\, dt\, d\boldsymbol{x} \\ V_{\gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot \boldsymbol{b}(\boldsymbol{x}) \int_{0}^{\tau^+(\boldsymbol{x})} J(t,\boldsymbol{x})\, dt\, d\boldsymbol{x}\,, \\ V_{\Gamma} = \oint_{\partial\gamma} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot \boldsymbol{b}(\boldsymbol{x}) \int_{\tau^-(\boldsymbol{x})}^{\tau^+(\boldsymbol{x})} J(t,\boldsymbol{x})\, dt\, d\boldsymbol{x} \end{cases} \tag{4.22}$$

and the volume of the basin of attraction V_Γ is estimated by Eq. 4.23 if starting points are drawn

uniformly on the reference surface,

$$V_\Gamma = V_\gamma + S_{\partial\gamma} \left\langle \hat{\boldsymbol{n}}\left(\boldsymbol{x}\right) \cdot \boldsymbol{b}\left(\boldsymbol{x}\right) \int_{\tau^-(\boldsymbol{x})}^{0} J\left(t, \boldsymbol{x}\right)\, dt \right\rangle_{\mu_0}, \tag{4.23}$$

where $J\left(t, \boldsymbol{x}\right)$ is the corresponding Jacobian of the dynamics $\boldsymbol{b}\left(\boldsymbol{x}\right)$, V_γ is the analytical expression of the volume of the reference hyperball γ, $S_{\partial\gamma}$ is the analytical expression of the surface of the reference hyperball γ and starting points $\boldsymbol{x}$ are drawn uniformly on the reference surface $\partial\gamma$.

Another estimator of the volume of the basin of attraction V_Γ is shown in Eq. 4.24 if starting points are drawn uniformly in the space $\Gamma - \gamma$.

$$V_\Gamma = V_\gamma + \frac{S_{\partial\gamma}}{\left\langle \left(\hat{\boldsymbol{n}}\left(\tilde{\boldsymbol{x}}\right) \cdot \boldsymbol{b}\left(\tilde{\boldsymbol{x}}\right) \int_0^{\tau(\tilde{\boldsymbol{x}})} J\left(-t, \tilde{\boldsymbol{x}}\right)\, dt \right)^{-1} \right\rangle_0}, \tag{4.24}$$

where $\tilde{\boldsymbol{x}} \equiv \boldsymbol{X}\left(\tau\left(\boldsymbol{x}\right), \boldsymbol{x}\right) \in \partial\gamma$ is the intersection point between the trajectory generated by the starting point $\boldsymbol{x}$ and the reference surface $\partial\gamma$, $\tau\left(\tilde{\boldsymbol{x}}\right)$ is the stopping time that satisfies $\boldsymbol{X}\left(\tau\left(\tilde{\boldsymbol{x}}\right), \tilde{\boldsymbol{x}}\right) \in \partial\Gamma$, $J\left(t, \boldsymbol{x}\right)$ is the corresponding Jacobian of the dynamics $\boldsymbol{b}\left(\boldsymbol{x}\right)$, V_γ is the analytical expression of the volume of the reference hyperball γ, $S_{\partial\gamma}$ is the analytical expression of the surface of the reference hyperball γ and starting points $\boldsymbol{x}$ are drawn uniformly in the space $\Gamma - \gamma$.

In this case, an umbrella-sampling (US)-like flow, as defined in Eq. 4.25, is used to exert some perturbation so that trajectories are more likely to terminate at points other than spikes of the hypercube.

$$\boldsymbol{b}^{\mathrm{US}}\left(\boldsymbol{x}\right) \equiv -\boldsymbol{\nabla} U\left(\boldsymbol{x}\right) - k\left(\boldsymbol{x} - \boldsymbol{c}\right), \tag{4.25}$$

where $\boldsymbol{c}$ is the center of the basin of attraction and $k \in \mathbb{R}$ is a biasing parameter. This flow can be considered as a flow of umbrella sampling in the window centered at $\boldsymbol{c}$. When $k = 0.0$, the US-like flow reduces to that of gradient descent dynamics used previously. When k is large, the US-like flow approaches to that of harmonic potential where density is propagated isotropically.

