CONSENSUS-BASED ADAPTIVE SAMPLING AND APPROXIMATION FOR HIGH-DIMENSIONAL ENERGY LANDSCAPES *

LIYAO LYU† AND HUAN LEI‡

Abstract. We present a consensus-based framework that unifies phase space exploration with posterior-residual-based adaptive sampling for surrogate construction in high-dimensional energy landscapes. Unlike standard approximation tasks where sampling points can be freely queried, systems with complex energy landscapes such as molecular dynamics (MD) do not have direct access to arbitrary sampling regions due to the physical constraints and energy barriers; the surrogate construction further relies on the dynamical exploration of phase space, posing a significant numerical challenge. We formulate the problem as a minimax optimization that jointly adapts both the surrogate approximation and residual-enhanced sampling. The construction of free energy surfaces (FESs) for high-dimensional collective variables (CVs) of MD systems is used as a motivating example to illustrate the essential idea. Specifically, the maximization step establishes a stochastic interacting particle system to impose adaptive sampling through both exploitation of a Laplace approximation of the max-residual region and exploration of uncharted phase space via temperature control. The minimization step updates the FES surrogate with the new sample set. Numerical results demonstrate the effectiveness of the present approach for biomolecular systems with up to 30 CVs. While we focus on the FES construction, the developed framework is general for efficient surrogate construction for complex systems with high-dimensional energy landscapes.

Key words. Consensus-based sampling, Adaptive sampling, Minimax optimization, Phase space exploration, High-dimensional free energy

MSC codes. 65C35, 65K10, 82C31

1. Introduction.

Many problems in computational science rely on the efficient approximation of physical quantities in systems governed by high-dimensional energy landscapes, where direct sampling and evaluation are constrained by complex system dynamics. Canonical examples include the computation of transition paths in gradient systems [49, 13], the construction of coarse-grained molecular dynamics (MD) models [35], and uncertainty quantification under energy-induced probability measures [29]. Unlike the standard surrogate construction tasks where the sampling points can be freely queried, these problems further rely on efficient exploration and sampling over the phase space, where the thermodynamically accessible regions are often unknown *a priori*. There are two essential challenges: (I) the prevalence of energy barriers, which makes direct sampling inefficient and prone to getting trapped in local minima; various enhanced sampling strategies are often required; (II) the high dimensionality of the surrogate model, which generally requires a large number of samples and motivates adaptive sampling strategies based on the approximation error of the target quantity. In practice, the efficient surrogate construction should account for both enhanced sampling with the complex energy landscapes and the residual-based adaptivity in constrained phase space — yet simultaneously addressing both remains a nontrivial and open computational challenge.

*Submitted to the editors DATE.

Funding: The work is supported in part by the National Science Foundation under Grant DMS-2110981 and the ACCESS program through allocation MTH210005.

†Department of Computational Mathematics, Science & Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, 48824, MI, USA (lyuliyao@msu.edu).

‡Department of Computational Mathematics, Science & Engineering, Michigan State University, 428 S Shaw Ln, Department of Statistics & Probability, Michigan State University, Wells Hall, 619 Red Cedar Road, East Lansing, 48824, MI, USA. (leihuan@msu.edu).

In this study, we aim to develop a unified approach that enables both efficient phase space exploration and residual-based adaptive sampling for constrained phase space learning problems. As a motivating application, we consider the construction of free energy surfaces (FESs) with respect to a set of collective variables (CVs) for MD systems. While this is a long-standing problem in computational science, the accurate construction of high-dimensional FESs remains a difficult task due to the aforementioned two challenges. Most existing methods primarily target the first challenge based on various importance sampling strategies to overcome energy barriers, such as umbrella sampling [44], histogram reweighting [26], metadynamics [27, 18], variationally enhanced sampling [46, 40, 4], and adaptive biasing force [10, 11, 28, 9]. Alternatively, temperature-accelerated [32] and adiabatic dynamics [39, 1] introduce an extended dynamics of the CVs with an artificially high temperature to facilitate the phase space exploration. Despite their broad applications, the computational efficiency of these methods generally degrades as the number of CVs increases. More importantly, these methods do not explicitly address the second challenge. The residual error is not incorporated into the adaptive sampling and FES construction, which limits their effectiveness in high-dimensional problems.

From a different perspective, several approaches related to the second challenge have been developed based on adaptive sampling [42, 51, 43, 16, 14, 23, 20] and adversarial learning [52, 3, 53] for solving high-dimensional partial differential equations (PDEs). The essential idea is to introduce certain residual-based distributions or weak formulations, where new collocation points or test basis functions can be adaptively updated during the training process. While they have shown promising results for high-dimensional PDEs, these methods rely on the free query of new sample points and their residual error within the domain. As such, they cannot be directly applied to the present problem, where the global residual error is unknown *a priori*. In particular, the new sample points cannot be freely placed within the phase space but need to be navigated through dynamical exploration of the thermodynamically accessible regions. Alternatively, the reinforced dynamics (RiD) [54] (see also Ref. [47]) proposes using the uncertainty indicator as a proxy for the residual error to bias MD simulations, which, however, relies on calculating the standard deviation of the predictions from an ensemble of neural network (NN) surrogates trained on the same dataset. Moreover, the phase space exploration is constrained by the underlying energy landscape, which typically requires small time steps due to the stiffness and roughness of the MD potential function.

To address the above two challenges, we present a consensus-based adaptive sampling (CAS) method to efficiently construct surrogates within high-dimensional energy landscapes with applications to FES construction in complex MD models. A unique feature is that the present method enables gradient-free residual-based adaptive sampling such that the FES approximation and the phase-space exploration can be simultaneously optimized. The method is formulated as a minimax optimization problem. The max-problem seeks a residual-based distribution to establish adaptive sampling in the vicinity of the explored phase space regime, while the min-problem optimizes the FES surrogate based on the new samples. For the maximization step, we emphasize that the establishment of the residual-based distribution is only formal; the analytical form of the distribution is unknown and the value at an arbitrary point can not be directly obtained. As a result, common sampling approaches based on Markovian Chain Monte Carlo (MCMC) [37] and Langevin dynamics [38] are not applicable, as explained in more detail in Sec. 2.2. Instead, we establish a consensus-based sampling [7] (see also Refs. [6, 8]) in the form of a stochastic interacting particle

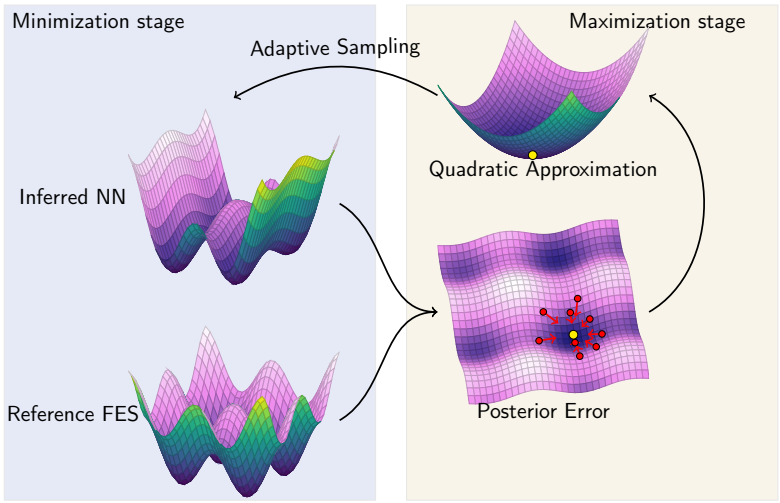

FIG. 1. *The workflow of the present CAS-based method. In the minimization step, given a collection of sampling points, the reference force* $F(\mathbf{z})$ *(i.e., the gradient of the underlying FES) can be calculated using the restrained molecular dynamics; a comparison with the force inferred from the NN surrogate* $A_{\mathcal{N}}(\mathbf{z})$ *yields the loss function* $L_{\mathcal{N}}(\mathbf{z}) = |\nabla_{\mathbf{z}} A_{\mathcal{N}}(\mathbf{z}) + F(\mathbf{z})|^2$*. In the maximization step, the loss function determines a residual-based distribution with entropy regularization* $q^*(\mathbf{z}) \propto \exp(\kappa_h \mathcal{L}_{\mathcal{N}}(\mathbf{z}))$*. An interacting particle system following the McKean SDE is used to achieve adaptive sampling of the max-residual region by modulating the exploitation of the Laplace approximation of* $q^*(\mathbf{z})$ *and the exploration of the uncharted phase space. The FES can be accurately reconstructed after several iterations of the minimization and maximization step.*

system governed by a McKean stochastic differential equation (SDE). The gradient-free nature enables us to collect new samples adaptive to the local residual without the analytical form of the target distribution. Specifically, a quadratic potential is adaptively constructed to probe the local max-residual regime by exploiting the Laplace approximation under a low-temperature limit. Meanwhile, a coherent noise term is introduced to efficiently explore the full CV space under a high-temperature limit and yield the updated sampling points used for the subsequent minimization step.

The present iterative procedure achieves adversarial learning of the FES pertaining to the thermodynamically important regions. In contrast to existing adaptive sampling methods for PDE solvers [42, 43, 16], the present method does not rely on the free query of arbitrary sample points. Instead, it enables us to establish a dynamical exploration of the phase space along with the surrogate construction. Moreover, unlike the reinforced dynamics [54] that navigates the sampling points through the biased MD simulations, the present sampling dynamics is governed by a smooth quadratic potential irrespective of the roughness of the underlying energy landscape, enabling much larger time steps and more efficient phase-space exploration. We demonstrate the effectiveness of the proposed method by constructing the FES of biomolecule systems involving up to 30 CVs. Fig. 1 sketches the workflow of the proposed method.

2. Methods.

2.1. Free energy and mean forces. We consider a full model with micro-scale coordinates $\mathbf{r} \in \mathbb{R}^N$ whose dynamics is governed by potential $U(\mathbf{r}) : \mathbb{R}^N \to \mathbb{R}$ under temperature T. Suppose we are interested in CVs $\mathbf{s}(\mathbf{r}) : \mathbb{R}^N \to \Gamma$ with $\Gamma \subset \mathbb{R}^M$, the

FES $A(\mathbf{z})$ of the CVs is defined by

$$A(\mathbf{z}) = -\frac{1}{\beta}\ln\rho(\mathbf{z}), \tag{2.1}$$

where $\beta = 1/k_BT$ is the inverse of the thermal temperature,

$$\rho(\mathbf{z}) = \frac{1}{Z}\int \exp\left(-\beta U(\mathbf{r})\right)\delta(\mathbf{s}(\mathbf{r}) - \mathbf{z})\mathrm{d}\mathbf{r} \tag{2.2}$$

is the marginal probability density function (PDF) for $\mathbf{s}(\mathbf{r}) = \mathbf{z}$, $\delta(\cdot)$ represents the Dirac delta function and $Z = \int \exp\left(-\beta U(\mathbf{r})\right)\mathrm{d}\mathbf{r}$ is the partition function; we refer to Ref. [41] for details. For high-dimensional CVs, direct estimation of $\rho(\mathbf{z})$ often becomes numerically challenging. An alternative approach is to fit the mean force $\mathbf{F}(\mathbf{z}) := -\nabla A(\mathbf{z})$ at various sample points, which can be estimated via the restrained MD [2] by introducing a harmonic term into the full potential, i.e.,

$$U_k(\mathbf{r}, \mathbf{z}) = U(\mathbf{r}) + \frac{k}{2}(\mathbf{s}(\mathbf{r}) - \mathbf{z})^\top(\mathbf{s}(\mathbf{r}) - \mathbf{z}), \tag{2.3}$$

where k represents the magnitude of the restrained potential. As shown in Ref [32], the mean force can be computed by $\mathrm{F}(\mathbf{z}) = \lim_{k\to\infty}\mathrm{F}^k(\mathbf{z})$, where $\mathrm{F}^k(\mathbf{z})$ is defined by

$$\mathrm{F}^k(\mathbf{z}) = \frac{1}{Z_k(\mathbf{z})}\int k(\mathbf{s}(\mathbf{r}) - \mathbf{z})\exp\left(-\beta U_k(\mathbf{r}, \mathbf{z})\right)\mathrm{d}\mathbf{r}, \tag{2.4}$$

and can be sampled as the first-moment estimation.

In principle, given a collection of sample points, $A(\mathbf{z})$ can be re-constructed (up to a constant) by matching the mean force $-\nabla A(\mathbf{z})$ at the individual points. However, for MD systems with complex energy landscape, the sampling over the phase space could be highly non-trivial due to the prevalence of energy local minima; the training set is often determined *a prior* as pre-selected points or in a greedy manner. As the number of CVs increases, the empirical random samples often introduce pronounced discretization error. To efficiently construct $A(\mathbf{z})$ in the thermodynamically accessible region, it is desirable to simultaneously optimize the training set and the FES surrogate approximation through certain adaptive sampling based on the *posterior* residual error. This motivates the present method illustrated below.

2.2. Min-Max formulation. Let $A_{\mathcal{N}}(\mathbf{z})$ denote the NN surrogate of the FES $A(\mathbf{z})$, which is parameterized by minimizing the loss function

$$\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = \left|\nabla_{\mathbf{z}}A_{\mathcal{N}}(\mathbf{z}) + \mathrm{F}(\mathbf{z})\right|^2 \tag{2.5}$$

for $\mathbf{z} \in \Gamma$. To solve the minimization problem, we introduce a sampling distribution $q(\mathbf{z})$ and define the weighted loss

$$(\mathcal{L}_{\mathcal{N}}, q) = \int_\Gamma \mathcal{L}_{\mathcal{N}}(\mathbf{z})q(\mathbf{z})\mathrm{d}\mathbf{z}. \tag{2.6}$$

A desired distribution intends to maximize the discrepancy in the dataset for the current NN surrogate $A_{\mathcal{N}}(\mathbf{z})$. Accordingly, we define the maximization problem as

$$J[\mathcal{L}_{\mathcal{N}}] = \max_q(\mathcal{L}_{\mathcal{N}}, q). \tag{2.7}$$

Since $(\mathcal{L}_{\mathcal{N}}, q)$ is always non-negative, a good approximation of the original free energy surface $A_{\mathcal{N}}(\mathbf{z})$ (up to a constant) can be obtained by solving the following problems

$$\min_{A_{\mathcal{N}}}\max_q(\mathcal{L}_{\mathcal{N}}, q). \tag{2.8}$$

PROPOSITION 2.1. *Assuming that there exists a solution of* $\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = 0$ *for* $\mathbf{z} \in \Gamma$*, then* $A^*_{\mathcal{N}}$ *is a solution if and only if it solves* (2.8).

Proof. Suppose $A^*_{\mathcal{N}}$ is the solution for $\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = 0$, it satisfies $(\mathcal{L}_{\mathcal{N}^*}, q) = 0$ for any q, i.e., $J[\mathcal{L}_{\mathcal{N}^*}] = 0$. Therefore, $A^*_{\mathcal{N}}$ is a solution for the minimax problem (2.8). On the other hand, if $\hat{A}_{\mathcal{N}}$ is the minimizer for problem (2.8) but not the solution for $\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = 0$, then there exists $\hat{q} \in V$ such that $(\mathcal{L}_{\hat{\mathcal{N}}}, \hat{q}) > 0$. However, $(\mathcal{L}_{\mathcal{N}^*}, q) = 0$ for all q, which contradicts the assumption that $\hat{A}_{\mathcal{N}}$ is the minimizer. ∎

Proposition 2.1 shows that direct construction of FES $A(\mathbf{z})$ can be reformulated as an adversarial learning of an optimal solution $A_{\mathcal{N}}(\mathbf{z})$ for the min-max problem (2.8). Accordingly, the training consists of two components: the minimization step optimizes the NN surrogate with the current training set; the maximization step explores the largest residual region for the current surrogate and essentially establishes an adaptive sampling of the training set based on the *posterior* residual $\mathcal{L}_{\mathcal{N}}(\mathbf{z})$.

To numerically solve the maximization problem, certain regularization needs to be introduced. Otherwise, the maximization problem will simply yield a delta measure, i.e., $\delta(\mathbf{z}-\mathbf{z}^*)$, where $\mathbf{z}^* = \arg\max \mathcal{L}_{\mathcal{N}}(\mathbf{z})$. Since the sampling needs to simultaneously identify the max-residual regime and explore the uncharted phase space, we introduce the entropy-based regularization [48, 15] (see also Refs. [19, 34, 43] for gradient-based regularization), and the max-problem is reformulated by

$$\min_{q} \int (-\mathcal{L}_{\mathcal{N}}(\mathbf{z}) + \kappa_h^{-1} \ln q(\mathbf{z}))q(\mathbf{z})\mathrm{d}\mathbf{z}. \tag{2.9}$$

This problem is convex with a unique global minimum at $q^*(\mathbf{z}) = \exp(-\kappa_h \mathcal{L}^-_{\mathcal{N}}(\mathbf{z}))/Z^*$, where $Z^* = \int \exp(-\kappa_h \mathcal{L}^-_{\mathcal{N}}(\mathbf{z}))\mathrm{d}\mathbf{z}$ and $\mathcal{L}^-_{\mathcal{N}}(\mathbf{z}) = -\mathcal{L}_{\mathcal{N}}(\mathbf{z})$. The parameter κ_h is a Lagrangian multiplier, balancing the weight between the max-residual concentration and the scope of exploration, and can be interpreted as the inverse of temperature in statistical physics. An elevated κ_h^{-1} induces a distribution closer to a uniform distribution. Conversely, a diminished κ_h^{-1} induces a distribution concentrated near the max-residual point.

Although $q^*(\mathbf{z}) = \exp(-\kappa_h \mathcal{L}^-_{\mathcal{N}}(\mathbf{z}))/Z^*$ solves Eq. (2.9), we emphasize that this solution is only formal. In particular, the analytical formulation is unknown since it further depends on the residual $\mathcal{L}_{\mathcal{N}}$ (2.5) and the exact FES $A(\mathbf{z})$. Unfortunately, the numerical evaluation of $\nabla_{\mathbf{z}} A(\mathbf{z}^*)$ at an arbitrary point relies on conducting restrained MD at $\mathbf{z} = \mathbf{z}^*$, which can be computationally expensive or even thermodynamically inaccessible. This constraint differs from conventional high-dimensional regression or PDE problems, where the residual can be freely queried within the domain. Consequently, common sampling approaches based on the MCMC [37] and Langevin dynamics [38] can not be directly used. Specifically, the effectiveness of the MCMC method relies on the variance of the underlying distribution. However, most configurations in the MD phase space are thermodynamically inaccessible, resulting in large variance and extremely low sampling efficiency. On the other hand, the Langevin-based sampling method depends on $-\nabla_{\mathbf{z}} \ln q^*(\mathbf{z})$, which, in turn, relies on computing $\nabla_{\mathbf{z}} \mathcal{L}^-_{\mathcal{N}}(\mathbf{z})$ and essentially the gradient of the mean force, i.e., $\nabla_{\mathbf{z}} F(\mathbf{z}) = -\nabla_{\mathbf{z}} \otimes \nabla_{\mathbf{z}} A(\mathbf{z})$ from restrained MD simulations. However, we can only sample the value of the mean force $F(\mathbf{z})$ rather than its gradient.

Inspired by the consensus-based sampling method [7], we propose a modified interacting particle system governed by the McKean SDE. The gradient-free nature enables us to efficiently construct the mean-field non-local conservative potential and

the coherent noise without the evaluation of $\nabla_{\mathbf{z}} F(\mathbf{z})$, and sample the particle distribution as an accurate approximation of $q^*(\mathbf{z})$ illustrated below.