The corresponding Jacobian of the US-like flow is

$$J^{\mathrm{US}}(t, \boldsymbol{x}) = J(t, \boldsymbol{x}) - kd, \tag{4.26}$$

where d is the dimension of system and $J(t, \boldsymbol{x})$ is the corresponding Jacobian of the gradient descent dynamics as defined in Eq. 4.3. Fig. 4.3 **A** shows the performance (measured by absolute relative error η) of the estimator in Eq. 4.23 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly on the reference surface under an US-like flow with various choices of k. Fig. 4.3 **B** shows the performance of the estimator in Eq. 4.24 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly in the space $\Gamma - \gamma$ under an US-like flow with various choices of k. The performance of both estimators under an US-like flow with $k = 1.0$ is significantly improved. Fig. 4.3 **C** shows the performance of the estimator in Eq. 4.23 with respect to dimension D and number of starting points on the reference surface N_T with radius $r = 0.01$ and tolerance $\Theta = 0.01$ under an US-like flow with $k = 1.0$. With more trajectories and thus a larger computational cost, the performance of the estimator under an US-flow with a proper k can be further improved. Fig. 4.3 **D** shows some sample trajectories on top of the plot of potential energy of the "PowSumCos" potential in 2D. The difference between contribution of trajectories under the US-like flow with $k = 1.0$ is much smaller than that of the original flow ($k = 0.0$). This suggests that an optimal flow could potentially be the one with smallest variance of contribution of trajectories, which is still under investigation.