2.3. Exploitation and exploration in the max-problem. To approximate the target distribution $q^*(\mathbf{z})$, particularly in the vicinity of the max-residual point $\mathbf{z}^*$, we exploit Laplace's principle in the large deviations theory, i.e.,

$$\lim_{\kappa\to\infty}\left(-\frac{1}{\kappa}\log\left(\int \exp\left(-\kappa f(\mathbf{z})\right)\mathrm{d}\rho^*(\mathbf{z})\right)\right) = f(\mathbf{z}^*), \tag{2.10}$$

which holds true for any compactly supported probability measure ρ^*, where $\mathbf{z}^* \in \mathrm{supp}(\rho^*)$ uniquely minimizes the function f. This enables us to identify the max-residual point from a collection of samples $\left\{\mathbf{z}^i\right\}_{i=1}^{N_w}$ by the first-order moment under the weighted density function $p(\mathbf{z})$, i.e.,

$$\mathbf{m}_{\kappa_l} = \int \mathbf{z}\frac{p(\mathbf{z})}{\int p(\mathbf{z}')\rho(\mathrm{d}\mathbf{z}')}\rho(\mathrm{d}\mathbf{z}) \approx \sum_{i=1}^{N_w}\mathbf{z}^i\hat{p}(\mathbf{z}^i) \quad \hat{p}(\mathbf{z}) = \frac{\exp\left(-\kappa_l\mathcal{L}_{\mathcal{N}}^-(\mathbf{z})\right)}{\sum_{i=1}^{N_w}\exp\left(-\kappa_l\mathcal{L}_{\mathcal{N}}^-(\mathbf{z}^i)\right)} \tag{2.11}$$

where κ_l^{-1} represents a low temperature limit, $p(\mathbf{z}) = \exp(-k_l\mathcal{L}_{\mathcal{N}}^-(\mathbf{z}))$ and $\rho = \frac{1}{N_w}\sum_{i=1}^{N_w}\delta_{\mathbf{z}^i}$ is the empirical measure. In order to sample the empirical measure towards q^*, we treat each sampler $\mathbf{z}^i$ as a random walker $\mathbf{z}_t^i$ governed by the following McKean SDE

$$\mathrm{d}\mathbf{z}_t^i = -\frac{1}{\gamma}\nabla_{\mathbf{z}}G(\mathbf{z}_t^i)\mathrm{d}t + \sqrt{\frac{2}{\kappa_h\gamma}}\mathrm{d}\mathbf{W}_t^i, \quad i = 1,\cdots,N_w, \tag{2.12}$$

where $G(\mathbf{z}_t) = \frac{1}{2}(\mathbf{z}_t - \mathbf{m}_{\kappa_l,t})^\top V_{\kappa_l,t}^{-1}(\mathbf{z}_t - \mathbf{m}_{\kappa_l,t})$ denotes an adaptively constructed mean-field conservative potential function. The formulations of $\mathbf{m}_{\kappa_l,t}, V_{\kappa_l,t}$ are specified in (2.13) with the rationale discussed in the next section. Consequently, $G(\mathbf{z}_t)$ navigates the random walkers (i.e., individual particles) towards $\mathbf{m}_{\kappa_l,t}$, which represents the region of large residual error. The second term in Eq. (2.12) represents a noise term where γ represents the friction coefficient and $\mathbf{W}_t^i$ represents the standard M-dimensional Brownian motion.

The coupling of the conservative and stochastic terms maintains a relatively high temperature κ_h^{-1}. Also, a large friction coefficient γ is applied such that the distribution of walkers $q_t(\mathbf{z})$ is almost always Gaussian during the evolution. As shown in Prop. 2.3, by properly choosing the form of $V_{\kappa_l,t}$, the distribution $q_t(\mathbf{z})$ converges to $\propto \exp(-\kappa_h G(\mathbf{z}))$ characterized by $\mathbf{m}_{\kappa_l,\infty}$ and $V_{\kappa_l,\infty}$. Accordingly, the balance between exploitation and exploration is controlled using two temperatures κ_l^{-1} and κ_h^{-1}. As κ_l^{-1} decreases, the distribution concentrates near the max-residual points, reflecting the role of exploitation. Conversely, as κ_h^{-1} increases, the distribution smoothens progressively, enhancing the exploration of the uncharted regions.

Remark 2.2. The present sampling dynamics (2.12) takes a different form from the one proposed in Ref. [7]. Specifically, the conservative potential $G(\mathbf{z})$ is constructed by both the first and second moment which enables us to conveniently characterize $\mathcal{L}_{\mathcal{N}}(\mathbf{z})$ near the max-residual region with Laplace's approximation, whereas it is only determined by the first moment in Ref. [7]. Also, κ_h is introduced as an independent parameter to modulate the exploration of the phase space. In principle, the consensus dynamics in Ref. [7] could be also used for adaptive sampling with the proper choice of the multiplicative noise term. We proposed the modified form (2.12) such that the sampling parameters can be chosen with a clear physical interpretation.

2.4. Convergence analysis. In this subsection, we analyze the long-time behavior of the sampling dynamics (2.12). We show that the particle distribution converges exponentially fast to a steady state as an approximation of the target residual-based distribution $q^*(\mathbf{z}) \propto \exp(-\kappa_h \mathcal{L}_{\mathcal{N}}^-(\mathbf{z}))$ under mild conditions. Specifically, we choose $\mathbf{m}_{\kappa_l,t}$ and $V_{\kappa_l,t}$ in the form of

$$\mathbf{m}_{\kappa_l,t} = \mathcal{M}_{\kappa_l}(\rho_t) \quad V_{\kappa_l,t} = \mathcal{V}_{\kappa_l}(\rho_t), \tag{2.13}$$

where $\rho_t = \frac{1}{N_w}\sum_{i=1}^{N_w} \delta_{\mathbf{z}_t^i}$ is the empirical measure and

$$\begin{aligned} \mathcal{M}_{\kappa_l}(\rho) &= \int \mathbf{z} \frac{p(\mathbf{z})}{\int p(\mathbf{z}')\rho(\mathrm{d}\mathbf{z}')}\rho(\mathrm{d}\mathbf{z}), \\ \mathcal{V}_{\kappa_l}(\rho) &= \kappa_t \int (\mathbf{z} - \mathcal{M}_{\kappa_l}(\rho)) \otimes (\mathbf{z} - \mathcal{M}_{\kappa_l}(\rho)) \frac{p(\mathbf{z})}{\int p(\mathbf{z}')\rho(\mathrm{d}\mathbf{z}')}\rho(\mathrm{d}\mathbf{z}). \end{aligned} \tag{2.14}$$

In particular, we show that by choosing $\kappa_t = \kappa_l + \kappa_h$, Eq. (2.12) converges to the target distribution $q^*(\mathbf{z})$.

PROPOSITION 2.3. *Suppose $\mathcal{L}_{\mathcal{N}}^-(\mathbf{z})$ takes a local quadratic approximation in form of $\frac{1}{2}(\mathbf{z}-\mu)^\top \Sigma^{-1}(\mathbf{z}-\mu)$. If the dynamics converge to an invariant distribution, then the stationary density is given by*

$$q_\infty = \frac{\exp\left(-\kappa_h \mathcal{L}_{\mathcal{N}}^-(\mathbf{z})\right)}{\int \exp\left(-\kappa_h \mathcal{L}_{\mathcal{N}}^-(\mathbf{z})\right)\mathrm{d}\mathbf{z}}, \tag{2.15}$$

by choosing $\kappa_t = \kappa_l + \kappa_h$.

Proof. Let $q_\infty(\mathbf{z})$ denote the invariant distribution of Eq. (2.12). Then $q_\infty(\mathbf{z})$ must be the invariant distribution of the following SDE

$$\mathrm{d}\mathbf{z} = -\frac{1}{\gamma} V_{\kappa_l,\infty}^{-1}(\mathbf{z} - \mathbf{m}_{\kappa_l,\infty})\mathrm{d}t + \sqrt{\frac{2}{\gamma\kappa_h}}\mathrm{d}\mathbf{W}_t, \tag{2.16}$$

where $\mathbf{m}_{\kappa_l,\infty}$ and $\kappa_t^{-1}V_{\kappa_l,\infty}$ are the mean and the covariance matrix of the re-weighted density $\propto\ q_\infty(\mathbf{z})e^{-\kappa_l \mathcal{L}_{\mathcal{N}}^-(\mathbf{z})}$. With the fluctuation-dissipation relation for Eq. (2.16), we can show $q_\infty(\mathbf{z})$ follows the Gaussian distribution with mean $\mathbf{m}_{\kappa_l,\infty}$ and covariance matrix $\kappa_h^{-1}V_{\kappa_l,\infty}$.

Since $\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mu)^\top\Sigma^{-1}(\mathbf{z}-\mu)$ is quadratic, the re-weighted density of a Gaussian distribution $q(\mathbf{z}) \sim \mathcal{N}(\mathbf{m}, V)$ remains Gaussian, i.e.,

$$q(\mathbf{z})e^{-\kappa_l \mathcal{L}_{\mathcal{N}}^-(\mathbf{z})} \propto\ \mathcal{N}(\mathbf{m}_{\kappa_l}, V_{\kappa_l}),$$

where $\mathbf{m}_{\kappa_l}$ and V_{κ_l} are defined by

$$\begin{aligned} \mathbf{m}_{\kappa_l}(\mathbf{m}, V) &= (V^{-1} + \kappa_l\Sigma^{-1})^{-1}(\kappa_l\Sigma^{-1}\mu + V^{-1}\mathbf{m}), \\ V_{\kappa_l}(\mathbf{m}, V) &= (V^{-1} + \kappa_l\Sigma^{-1})^{-1}. \end{aligned} \tag{2.17}$$

Therefore, the mean and covariance of the steady-state Gaussian distribution satisfies

$$\begin{aligned} \mathbf{m}_{\kappa_l,\infty} &= (\kappa_h V_{\kappa_l,\infty}^{-1} + \kappa_l\Sigma^{-1})(\kappa_l\Sigma^{-1}\mu + \kappa_h V_{\kappa_l,\infty}^{-1}\mathbf{m}_{\kappa_l,\infty}), \\ V_{\kappa_l,\infty} &= \kappa_t(\kappa_h V_{\kappa_l,\infty}^{-1} + \kappa_l\Sigma^{-1})^{-1}. \end{aligned}$$

It is easy to show that by choosing $\kappa_t = \kappa_l + \kappa_h$, $\mathbf{m}_{\kappa_l,\infty}$ and $V_{\kappa_l,\infty}$ recovers μ and Σ, respectively, and the invariant distribution takes the form

$$q_\infty(\mathbf{z}) \sim \mathcal{N}\left(\mu, \kappa_h^{-1}\Sigma\right).$$

□

We emphasize that Eqs. (2.12) and (2.13) differ from the standard Langevin dynamics; the sampling relies on neither the explicit knowledge of the target distribution $q^*(\mathbf{z}) \propto \exp(-\kappa_h \mathcal{L}_{\mathcal{N}}^-(\mathbf{z}))$ nor the numerical evaluation of $\nabla_{\mathbf{z}} q^*(\mathbf{z})$ for individual sampling points. The quadratic assumption of the loss function $\mathcal{L}_{\mathcal{N}}^-(\mathbf{z})$ is due to the fact that we are mainly interested in the thermodynamically accessible region near the max-residual point. Under the low-temperature limit, the local regime can be well characterized by the first and second moment following Laplace's principle.