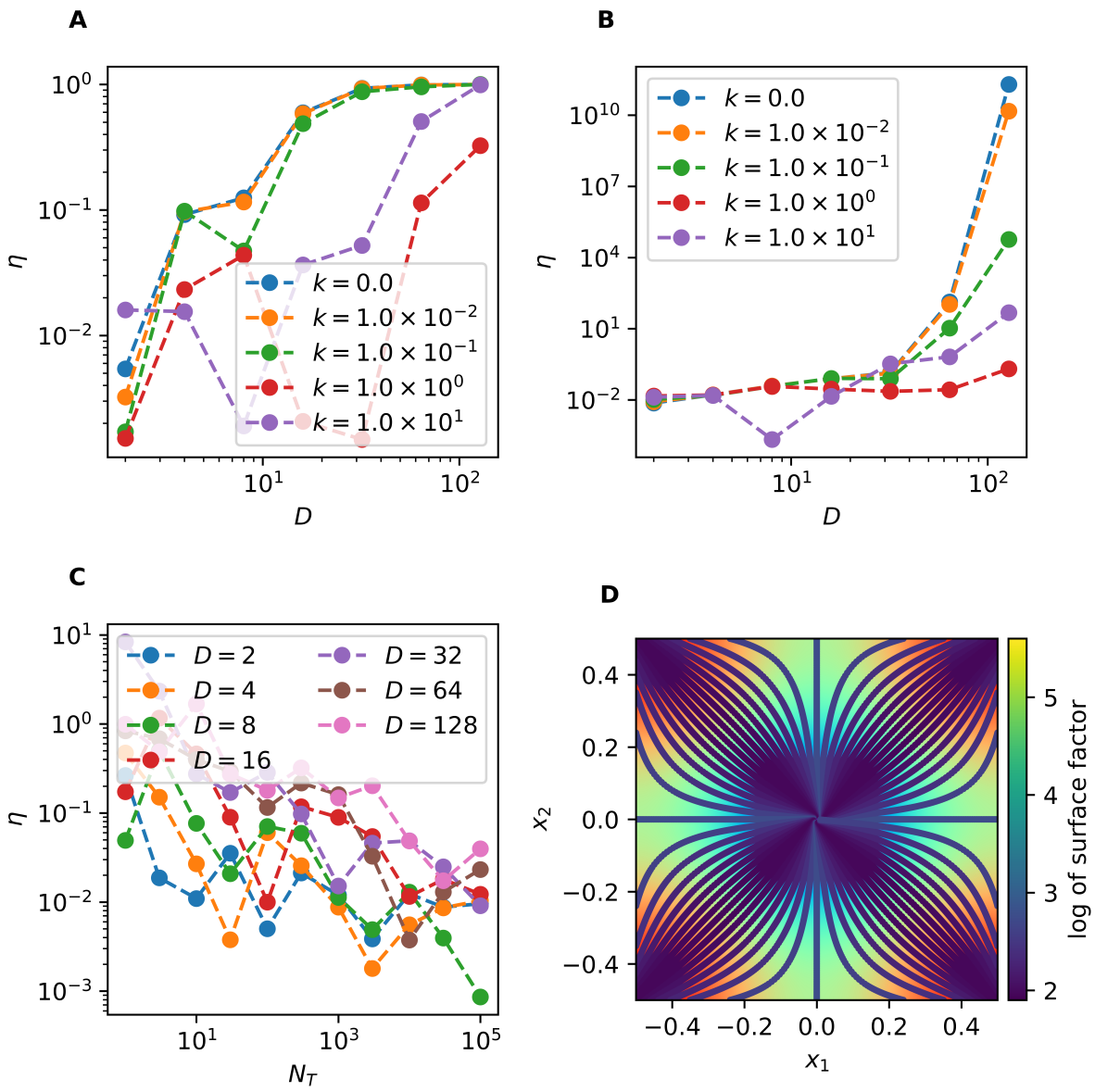

Figure 4.3: **A** Absolute relative error η of the estimator in Eq. 4.23 with respect to dimension D with tolerance $\Theta = 0.01$, radius of the reference hyperball $r = 0.01$ and 2000 starting points drawn uniformly on the reference surface under an US-like flow with various ks. **B** Absolute relative error η of the estimator in Eq. 4.24 with respect to dimension D with tolerance $\Theta = 0.01$, radius $r = 0.01$ and 2000 starting points drawn uniformly in the space $\Gamma - \gamma$ under an US-like flow with various ks. **C** Absolute relative error η of the estimator in Eq. 4.23 with respect to dimension D and number of starting points N_T drawn uniformly on the reference surface with tolerance $\Theta = 0.01$ and radius $r = 0.01$ under an US-like flow with $k = 1.0$. **D** Sample trajectories generated by starting points on the reference surface with radius $r = 0.01$ under an US-like flow with $k = 1.0$ on top of potential energy plot of the "PowSumCos" potential in 2D. The surface factor is the term in the bracket of Eq. 4.23.

4.5 Conclusions

A density propagation method is used to make use of the whole trajectory for estimating basin volumes. Following the idea in the previous chapter, a key equality (Eq. 4.7) is obtained, which can be understood as the "mass" contribution of the trajectory. With this critical equality (Eq. 4.7), there are two ways to estimate the basin volume. (**1**) From trajectories generated by starting points in space V_Γ. (**2**) From trajectories generated by starting points on the reference surface $S_{\partial\gamma}$. Moreover, the unbiased ITNEIS estimators in the previous chapter can be considered as a special case of the second way with starting points drawn under the surface density ν, defined in Eq. 4.8. The method works well in gaussian potentials and the performance of both methods are controlled by width Θ in the integral of Jacobian. However, both methods fail in "PowSumCos" potential, where the basin of attraction is a hypercube with 2^d spikes in d dimensions. A generalized version of Eq. 4.7 has been proposed in Eq. 4.21, and the performance of both methods is considerably improved by umbrella sampling (US)-like flow with proper spring constant k. However, the motivation for selecting k to achieve the optimal performance is still under investigation.

5 | Conclusion

In this thesis, we take advantages of novel mathematical results of density propagation and have made three main contributions.

Firstly, we applied the original unbiased nonequilibrium importance sampling (NEIS) estimators from the work of Rotskoff and Vanden-Eijnden under "quench" dynamics to two systems: independent harmonic springs and alanine dipeptide. In the system of independent harmonic springs, a nearly exponential decay of mean total energy has been observed when γ in "quench" dynamics is adequately small. A proper heuristic for how long simulation of "quench" dynamics should be performed can be made with this observation. "Quench" dynamics itself works well in system with free energy surface (FES) that has single basin while it solely is not the optimal choice for system with FES that has multiple basins, like the case in alanine dipeptide in vacuum. Combining "quench" dynamics with umbrella sampling can significantly improve the performance of both sides after reweighting biased results aptly. Regarding "quench" dynamics, the combined method gives decent results for computing FES with acceptable errors. Regarding umbrella sampling, the combined method has three advantages. (**1**) It is more efficient since high energy regions are more likely to be sampled in backward "quench" dynamics and there are more overlap between umbrella windows. (**2**) It can get better estimations when chosen collective variable (CV) is bad. (**3**) It can make good estimations at various temperatures using the same trajectories while a normal simulation can only make proper estimation at the same temperature when sample points are drawn. Furthermore, "quench" dynamics can also be coupled with other

sophisticated enhanced sampling techniques readily such as solute tempering, as we have shown in the system of alanine dipeptide in water.

Secondly, we extended the theory of the unbiased NEIS estimators with given trajectories, standing on the shoulder of Cao and Vanden-Eijnden. After generalizing the unbiased NEIS estimators, we not only gave mathematical and physical interpretation of the generalized unbiased NEIS estimators but also figured out that conventional way of computing the corresponding variance is not always true and proposed the correct version of the variance of the unbiased generalized NEIS estimators with the acknowledgment that the source of unbiased property of NEIS estimators at the sample point $\boldsymbol{x}$ comes from the unbiased contribution of the infinitely long trajectory generated from any starting point $\boldsymbol{x}$. It can be further proved by Jensen's inequality that ITNEIS estimator is the optimal unbiased NEIS estimator with given trajectory, which is intuitive. Therefore, the only way to improve the performance of NEIS estimators is choosing better dynamics or flow, either by solving optimal flow condition or by training flow using deep learning techniques. In practice, simulations can be terminated when both the numerator and denominator of ITNEIS estimator converge, although the truncated NEIS estimator is not strictly unbiased.

Thirdly, we applied so-called density propagation method and took full advantage of the trajectory that determines whether a point in the space belongs to the basin of attraction or not, which was wasted in previous methods. Several estimators of the volume of the basin of attraction have been proposed, using trajectories generated by starting points drawn either from space or from a reference surface. We also recognized that the unbiased ITNEIS estimator used in the previous chapters can be considered as a special case of density propagation method with a specially chosen surface density. The method works in the simple case of gaussian potentials but fails in the more complex case of "PowSumCos" potential, where the basin of attraction has multiple spikes in high dimensions. A generalized equality that captures "mass" contribution of all trajectories from the reference surface was obtained, which provides space of further improvements of

performance of the estimators with better dynamics or flow. An umbrella sampling-like dynamics with proper spring constant can considerably improve the performance of the estimators, but optimizing this approach and determining whether it is more efficient and accurate than other approaches requires further investigation.

Bibliography

[1] Berni Julian Alder, Thomas Everett Wainwright, et al. "Phase transition for a hard sphere system". In: *The Journal of chemical physics* 27.5 (1957), p. 1208.

[2] Shneior Lifson and Arieh Warshel. "Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules". In: *The Journal of Chemical Physics* 49.11 (1968), pp. 5116–5129.

[3] Michael Levitt and Shneior Lifson. "Refinement of protein conformations using a macromolecular energy minimization procedure". In: *Journal of molecular biology* 46.2 (1969), pp. 269–279.

[4] J Andrew McCammon, Bruce R Gelin, and Martin Karplus. "Dynamics of folded proteins". In: *nature* 267.5612 (1977), pp. 585–590.

[5] Romelia Salomon-Ferrer et al. "Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald". In: *Journal of chemical theory and computation* 9.9 (2013), pp. 3878–3888.

[6] John E Stone et al. "Evaluation of emerging energy-efficient heterogeneous computing platforms for biomolecular and cellular simulation workloads". In: *2016 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE. 2016, pp. 89–100.

[7] Adam Hospital et al. “Molecular dynamics simulations: advances and applications”. In: *Advances and Applications in Bioinformatics and Chemistry* (2015), pp. 37–47.

[8] Scott A Hollingsworth and Ron O Dror. “Molecular dynamics simulation for all”. In: *Neuron* 99.6 (2018), pp. 1129–1143.