Next, we show that, under appropriate conditions, the sampling dynamics (2.12) converges to the target distribution exponentially fast. In particular, we consider the time discretization given by

$$\mathbf{z}_{t+1}^i = \mathbf{z}_t^i - \frac{\kappa_t^{-1}\mathcal{V}_{\kappa_l}^{-1}(\rho_t)}{\gamma}\left(\mathbf{z}_t^i - \mathcal{M}_{\kappa_l}(\rho_t)\right)\delta t + \sqrt{\frac{2\delta t}{\gamma\kappa_h}}\eta_t^i, \tag{2.18}$$

where η_t^i are independent $\mathcal{N}(0, \mathbb{I}_M)$ random variables and δt is the time step.

LEMMA 2.4. *If the initial law ρ_0 for (2.18) is Gaussian, so is the law for any $t \in \mathbb{N}_{>0}$. Also, its evolution is characterized by the first and second moments $(\mathbf{m}_t, V_t)$ of ρ_t, which are governed by the recurrence relation*

$$\begin{aligned} \mathbf{m}_{t+1} &= \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1} V_{k_l}^{-1}(\mathbf{m}_t, V_t)}{\gamma}\right)\mathbf{m}_t + \frac{\delta t \kappa_t^{-1} V_{k_l}^{-1}(\mathbf{m}_t, V_t)}{\gamma}\mathbf{m}_{k_l}(\mathbf{m}_t, V_t), \\ V_{t+1} &= \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1} V_{k_l}^{-1}(\mathbf{m}_t, V_t)}{\gamma}\right)^2 V_t + \frac{2\delta t}{\gamma\kappa_h}, \end{aligned} \tag{2.19}$$

where $\mathbf{m}_{k_l}(\mathbf{m}, V) = \mathcal{M}_{\kappa_l}(g(\cdot; \mathbf{m}, V))$ and $V_{k_l}(\mathbf{m}, V) = \mathcal{V}_{\kappa_l}(g(\cdot; \mathbf{m}, V))$.

Proof. The time evolution of the law of the solution (2.18) is governed by the following discrete-time dynamics on probability density:

$$\rho_{t+1}(\mathbf{z}) = \int_{\mathbb{R}^M} g\left(\mathbf{z}; \mathbf{z}' - \frac{\delta t \kappa_t^{-1} V_{\kappa_l}^{-1}(\rho_t)}{\gamma}\left(\mathbf{z}' - \mathcal{M}_{\kappa_l}(\rho_t)\right), \frac{2\delta t}{\gamma\kappa_h}\right)\rho_t(\mathbf{z}')\mathrm{d}\mathbf{z}', \tag{2.20}$$

where $g(\mathbf{z}; \mathbf{m}, V) = \frac{1}{\sqrt{(2\pi)^M \det(V)}}\exp(-\frac{1}{2}\left(\mathbf{z}-\mathbf{m}\right)^\top V^{-1}(\mathbf{z}-\mathbf{m})\big)$. Since both ρ_t and g are Gaussian, the convolution in (2.20) results in another Gaussian distribution. Therefore, ρ_t remains Gaussian for all $t > 0$. The update rules for the mean and covariance can be derived by explicitly computing the first and second moments of ρ_{t+1} from the integral form (2.20). ∎

PROPOSITION 2.5. *Assume that $\mathcal{L}_{\mathcal{N}}^-(\mathbf{z})$ takes a local quadratic approximation in form of $\frac{1}{2}(\mathbf{z}-\mu)^\top\Sigma^{-1}(\mathbf{z}-\mu)$ and initial condition $(\mathbf{m}_0, V_0) \in \mathbb{R}^M \times S_{++}^M$, where S_{++}^M denotes the set of symmetric strictly positive definite matrices in $\mathbb{R}^{M\times M}$. Then the first and second moment of the empirical distribution $\mathbf{m}_t$ and V_t converge to μ and $\kappa_h^{-1}\Sigma$ exponentially fast if δt is sufficiently small. To be more specific, we have*

$$\begin{aligned} \mathbf{m}_t &= \mu + \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1}\kappa_l\Sigma^{-1}}{\gamma}\right)^t(\mathbf{m}_0 - \mu) \\ V_t &= \kappa_h^{-1}\Sigma + \Sigma\left(\mathbb{I}_M - \frac{2\delta t\kappa_l\Sigma^{-1}}{\gamma(\kappa_l+\kappa_h)}\right)^t\left(\Sigma^{-1}V_0 - \kappa_h^{-1}\mathbb{I}_M\right). \end{aligned} \tag{2.21}$$

Proof. For the first moment, we have the following recurrence relation

$$\begin{aligned}\mathbf{m}_{t+1} &= \mathbf{m}_t - \frac{\delta t \kappa_t^{-1}(V_t^{-1} + \kappa_l \Sigma^{-1})}{\gamma}\left(\mathbf{m}_t - (V_t^{-1} + \kappa_l\Sigma^{-1})^{-1}(\kappa_l\Sigma^{-1}\mu + V_t^{-1}\mathbf{m}_t)\right) \\ &= \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1}\kappa_l\Sigma^{-1}}{\gamma}\right)\mathbf{m}_t + \frac{\delta t\kappa_t^{-1}\kappa_l\Sigma^{-1}}{\gamma}\mu.\end{aligned} \tag{2.22}$$

Therefore, we have that

$$\begin{aligned}\mathbf{m}_{t+1} - \mu &= \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1}\kappa_l\Sigma^{-1}}{\gamma}\right)(\mathbf{m}_t - \mu) \\ &= \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1}\kappa_l\Sigma^{-1}}{\gamma}\right)^{t+1}(\mathbf{m}_0 - \mu).\end{aligned} \tag{2.23}$$

By chossing the timestep $\delta t \in \left(0, \frac{\gamma}{\kappa_t^{-1}\kappa_l e(\Sigma)^{-1}}\right)$, where $e(\Sigma)$ denote the smallest eigenvalue of Σ, the first moment will converge to μ exponentially fast. For the second moment, we have that

$$V_{t+1} = \left(\mathbb{I}_M - \frac{\delta t \kappa_t^{-1}(V_t^{-1} + \kappa_l\Sigma^{-1})}{\gamma}\right)^2 V_t + \frac{2\delta t}{\gamma\kappa_h} \tag{2.24}$$

By assuming that δt is small, we have the following approximation

$$\begin{aligned}V_{t+1} &= \left(\mathbb{I}_M - \frac{2\delta t \kappa_t^{-1}(V_t^{-1} + \kappa_l\Sigma^{-1})}{\gamma}\right)V_t + \frac{2\delta t}{\gamma\kappa_h} \\ &= V_t - \frac{2\delta t\kappa_l\Sigma^{-1}V_t}{\gamma(\kappa_l + \kappa_h)} + \frac{2\delta t\kappa_l}{\gamma\kappa_h(\kappa_l+\kappa_h)} \\ &= V_t - \frac{2\delta t\kappa_l}{\gamma(\kappa_l+\kappa_h)}\left(\Sigma^{-1}V_t - \kappa_h^{-1}\mathbb{I}_M\right).\end{aligned} \tag{2.25}$$

Therefore, we have that

$$\begin{aligned}\Sigma^{-1}V_{t+1} - \kappa_h^{-1}\mathbb{I}_M &= \left(\mathbb{I}_M - \frac{2\delta t\kappa_l\Sigma^{-1}}{\gamma(\kappa_l+\kappa_h)}\right)\left(\Sigma^{-1}V_t - \kappa_h^{-1}\mathbb{I}_M\right) \\ &= \left(\mathbb{I}_M - \frac{2\delta t\kappa_l\Sigma^{-1}}{\gamma(\kappa_l+\kappa_h)}\right)^{t+1}\left(\Sigma^{-1}V_0 - \kappa_h^{-1}\mathbb{I}_M\right).\end{aligned} \tag{2.26}$$

By choosing the timestep $\delta t \in \left(0, \frac{\gamma}{2\kappa_t^{-1}\kappa_l e(\Sigma)^{-1}}\right)$, the second moment will converge to $\kappa_h^{-1}\Sigma$ exponentially fast. ∎

2.5. Practical stabilization modifications. In this study, for the sake of computational efficiency, we further simplify V by only considering the diagonal entries denoted as $\mathbf{v}$. Moreover, we utilize a moving average to ensure stable estimation

$$\begin{aligned}\mathbf{m}_{t+1} &= \beta_1\mathbf{m}_t + (1-\beta_1)\sum_{i=1}^{N_w}\mathbf{z}_t^i\hat{p}(\mathbf{z}_t^i), \\ \mathbf{v}_{t+1} &= \beta_2\mathbf{v}_t + (1-\beta_2)\kappa_t\sum_{i=1}^{N_w}(\mathbf{z}_t^i - \mathbf{m})\odot(\mathbf{z}_t^i - \mathbf{m})\hat{p}(\mathbf{z}_t^i),\end{aligned} \tag{2.27}$$

where β_1 and β_2 are hyper-parameters. To maintain unbiased estimation, normalization is implemented as follows: $\mathbf{m} = \frac{\mathbf{m}_t}{1-\beta_1^t}$ and $\mathbf{v} = \frac{\mathbf{v}_t}{1-\beta_2^t}$.

Furthermore, we note that the kinetic processes of a molecular system are generally characterized by the local minima and saddle points of the FES. On the other hand, the regimes of high free energy are less relevant. To accurately construct these thermodynamically accessible regimes, we modify the loss function as

$$\mathcal{L}_{\mathcal{N}}(\mathbf{z}) = \frac{\left|\nabla_{\mathbf{z}} A_{\mathcal{N}}(\mathbf{z}) + \mathrm{F}(\mathbf{z})\right|^2}{\left|\mathrm{F}(\mathbf{z})\right|^2 + e}, \tag{2.28}$$

for all the biomolecule systems except the toy example of the 1D Rastrigin function. e is a small value to regularize the denominator.

With the training samples obtained from the aforementioned maximization step, the NN surrogate is optimized using the Adam stochastic gradient descent method [25] for the minimization step. The loss function of the updated $A_{\mathcal{N}}(\mathbf{z})$, in turn, navigates the consensus-based adaptive sampling for the updated maximization step. The min-max problem is solved iteratively to achieve comprehensive sampling of the full phase space. A detailed algorithm is presented in Alg. 2.1.

Algorithm 2.1 Consensus-based adaptive sampling.

Require: Initial sampling point $\mathbf{z}_0^i$, for $i = 1, \ldots, N_w$
Require: Initial NN parameter θ_0
Require: The number of training iterations N_{train}
Require: The number of data collected N_{data} in each training iteration
  $j \leftarrow 0, t \leftarrow 0$
  $T \leftarrow \lceil \frac{N_{data}}{N_w} \rceil$
  while $j < N_{train}$ **do**
    while $t \leq T$ **do**
      calculate the mean force F_t^i at $\mathbf{z}_t^i$
      calculate the predicted force $\mathcal{F}_\theta(\mathbf{z}_t^i) = \nabla_{\mathbf{z}} A_{\mathcal{N}}(\mathbf{z}_t^i; \theta_j)$
      $L^i \leftarrow \mathcal{L}_{\mathcal{N}}(\mathbf{z}_t^i)$
      $w^i \leftarrow \frac{\exp(\kappa_l L^i)}{\sum_i \exp(\kappa_l L^i)}$
      $\mathbf{m}_{t+1} \leftarrow \beta_1 \mathbf{m}_t + (1-\beta_1) \sum_i \mathbf{z}_t^i w^i$
      $\mathbf{v}_{t+1} \leftarrow \beta_2 \mathbf{v}_t + (\kappa_l + \kappa_h)(1-\beta_2) \sum_i (\mathbf{z}_t^i - \mathbf{m}_t)^2 w^i$
      $\mathbf{m} \leftarrow \frac{\mathbf{m}_{t+1}}{1-\beta_1^t}$
      $\mathbf{v} \leftarrow \frac{\mathbf{v}_{t+1}}{1-\beta_2^t}$
      $\mathbf{z}_{t+1}^i \leftarrow \mathbf{z}_t^i - \frac{\delta t}{\gamma}(\mathbf{z}_t^i - \mathbf{m}) \oslash \mathbf{v} + \sqrt{\frac{2\delta t}{\gamma \kappa_h}} \eta_t^i, \eta_t^i \sim \mathcal{N}(0,1)$
      $t \leftarrow t + 1$
    end while
    Save the training dataset $\mathcal{D}_j = \{\mathbf{z}_t^i, \mathrm{F}_t^i\}_{t=0}^T$
    Optimize θ_{j+1} using the generated training set $\mathcal{D}_l$ for $l = 0, \ldots, j$.
    $j \leftarrow j + 1$
  end while

2.6. Related methods. Before we present the numerical results, we further discuss the differences between the present CAS method and the existing approaches. Since the standard sampling methods such as MCMC [37], Langevin dynamics [38], and generative model [12, 42] are not applicable, we focus on two major approaches

based on the enhanced sampling such as the VES method [4] as well as the active learning such as the RiD method [54].

The VES method provides an efficient approach to establish the re-weighted sampling of the phase space. However, the surrogate construction of the FES $A(\mathbf{z})$ further relies on the estimation of the PDF $\rho(\mathbf{z})$ from the obtained reweighted samples, which could become computationally expensive for high-dimensional problems. In contrast, the present method circumvents the PDF estimation and directly constructs $A(\mathbf{z})$ through the adaptive sampling of the mean force $F(\mathbf{z}) = -\nabla A(\mathbf{z})$.

Furthermore, the present method differs from the RiD method in the following two aspects: (I) The present method imposes the sampling adaptivity by explicitly using the residual error $\mathcal{L}_{\mathcal{N}}(\mathbf{z})$ while the RiD method does not directly account for the approximation error. Instead, the RiD method trains a replica of NNs on the same sample set and uses the standard deviation of multiple NN predictions as an indirect measure of the $\mathcal{L}_{\mathcal{N}}(\mathbf{z})$. As a result, the sampling error of the mean force $F(\mathbf{z})$ may lead to a consistent biased prediction among multiple NNs with a small standard deviation. (II) The RiD method drives the sampling dynamics with a biased MD potential $U(\mathbf{r})$ similar to Eq. (2.3). Accordingly, the time step needs to be sufficiently small due to the stiffness of $U(\mathbf{r})$. In contrast, the dynamics of the random walkers for the present method is governed by a mean-field quadratic potential $G(\mathbf{z})$ in Eq. (2.12). In particular, $G(\mathbf{z})$ is decoupled from the MD potential $U(\mathbf{r})$. This enables us to choose a much larger time step irrespective of the roughness of $U(\mathbf{r})$ and achieve a more efficient exploration of the phase space. Below we compare the performance of the present method with these two methods in more detail.

3. Numerical Results.

3.1. One-dimensional Rastrigin function. To illustrate the essential idea of the present method, we start with the one-dimensional Rastrigin function:

$$f(z) = z^2 - \cos(2\pi z), \quad z \in [-3, 3]. \tag{3.1}$$

Instead of a neural network, we simply use a piecewise polynomial function $f_\theta(z)$ for this 1D problem. Accordingly, the residual is directly defined as $|f - f_{\theta_i}|$, where θ_i represents the fitting parameters for the i-th iteration. Initially, we set $f_{\theta_0}(z) \equiv 8$ with consistent boundary condition. We use the proposed sampling method with 10 walkers to estimate the first and second moments, yielding $m_1 = 1 \times 10^{-4}$ and $V_1 = 0.022$ and therefore the first and second derivative $f'(m_1) = 0$ and $f''(m_1) = V_1^{-1} = 40.97$. The obtained m_1 is very close to the true values of the max-residual point $z_1^* = 0$ and the second derivative $f''(z_1^*) = 41.47$. Accordingly, we add new data points near $x = 1 \times 10^{-4}$ into the training set, which enables us to construct an updated approximation $f_{\theta_1}(z)$. Similarly, with the approximation $f_{\theta_i}(z)$, we conduct the sampling process (i.e., the maximization step) and include the obtained samples near m_{i+1} into the training set, yielding an updated approximation $f_{\theta_{i+1}}(z)$. As shown in Fig. 2, for each iteration, the sampling step can accurately pinpoint the max-residual region. Furthermore, as shown in Appendix C.1, the second moment V_i yields a consistent estimation of the second derivative near the max-residual region. The underlying function $f(z)$ can be accurately constructed after 12 iterations.

3.2. Two-dimensional FES. We use the alanine dipeptide (Ace-Ala-Nme), referred to as Ala2, as a benchmark problem. The molecule is solvated in 383 TIP3P water molecules similar to Ref. [54]. The full MD system is simulated in a canonical

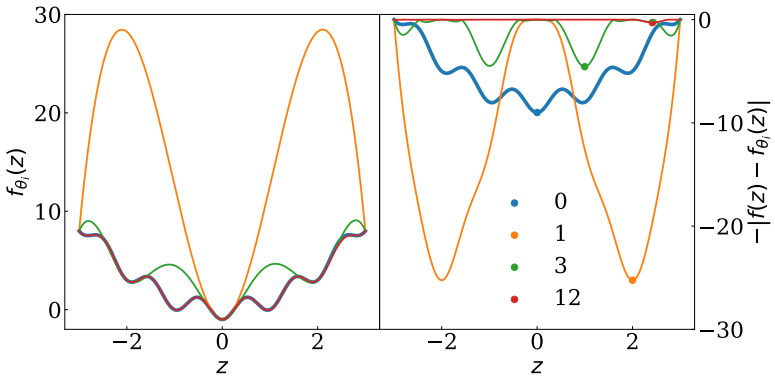

FIG. 2. *Adaptive sampling and construction of the 1D Rastrigin function. Left: The reference function $f(z)$ and the constructed approximations $f_{\theta_i}(z)$ obtained at different iteration steps. The relative l_2 error is less than 6×10^{-3} after 12 iterations. Right: The residual function $|f(z)-f_{\theta_i}(z)|$. The symbols represent the locations identified by each adaptive sampling (i.e., the maximization) step where new points will be added for the next training (i.e., the minimization) step.*

ensemble under temperature 300K using the Amber99-SB force field [22] with a time step 2 fs. We refer to Appendix A for the simulation details.

We choose the CVs as the two torsion angles ϕ (C, N, C_α, C) and ψ (N, C_α, C, N). For comparison, we also construct the FESs using the VES [4] and RiD [54] method with the same setup parameters presented therein. For the present CAS method, we use 10 walkers to explore the configuration space. The initial points of the walkers at the k-th iteration are chosen to be the final states of the previous iteration. In the sampling stage, the inverse of the low and high temperatures are set to be $\kappa_l = 10$ and $\kappa_h = 1$, respectively. For each sample point, restrained MD is conducted with $\kappa = 500$ for 5000 steps to compute the average force. The time step for the dynamics of the random walkers (2.12) is set to be 0.1. Fig. 3 shows the FESs constructed by the three different methods and the reference obtained by the metadynamics [27] using a long simulation time. As shown in Tab. 1, the present CAS method yields a smaller approximation error and meanwhile requires lower computational cost than the VES and the RiD method.

TABLE 1

The accuracy of the constructed 2D FES (the Ala2 molecule) and computational time (in hours, the same below) for the VES, RiD, and CAS methods. The l_2 and l_∞ error are computed up to 40 KJ/mol. The reference solution is constructed by the metadynamics. The simulation time of the CAS method is multiplied by 10 since 10 random walkers are used.