[9] Gaurav Mitra et al. “A coarse-grained simulation model for colloidal self-assembly via explicit mobile binders”. In: *Soft Matter* 19.23 (2023), pp. 4223–4236.

[10] Shihao Zang et al. “Direct observation and control of non-classical crystallization pathways in binary colloidal systems”. In: *Nature communications* 16.1 (2025), p. 3645.

[11] Yuvraj Singh, Glen M Hocky, and Brad J Nolen. “Molecular dynamics simulations support a multistep pathway for activation of branched actin filament nucleation by Arp2/3 complex”. In: *Journal of Biological Chemistry* 299.9 (2023).

[12] Willmor J Peña Ccoa and Glen M Hocky. “Assessing models of force-dependent unbinding rates via infrequent metadynamics”. In: *The Journal of Chemical Physics* 156.12 (2022).

[13] Willmor J Peña Ccoa et al. “A direct computational assessment of vinculin–actin unbinding kinetics reveals catch-bonding behavior”. In: *Proceedings of the National Academy of Sciences* 122.21 (2025), e2425982122.

[14] Desiree Mae Prado et al. “Breakthrough Conductivity Enhancement in Deep Eutectic Solvents via Grotthuss-Type Proton Transport”. In: *Advanced Materials Interfaces* 11.36 (2024), p. 2400508.

[15] Daisuke Kuroshima et al. “Machine learning classification of local environments in molecular crystals”. In: *Journal of chemical theory and computation* 20.14 (2024), pp. 6197–6206.

[16] Song Xia, Eric Chen, and Yingkai Zhang. “Integrated molecular modeling and machine learning for drug design”. In: *Journal of chemical theory and computation* 19.21 (2023), pp. 7478–7495.

[17] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2010.

[18] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[19] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.

[20] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.

[21] Lutz Molgedey and Heinz Georg Schuster. "Separation of a mixture of independent signals using time delayed correlations". In: *Physical review letters* 72.23 (1994), p. 3634.

[22] Laurens Van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE." In: *Journal of machine learning research* 9.11 (2008).

[23] Leland McInnes, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction". In: *arXiv preprint arXiv:1802.03426* (2018).

[24] Ronald R Coifman et al. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps". In: *Proceedings of the national academy of sciences* 102.21 (2005), pp. 7426–7431.

[25] Joshua B Tenenbaum, Vin de Silva, and John C Langford. "A global geometric framework for nonlinear dimensionality reduction". In: *science* 290.5500 (2000), pp. 2319–2323.

[26] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. "Nonlinear component analysis as a kernel eigenvalue problem". In: *Neural computation* 10.5 (1998), pp. 1299–1319.

[27] Nicodemo Mazzaferro et al. "Good rates from bad coordinates: the exponential average time-dependent rate approach". In: *Journal of chemical theory and computation* 20.14 (2024), pp. 5901–5912.

[28] Subarna Sasmal, Martin McCullagh, and Glen M Hocky. “Reaction coordinates for conformational transitions using linear discriminant analysis on positions”. In: *Journal of chemical theory and computation* 19.14 (2023), pp. 4427–4435.

[29] Subarna Sasmal et al. “Quantifying Unbiased Conformational Ensembles from Biased Simulations Using ShapeGMM”. In: *Journal of chemical theory and computation* 20.9 (2024), pp. 3492–3502.

[30] Herman Kahn and Andy W Marshall. “Methods of reducing sample size in Monte Carlo computations”. In: *Journal of the Operations Research Society of America* 1.5 (1953), pp. 263–278.

[31] Augustine Kong. “A note on importance sampling using standardized weights”. In: *University of Chicago, Dept. of Statistics, Tech. Rep* 348 (1992), p. 14.

[32] Jun S Liu. “Metropolized independent sampling with comparisons to rejection sampling and importance sampling”. In: *Statistics and computing* 6 (1996), pp. 113–119.

[33] Glenn M Torrie and John P Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *J. Comp. Phys.* 23.2 (1977), pp. 187–199.

[34] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the national academy of sciences* 99.20 (2002), pp. 12562–12566.

[35] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-tempered metadynamics: a smoothly converging and tunable free-energy method”. In: *Phys. Rev. Lett.* 100.2 (2008), p. 020603.