Method	Accuracy		Time	
	l_2 error	l_∞ error	Simulation	Train
VES	5.39	21.03	47.5	
RiD	3.15	11.04	17.98	0.22 (GPU)
CAS	1.88	10.68	0.23×10	0.18 (CPU)
				0.13 (GPU)

3.3. Three-dimensional FES. Next, we consider a s-(1)-phenylethyl (s1pe) peptoid in an aqueous environment similar to Refs. [50, 48]. The full system consists of one biomolecule and 546 water molecules in a $(2.9\text{nm})^3$ dodecahedron box. The

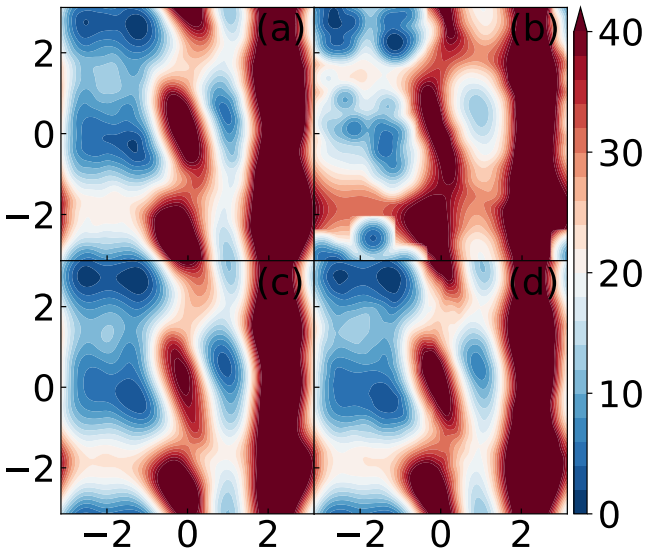

FIG. 3. *The 2D FES for the Ala2 molecule on the $\phi-\psi$ plane constructed by (a) Metadynamics (reference) (b) VES (c) RiD (d) CAS (the present method). The accuracy and the computational cost are shown in Table 1.*

CHARMM general force field (CGenFF) [50] is used for the biomolecule and the TIP3P model [24] is used for the water molecules; see Appendix A for details.

The CVs are the three torsion angles ω spanned by atoms (C_α, C, N, C_α), ϕ, and ψ, where the latter two are the same as the Ala2 molecule. The FES is constructed by both the RiD and CAS methods. The setup of the RiD method is the same as [48]. For the CAS method, we use 20 walkers and set the inverse of the low and high temperatures to be $\kappa_l = 10$ and $\kappa_h = 2$, respectively. The initial points of these walkers at each iteration are chosen in the same method as the two-dimensional problem. The timestep δt for the dynamics of the random walkers (2.12) is set to be 0.1. For each sample point, restrained MD with $\kappa = 500$ is conducted for 10000 steps to compute the mean force.

For visualization, we project the constructed FES onto a two-dimensional plane and fix the third variable. Fig. 4 shows the projected FES on the $\omega - \phi$ and $\phi - \psi$ plane obtained from the CAS and the RiD methods. For each projection, the reference is constructed as a 2D FES using the metadynamics [27]. Similar to the previous 2D case, the present CAS method yields higher accuracy with lower computational cost.

3.4. Nine-dimensional FES. Furthermore, we consider a more complex molecule, the peptoid trimer $(s1pe)_3$, solvated in a $(4.2\text{ nm})^3$ dodecahedron box with 1622 TIP3P water molecules. The chosen CVs consist of the 9 torsion angles ω, ϕ, ψ that are defined in s1pe case, associated with the different C_α atoms and denoted as $\omega_1, \phi_1, \psi_1, \omega_2, \phi_2, \psi_2, \omega_3, \phi_3, \psi_3$. We use 64 walkers for this case and the initial conditions of the walkers at each iteration are chosen with the same method as the previous cases. The inverse of the low and the high temperature are set to be $\kappa_l = 100$ and $\kappa_h = 2$, respectively. The timestep δt of the dynamics of the random walkers is set to

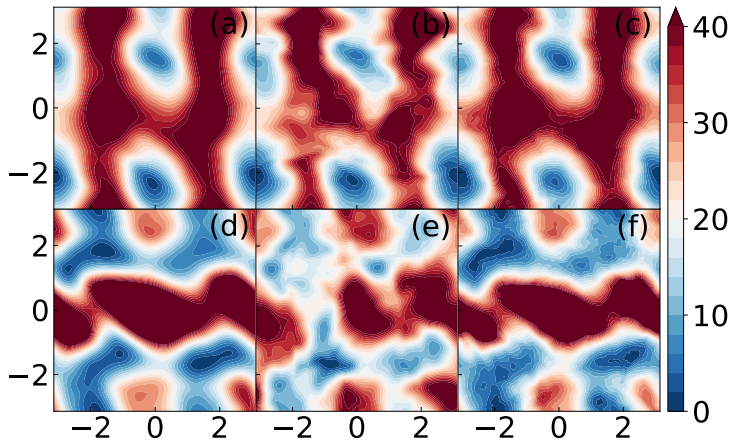

FIG. 4. *The 3D FES for molecule s1pe projected on the $\omega - \phi$ plane by fixing $\psi = 1.5$ (first row) and the $\phi - \psi$ planes by fixing $\omega = 1.5$ (second-row). (a-d) The 2D FES constructed using metadynamics with the third variable restrained (reference); (b-e) 2D projection of the 3D FES constructed by the RiD method; (c-f) 2D projection of the 3D FES constructed by the present CAS method.*

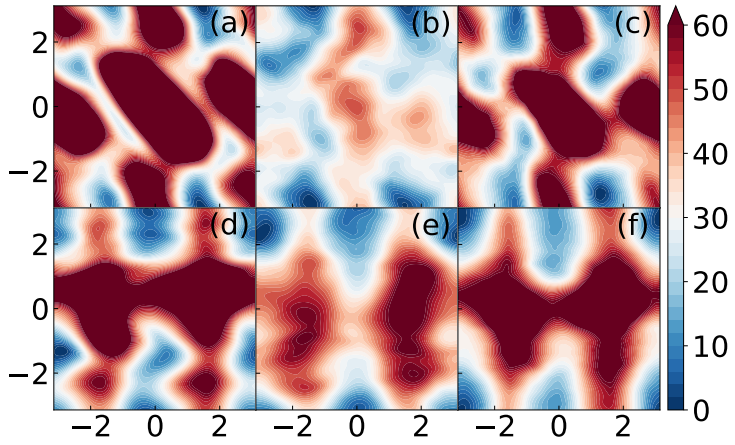

FIG. 5. *The 9D FES for molecule $(s1pe)_3$ projected on the $\omega_1 - \phi_1$ (first row) and $\omega_1 - \psi_1$ plane (second row). (a-d) The 2D FES constructed using metadynamics with the remaining variable restrained (reference solution); (b-e) 2D projection of the 9D FES constructed by the RiD method; (c-f) 2D projection of the 9D FES constructed by the present CAS method.*

be 0.1. The FES is constructed by the CAS method using 28 iterations of sampling and training, which requires $225.53 \times 64 = 14434.35$ CPU hours for simulation and 6.06 GPU hours for training. For comparison, the RiD method uses 17900 CPU hours for simulation and 15.44 GPU hours for training. Similar to the above 3D problem, the constructed 9D FES is projected onto various 2D planes with the remaining vari-

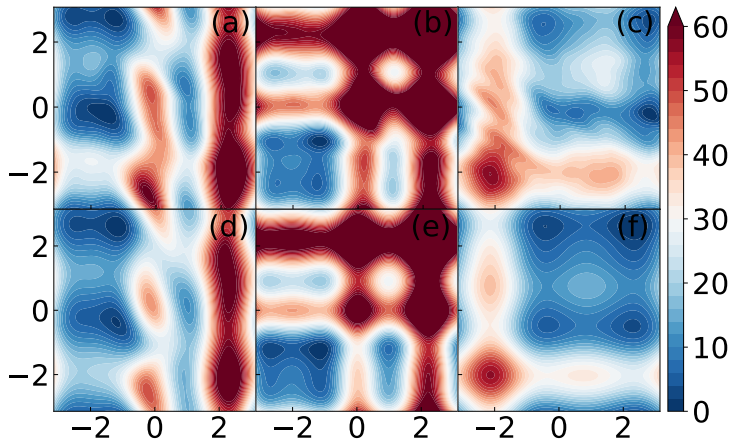

FIG. 6. *The 30D FES for molecule Ala16 projected on various 2D planes. (a-d) $\phi_1 - \psi_1$ (b-e) $\phi_2 - \phi_3$ (c-f) $\psi_2 - \psi_3$. (a-b-c) 2D FES constructed by metadynamics (reference); (d-e-f) 2D projection of the 30D FES constructed by the present CAS method.*

ables fixed. For each 2D projection, the reference is constructed as a 2D FES using metadynamics. Fig. 5 shows the projection on the $\omega_1 - \phi_1$ and $\omega_1 - \psi_1$ plane (see Appendix C.4 for other 2D projections). Compared with the Rid method, the present CAS method yields higher accuracy and lower computational cost.

3.5. Thirty-dimensional FES. Finally, we consider the polyalanine-15 (Ace-$(Ala)_{15}$-Nme), denoted as Ala16, solvated in 2258 water molecules. The chosen CVs consist of torsion angles ϕ and ψ, defined in Ala2 case, associated with the different C_αs, denoted as $\{\phi_i, \psi_i\}_{i=1}^{15}$. We use 64 walkers for this case and the initial value of the walkers at each iteration is given by the biased simulation before. The inverse of the low and high temperatures are set to be $\kappa_l = 20$ and $\kappa_h = 5$, respectively. The timestep α for dynamics of the random walkers Eq. (2.12) is set to be 0.1. The FES is constructed using 100 iterations of sampling and training. Similar to the previous case, the obtained FES is plotted on a two-dimensional plane while the remaining variables are fixed and the reference is constructed as a 2D FES using the metadynamics. Fig. 6 shows the projection on the $\phi_1 - \psi_1$, $\phi_2 - \phi_3$ and $\psi_2 - \psi_3$ plane. For all the cases, the projection of the 30-dimensional FES shows good agreement with the 2-dimensional reference solution. We have also examined the projection on other planes; the prediction shows good agreement with the reference solution as well.