[36] Michele Invernizzi and Michele Parrinello. “Rethinking metadynamics: from bias potentials to probability distributions”. In: *The journal of physical chemistry letters* 11.7 (2020), pp. 2731–2736.

[37] Lula Rosso et al. “On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles”. In: *The Journal of chemical physics* 116.11 (2002), pp. 4389–4402.

[38] Jerry B Abrams and Mark E Tuckerman. “Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations”. In: *The Journal of Physical Chemistry B* 112.49 (2008), pp. 15742–15757.

[39] Luca Maragliano and Eric Vanden-Eijnden. “A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations”. In: *Chemical physics letters* 426.1-3 (2006), pp. 168–175.

[40] Michael J Hartmann et al. “Infinite switch simulated tempering in force (FISST)”. In: *The Journal of Chemical Physics* 152.24 (2020).

[41] Yuji Sugita and Yuko Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chem. Phys. Lett.* 314.1-2 (1999), pp. 141–151.

[42] Pu Liu et al. “Replica exchange with solute tempering: A method for sampling biological systems in explicit water”. In: *Proc. Natl. Acad. Sci.* 102.39 (2005), pp. 13749–13754.

[43] Lingle Wang, Richard A Friesner, and BJ Berne. “Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2)”. In: *The Journal of Physical Chemistry B* 115.30 (2011), pp. 9431–9438.

[44] Shankar Kumar et al. “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *J. Comp. Chem.* 13.8 (1992), pp. 1011–1021.

[45] Michael R Shirts and John D Chodera. “Statistically optimal analysis of samples from multiple equilibrium states”. In: *The Journal of chemical physics* 129.12 (2008).

[46] Erik H Thiede et al. “Eigenvector method for umbrella sampling enables error analysis”. In: *J. Chem. Phys.* 145.8 (2016), p. 084115.

[47] Johannes Kästner and Walter Thiel. “Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method:“Umbrella integration””. In: *The Journal of chemical physics* 123.14 (2005).

[48] Johannes Kästner. “Umbrella integration in two or more reaction coordinates”. In: *The Journal of chemical physics* 131.3 (2009).

[49] Marc Souaille and Benoıt Roux. “Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations”. In: *Computer physics communications* 135.1 (2001), pp. 40–57.

[50] Emilio Gallicchio et al. “Temperature weighted histogram analysis method, replica exchange, and transition paths”. In: *J. Phys. Chem. B* 109.14 (2005), pp. 6722–6731.

[51] Gene H Golub and Carl D Meyer Jr. “Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains”. In: *SIAM Journal on Algebraic Discrete Methods* 7.2 (1986), pp. 273–281.

[52] Yehuda Vardi. “Empirical distributions in selection bias models”. In: *The Annals of Statistics* (1985), pp. 178–203.

[53] Christopher Jarzynski. “Nonequilibrium equality for free energy differences”. In: *Physical Review Letters* 78.14 (1997), p. 2690.

[54] Gerhard Hummer and Attila Szabo. “Free energy reconstruction from nonequilibrium single-molecule pulling experiments”. In: *Proceedings of the National Academy of Sciences* 98.7 (2001), pp. 3658–3661.

[55] Sean X Sun. “Equilibrium free energies from path sampling of nonequilibrium trajectories”. In: *The Journal of chemical physics* 118.13 (2003), pp. 5769–5775.

[56] Manuel Athenes. “A path-sampling scheme for computing thermodynamic properties of a many-body system in a generalized ensemble”. In: *The European Physical Journal B-Condensed Matter and Complex Systems* 38 (2004), pp. 651–663.

[57] Jerome P Nilmeier et al. “Nonequilibrium candidate Monte Carlo is an efficient tool for equilibrium simulation”. In: *Proceedings of the National Academy of Sciences* 108.45 (2011), E1009–E1018.

[58] Christoph Dellago and Gerhard Hummer. “Computing equilibrium free energies using non-equilibrium molecular dynamics”. In: *Entropy* 16.1 (2013), pp. 41–61.

[59] Radford M Neal. “Annealed importance sampling”. In: *Statistics and computing* 11.2 (2001), pp. 125–139.

[60] Grant M Rotskoff and Eric Vanden-Eijnden. “Dynamical Computation of the Density of States and Bayes Factors Using Nonequilibrium Importance Sampling”. In: *Phys. Rev. Lett.* 122.15 (2019), p. 150602.

[61] ME Tuckerman, CJ Mundy, and GJ Martyna. “On the classical statistical mechanics of non-Hamiltonian systems”. In: *Europhysics Letters* 45.2 (1999), p. 149.

[62] Mark E Tuckerman et al. “Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems”. In: *The Journal of Chemical Physics* 115.4 (2001), pp. 1678–1702.

[63] Johan Ludwig William Valdemar Jensen. “Sur les fonctions convexes et les inégalités entre les valeurs moyennes”. In: *Acta mathematica* 30.1 (1906), pp. 175–193.

[64] Stephen Abbott. *Understanding analysis*. Springer, 2015.

[65] Yu Cao and Eric Vanden-Eijnden. “Learning Optimal Flows for Non-Equilibrium Importance Sampling”. In: *Adv. Neural Inf. Process.* Ed. by Alice H. Oh et al. 2022.

[66] Kangxin Liu et al. “Computing equilibrium free energies through a nonequilibrium quench”. In: *The Journal of Chemical Physics* 160.3 (2024).

[67] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Elsevier, 2001.

[68] Jérôme Hénin et al. “Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1. 0]”. In: *LiveCoMS* 4.1 (2022), pp. 1583–1583.

[69] Christopher Jarzynski. “Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach”. In: *Phys. Rev. E* 56.5 (1997), p. 5018.

[70] Suriyanarayanan Vaikuntanathan and Christopher Jarzynski. “Escorted free energy simulations: Improving convergence by reducing dissipation”. In: *Phys. Rev. Lett.* 100.19 (2008), p. 190601.

[71] Stefano Martiniani. “On the complexity of energy landscapes: algorithms and a direct test of the Edwards conjecture”. PhD thesis. University of Cambridge, 2017.

[72] Achille Thin et al. “NEO: Non equilibrium sampling on the orbits of a deterministic transform”. In: *Adv. Neural Inf. Process.* Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 17060–17071.

[73] Steve Plimpton. “Fast parallel algorithms for short-range molecular dynamics”. In: *J. Comp. Phys.* 117.1 (1995), pp. 1–19.

[74] Aidan P Thompson et al. “LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales”. In: *Comp. Phys. Comm.* 271 (2022), p. 108171.

[75] Benedict Leimkuhler and Charles Matthews. “Robust and efficient configurational molecular sampling via Langevin dynamics”. In: *J. Chem. Phys.* 138.17 (2013), 05B601_1.

[76] Scott Kirkpatrick, C Daniel Gelatt Jr, and Mario P Vecchi. “Optimization by simulated annealing”. In: *Science* 220.4598 (1983), pp. 671–680.

[77] F Aluffi-Pentini, V Parisi, and F Zirilli. “Global optimization and stochastic differential equations”. In: *J Optim Theory Appl* 47 (1985), pp. 1–16.

[78] Enzo Marinari and Giorgio Parisi. “Simulated tempering: a new Monte Carlo scheme”. In: *Europhys. Lett.* 19.6 (1992), p. 451.

[79] Yadu Babuji et al. “Parsl: Pervasive parallel programming in python”. In: *Proceedings of the 28th International Symposium on High-Performance Parallel and Distributed Computing*. 2019, pp. 25–36.

[80] Lucas Frese Grønbech Jensen and Niels Grønbech-Jensen. “Accurate configurational and kinetic statistics in discrete-time Langevin systems”. In: *Mol. Phys.* 117.18 (2019), pp. 2511–2526.

[81] Alexander D MacKerell Jr, Nilesh Banavali, and Nicolas Foloppe. “Development and current status of the CHARMM force field for nucleic acids”. In: *Biopolymers: Original Research on Biomolecules* 56.4 (2000), pp. 257–265.

[82] Massimiliano Bonomi et al. “Promoting transparency and reproducibility in enhanced molecular simulations”. In: *Nature methods* 16.8 (2019), pp. 670–673.

[83] Gareth A Tribello et al. “PLUMED 2: New feathers for an old bird”. In: *Comp. Phys. Comm.* 185.2 (2014), pp. 604–613.