4. Conclusions. We developed a consensus-based sampling approach for adaptive surrogate construction in systems with complex energy landscapes. The method reformulates the construction task as a minimax optimization problem by simultaneously optimizing the target function approximation and the training set. A unique feature of the method is its ability to unify the dynamical exploration of high-dimensional phase space in the presence of prevalent energy barriers with efficient sampling adaptive to the posterior residual error. As a motivating example, we considered the long-standing problem of constructing high-dimensional FESs for MD systems. Compared with existing approaches [41] focusing on various enhanced sampling strategies

to overcome energy barriers, the present method further enables the residual-based adaptive sampling that remains an open problem in FES construction. On the other hand, unlike the common adaptive sampling strategies based on the generative model [43] and important sampling [16] used for solving high-dimensional PDEs, the present method achieves efficient sampling of the max-residual region in the form of an interacting particle system that does not rely on the free query of the arbitrary point in the phase space and the gradient computation of the residual function. Given the fact the sampling only relies on the first and second-moment estimation, the method could be particularly efficient for high-dimensional problems. While the numerical results of biomolecular systems have demonstrated the effectiveness for the high-dimensional FES construction, the present method provides a framework for unifying the phase space exploration and residual-based adaptive sampling, and can be broadly applied to multiscale systems involving constrained sampling dynamics and surrogate modeling, such as rare event sampling [13], stochastic reduced dynamics [31, 17] and uncertainty quantification in high-dimensional constrained phase space [29].

Appendix A. Simulation setup. All the MD simulations are performed using the package GROMACS 2019.2 [30] and open-source, community-developed PLUMED library [45]. The simulation is carried out on Intel(R) Xeon(R) Platinum 8260 CPU.

A.1. ala2. The Ace-Ala-Nme (ala2) molecule is modeled by the Amber99SB force field [22]. The molecule is solvated in an aqueous environment with 383 TIP3P water molecules. Periodic boundary conditions are imposed along each direction. The cut-off radius of the van der Waals interaction is 0.9 nm. The long-range Coulomb interaction is treated with the smooth particle mesh Ewald method with a real space cutoff of 0.9 nm and a reciprocal space grid spacing of 0.12 nm. The system is integrated with the leap-frog scheme at time step 2 fs. The temperature is set to be 300 K by a velocity-rescale thermostat [5] with a relaxation time of 0.2 ps. The Parrinello-Rahman barostat [36] with a relaxation time scale of 1.5 ps and a compressibility of 4.5×10^{-5} bar^{-1} is coupled to the system to control the pressure to 1 bar. The hydrogen atom is constrained by the LINCS algorithm [21] and the H–O bond and H–O–H angle of water molecules are constrained by the SETTLE algorithm [33].

A.2. s1pe. The s-(1)-phenylethyl (s1pe) molecule is modeled by the CHARMM general force field (CGenFF) [50]. The molecule is dissolved in 546 TIP3P water molecules in a (2.69 nm)^3 dodecahedron box. The cut-off radius of the van der Waals interaction is 1 nm. Other setups are similar to the ala2 molecule.

A.3. (s1pe)$_3$. The (s1pe)$_3$ molecule is modeled by the CHARMM general force field (CGenFF) [50]. The molecule is solvated in a (4.2 nm)^3 dodecahedron box with 1622 TIP3P water molecules. Other setups are similar to the ala2 molecule.

A.4. ala16. The Ace-(Ala)$_{15}$-Nme (ala16) molecule is modeled by the Amber99SB force field [22]. The molecule is solvated in a (4.62 nm)^3 dodecahedron box with 2258 TIP3P water molecules. Other setups are similar to the ala2 molecule.

Appendix B. Training. The training data is collected by restrained molecular dynamics. The initial 10% steps of the restrained dynamics are used as equilibrium and the rest 90% steps are used to calculate the mean force. The FES are parameterized as a fully connected neural network. The depth and width of the NN are shown in Table 2. The NNs are trained by Adam for 100000 steps with a learning rate 1×10^{-3}. For each training step, 5000 sampling points are randomly selected from the data set. All the training processes are conducted using an Nvidia GPU V100 with 32GB memory.

TABLE 2
The depth and width of the NN used to parameterize the FES of different molecules.

molecule	depth	width
ala2	3	48
s1pe	4	64
(s1pe)$_3$	4	512
ala16	4	640

Appendix C. Additional numerical results.

C.1. 1D Rastrigin function. Tab. 3 presents the numerical estimations of the first and second moments along with the construction of the 1D Rastrigin function.

As shown in Prop. 2.3, the first moment m_i provides the prediction of the max-residual point z_i^* for the i-th iteration. The second moment V_i characterizes the second derivative of the residual $|f(z) - f_{\theta_i}(z)|$ at $z = z_i^*$. As shown in Tab. 3, the predictions of m_i and $f''_{\theta_{i-1}} \pm V_i^{-1}$ from the present method show good agreement with the max-residual point z_i^* and the second derivative $f''(z_i^*)$ for each iteration.

TABLE 3

The max-residual point z_i^* and its second derivative $f''(z_i^*)$, as well as their numerical estimations from the present sampling method at each iteration. Specifically, the max-residual point z_i^* can be estimated by m_i and the second derivative $f''(z_i^*)$ can be estimated by $f''_{\theta_i} \pm V^{-1}$.

i	0	1	2	3	4	5	6
z_i^*	0.000	2.000	-2.000	0.996	-0.996	2.805	-2.806
m_i	-0.000	2.000	-2.000	0.996	-0.995	2.805	-2.806
$f''(z_i^*)$	41.478	41.478	41.478	41.468	41.462	15.273	15.539
$f''_{\theta_i} \pm V^{-1}$	40.974	37.442	36.916	38.283	39.519	11.841	13.380
i	6	7	8	9	10	11	12
z_i^*	-2.806	1.500	-1.502	0.503	-0.501	2.416	-2.413
m_i	-2.806	1.499	-1.503	0.502	-0.501	2.414	-2.413
$f''(z_i^*)$	15.539	-37.478	-37.473	-37.474	-37.478	-31.901	-31.688
$f''_{\theta_i} \pm V_i^{-1}$	15.380	-36.796	-36.271	-37.515	-37.285	-31.521	-31.189

C.2. The 2D FES for molecule ala2. Tab. 4 shows the approximation error of the constructed FES surrogate at each iteration. The total computational cost is presented in Tab. 1 in the main manuscript.

C.3. The 3D FES for molecule s1pe. Fig. 7-9 shows the additional 2D projections of the 3D FES (molecule s1pe) constructed by the Rid and presented CAS method. For each case, the reference is constructed as a 2D FES using the metadynamics. The computational cost and accuracy of CAS and RiD are shown in Tab. 5. It shows that the CAS method achieves better accuracy with less computational cost.

C.4. The 9D FES for molecule $(\text{s1pe})_3$. The addition 2D projections of the 9D FES for molecule $(\text{s1pe})_3$ are presented in Fig. 10. Similar to the 3D case, the FES constructed by the present CAS method shows a better agreement with the reference constructed as a 2D FES using metadynamics.

C.5. The 30D FES for molecule ala16. The addition 2D projections of the 30D FES for molecule ala16 are presented in Fig. 11. We note that the projection on the $\phi_5 - \psi_5$ plane is significantly different from other projections such as the $\phi_1 - \psi_1$ plane in the main context. While the ala16 molecule consists of 15 sequential alanine residues, the FES for individual $\phi - \psi$ projections shows different features. The numerical results of the present CAS method show good agreement with the reference.

Appendix D. Code availablity. Python implementation of our codes are available at github https://github.com/Lyuliyao/consensus-sampling-method-for-expolering-high-dimensional-energy-surface.

Acknowledgments. The work is supported in part by the National Science Foundation under Grant DMS-2110981 and the ACCESS program through allocation MTH210005.

TABLE 4
The l_2 and l_∞ error of the FES surrogate by the present CAS method for each sampling iteration.

iteration	1	2	3	4	5	6	7
l_∞ error	35.21	30.62	22.38	16.41	14.83	9.86	10.68
l_2 error	9.82	7.52	6.84	3.77	2.36	3.37	1.88

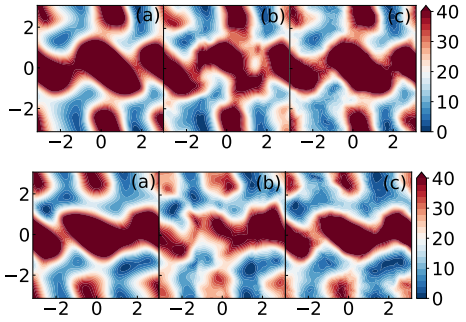

FIG. 7. *The 3D FES for the s1pe molecule projected on the $\phi - \psi$ plane with the third variable $\omega = -1$ (first row) and $\omega = -2$ (second row). (a) 2D FES by metadynamics (reference) (b) RiD (c) CAS.*

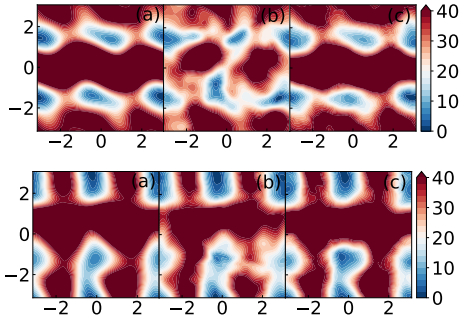

FIG. 8. *The 3D FES of the s1pe molecule on the $\omega - \psi$ plane with the third variable $\phi = 0$ (first row) and $\phi = -1$ (second row). (a) 2D FES by metadynamics (reference) (b) RiD (c) CAS.*

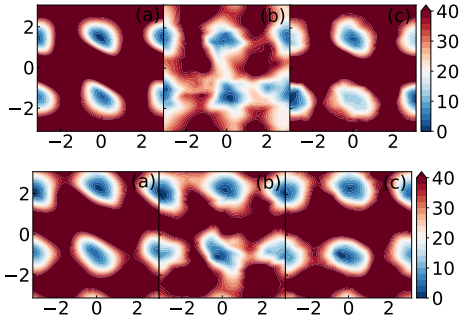

FIG. 9. *The 3D FES of the s1pe molecule on the $\omega - \phi$ plane with the third variable $\psi = 0$ (first row) and $\psi = -1$ (second row). (a) 2D FES by metadynamics (reference) (b) RiD (c) CAS.*

TABLE 5

The accuracy of the constructed 3D FES (the s1pe molecule) and computational time for RiD and the present CAS method. The l_2 and l_∞ error are calculated up to 40 KJ/mol. For each case, the FES is projected onto a 2D plane with the third variable fixed; the reference solution is constructed as a 2D FES by the metadynamics. The simulation time of the CAS method is multiplied by 20 since 20 walkers are used.