[84] Massimiliano Bonomi et al. “PLUMED: A portable plugin for free-energy calculations with molecular dynamics”. In: *Computer Physics Communications* 180.10 (2009), pp. 1961–1972.

[85] Benoıˆt Roux. “The calculation of the potential of mean force using computer simulations”. In: *Comp. Phys. Comm.* 91.1-3 (1995), pp. 275–282.

[86] Xinqiang Ding, Jonah Z Vilseck, and Charles L Brooks III. “Fast solver for large scale multistate Bennett acceptance ratio equations”. In: *Journal of chemical theory and computation* 15.2 (2019), pp. 799–802.

[87] Sunhwan Jo et al. “CHARMM-GUI: a web-based graphical user interface for CHARMM”. In: *J. Comp. Chem.* 29.11 (2008), pp. 1859–1865.

[88] Jumin Lee et al. “CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field”. In: *J. Chem. Theor. Comput.* 12.1 (2016), pp. 405–413.

[89] Jumin Lee et al. “CHARMM-GUI supports the Amber force fields”. In: *J. Chem. Phys.* 153.3 (2020), p. 035103.

[90] Stefano Martiniani and Mathias Casiulis. “When you can’t count, sample! Computable entropies beyond equilibrium from basin volumes”. In: *Papers in Physics* 15 (2023), pp. 150001–150001.

[91] Ludwig Boltzmann. “On the relationship between the second fundamental theorem of the mechanical theory of heat and probability calculations regarding the conditions for thermal equilibrium”. In: *Entropy* 17.4 (2015), pp. 1971–2009.

[92] Ning Xu, Daan Frenkel, and Andrea J Liu. “Direct determination of the size of basins of attraction of jammed solids”. In: *Physical Review Letters* 106.24 (2011), p. 245502.

[93] Daniel Asenjo, Fabien Paillusson, and Daan Frenkel. “Numerical calculation of granular entropy”. In: *Physical review letters* 112.9 (2014), p. 098002.

[94] Stefano Martiniani et al. “Turning intractable counting into sampling: Computing the configurational entropy of three-dimensional jammed packings”. In: *Physical Review E* 93.1 (2016), p. 012906.

[95] Stefano Martiniani et al. “Numerical test of the Edwards conjecture shows that all packings are equally probable at jamming”. In: *Nature physics* 13.9 (2017), pp. 848–851.

[96] David J Smith and Mavina K Vamanamurthy. “How small is a unit ball?” In: *Mathematics Magazine* 62.2 (1989), pp. 101–107.

[97] David Edwin George Hare. “Computing the principal branch of log-Gamma”. In: *Journal of Algorithms* 25.2 (1997), pp. 221–236.

[98] Daniel W Lozier. “NIST digital library of mathematical functions”. In: *Annals of Mathematics and Artificial Intelligence* 38 (2003), pp. 105–119.

[99] M Ostrogradsky. “Mémoire sur le calcul des variations des intégrales multiples.” In: (1836).

[100] Paul C Matthews. *Vector calculus*. Springer Science & Business Media, 2000.

[101] Ramamurti Shankar. *Principles of quantum mechanics*. Springer Science & Business Media, 2012.

[102] Sheldon Axler. *Linear algebra done right*. Springer, 2015.

[103] Alan C Hindmarsh et al. “SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers”. In: *ACM Transactions on Mathematical Software (TOMS)* 31.3 (2005), pp. 363–396. DOI: 10.1145/1089014.1089020.

[104] Cody J Balos et al. “Enabling GPU accelerated computing in the SUNDIALS time integration library”. In: *Parallel Computing* 108 (2021), p. 102836.

[105] David J Gardner et al. “Enabling new flexibility in the SUNDIALS suite of nonlinear and differential/algebraic equation solvers”. In: *ACM Transactions on Mathematical Software (TOMS)* (2022). DOI: 10.1145/3539801.

[106] Praharsh Suryadevara, Mathias Casiulis, and Stefano Martiniani. “Mirages in the Energy Landscape of Soft Sphere Packings”. In: *arXiv preprint arXiv:2409.12113* (2024).