Method	Restraint	Accuracy		Time	
		l_2 error	l_∞ error	Sampling	Train
RiD	$\psi = 1.5$	5.76	25.72	423.33	8 (GPU)
	$\omega = 1.5$	12.04	49.13		
CAS	$\psi = 1.5$	2.44	11.21	4.81×20	0.84 (GPU)
	$\omega = 1.5$	3.89	28.80		

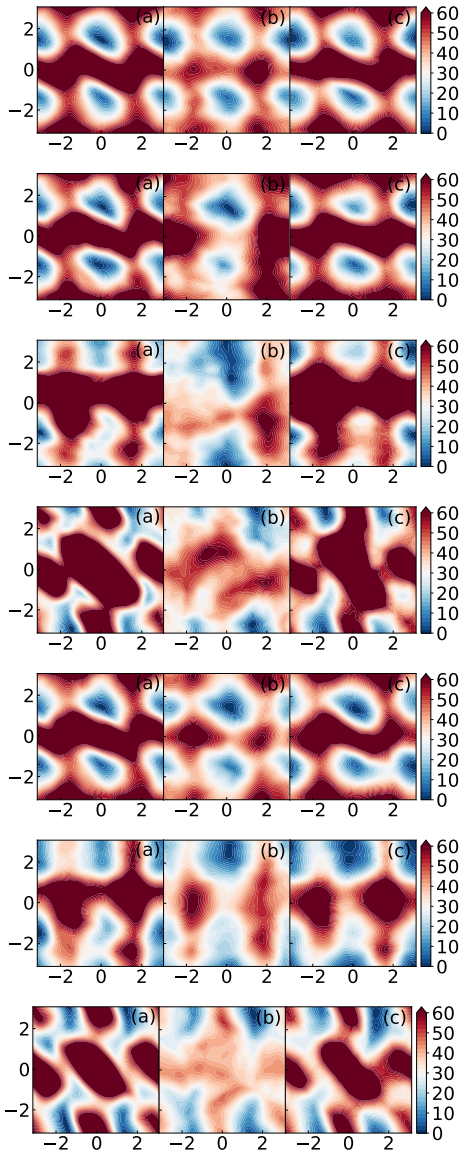

FIG. 10. *The 9D FES for the (s1pe)$_3$ molecule projected on the $\omega_1 - \phi_1, \omega_2 - \phi_2, \omega_2 - \psi_2, \phi_2 - \psi_2, \omega_3 - \phi_3, \omega_3 - \psi_3, \phi_3 - \psi_3$ plane from top to bottom. (a) 2D FES by metadynamics (reference) (b) RiD (c) CAS.*

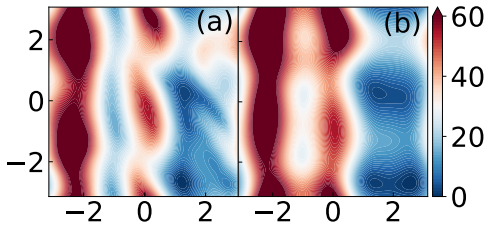

FIG. 11. *The 30D FES of the ala16 projected on the $\phi_5 - \psi_5$ plane. (a) 2D FES constructed by metadynamics (reference) (b) 30D FES constructed by the present CAS method.*

REFERENCES

[1] J. B. Abrams and M. E. Tuckerman, *Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations*, The Journal of Physical Chemistry B, 112 (2008), pp. 15742–15757.

[2] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, 2017.

[3] G. Bao, X. Ye, Y. Zang, and H. Zhou, *Numerical solution of inverse problems by weak adversarial networks*, Inverse Problems, 36 (2020), p. 115003.

[4] L. Bonati, Y.-Y. Zhang, and M. Parrinello, *Neural networks-based variationally enhanced sampling*, Proceedings of the National Academy of Sciences, 116 (2019), pp. 17641–17647.

[5] G. Bussi, D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*, The Journal of Chemical Physics, 126 (2007).

[6] J. A. Carrillo, Y.-P. Choi, C. Totzeck, and O. Tse, *An analytical framework for consensus-based global optimization method*, Mathematical Models and Methods in Applied Sciences, 28 (2018), pp. 1037–1066.

[7] J. A. Carrillo, F. Hoffmann, A. M. Stuart, and U. Vaes, *Consensus-based sampling*, Studies in Applied Mathematics, 148 (2022), pp. 1069–1140.

[8] J. Chen, S. Jin, and L. Lyu, *A consensus-based global optimization method with adaptive momentum estimation*, Communications in Computational Physics, 31 (2022), pp. 1296–1316.

[9] C. Chipot and T. Lelièvre, *Enhanced sampling of multidimensional free-energy landscapes using adaptive biasing forces*, SIAM Journal on Applied Mathematics, 71 (2011), pp. 1673–1695.

[10] E. Darve and A. Pohorille, *Calculating free energies using average force*, The Journal of Chemical Physics, 115 (2001), pp. 9169–9183.

[11] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, *Adaptive biasing force method for scalar and vector free energy calculations*, The Journal of Chemical Physics, 128 (2008).

[12] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using real nvp*, arXiv preprint arXiv:1605.08803, (2016).

[13] W. E and E. Vanden-Eijnden, *Transition-path theory and path-finding algorithms for the study of rare events*, Annual Review of Physical Chemistry, 61 (2010), pp. 391–420.

[14] W. Gao and C. Wang, *Active learning based sampling for high-dimensional nonlinear partial differential equations*, Journal of Computational Physics, 475 (2023), p. 111848.

[15] X. Gao, Z. Q. Xu, and X. Y. Zhou, *State-dependent temperature control for langevin diffusions*, SIAM Journal on Control and Optimization, 60 (2022), pp. 1250–1268.

[16] Z. Gao, L. Yan, and T. Zhou, *Failure-informed adaptive sampling for pinns*, SIAM Journal on Scientific Computing, 45 (2023), pp. A1971–A1994.

[17] P. Ge, Z. Zhang, and H. Lei, *Data-driven learning of the generalized langevin equation with state-dependent memory*, Phys. Rev. Lett., 133 (2024), p. 077301.

[18] T. Grafke and A. Laio, *Metadynamics for transition paths in irreversible dynamics*, Multiscale Modeling & Simulation, 22 (2024), pp. 125–141.

[19] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, *Improved training of wasserstein gans*, Advances in neural information processing systems, 30 (2017).

[20] J. Han, S. Gu, and X. Zhou, *StringNET: Neural network based variational method for transition pathways*, arXiv preprint arXiv:2408.12621, (2024).

[21] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, *Lincs: A linear constraint solver for molecular simulations*, Journal of Computational Chemistry, 18 (1997), pp. 1463–1472.

[22] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple amber force fields and development of improved protein backbone parameters*, Proteins: Structure, Function, and Bioinformatics, 65 (2006), pp. 712–725.

[23] Y. Jiao, D. Li, X. Lu, J. Z. Yang, and C. Yuan, *Gas: A gaussian mixture distribution-based adaptive sampling method for pinns*, arXiv preprint arXiv:2303.15849, (2023).

[24] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water*, The Journal of Chemical Physics, 79 (1983), pp. 926–935.

[25] D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, International Conference on Learning Representations (ICLR), (2015).

[26] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method*, Journal of Computational Chemistry, 13 (1992), pp. 1011–1021.

[27] A. Laio and M. Parrinello, *Escaping free-energy minima*, Proceedings of the National Acad-

emy of Sciences, 99 (2002), pp. 12562–12566, https://doi.org/10.1073/pnas.202427399.
[28] T. Lelièvre, M. Rousset, and G. Stoltz, *Long-time convergence of an adaptive biasing force method*, Nonlinearity, 21 (2008), p. 1155.
[29] T. Lelièvre, M. Rousset, and G. Stoltz, *Partial Differential Equations and Stochastic Methods in Molecular Dynamics*, vol. 1970 of Lecture Notes in Mathematics, Springer, 2016.
[30] Lindahl, Abraham, Hess, and van der Spoel, *Gromacs 2019.2 source code*, Apr. 2019.
[31] L. Lyu and H. Lei, *Construction of coarse-grained molecular dynamics with many-body non-markovian memory*, Phys. Rev. Lett., 131 (2023), p. 177301.
[32] L. Maragliano and E. Vanden-Eijnden, *A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations*, Chemical physics letters, 426 (2006), pp. 168–175.
[33] S. Miyamoto and P. A. Kollman, *Settle: An analytical version of the shake and rattle algorithm for rigid water models*, Journal of Computational Chemistry, 13 (1992), pp. 952–962.
[34] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, *Spectral normalization for generative adversarial networks*, arXiv preprint arXiv:1802.05957, (2018).
[35] W. G. Noid, *Perspective: Coarse-grained models for biomolecular systems*, The Journal of Chemical Physics, 139 (2013), p. 090901.
[36] M. Parrinello and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*, Journal of Applied Physics, 52 (1981), pp. 7182–7190.
[37] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2nd ed., 2004.
[38] G. O. Roberts and R. L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli, 2 (1996), pp. 341 – 363.
[39] L. Rosso, P. Minárý, Z. Zhu, and M. E. Tuckerman, *On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles*, The Journal of Chemical Physics, 116 (2002), pp. 4389–4402.
[40] P. Shaffer, O. Valsson, and M. Parrinello, *Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin*, Proceedings of the National Academy of Sciences, 113 (2016), pp. 1150–1155.
[41] G. Stoltz, M. Rousset, et al., *Free energy computations: A mathematical perspective*, World Scientific, 2010.
[42] K. Tang, X. Wan, and Q. Liao, *Adaptive deep density approximation for fokker-planck equations*, Journal of Computational Physics, 457 (2022), p. 111080.
[43] K. Tang, J. Zhai, X. Wan, and C. Yang, *Adversarial adaptive sampling: Unify pinn and optimal transport for the approximation of pdes*, arXiv preprint arXiv:2305.18702, (2023).
[44] G. M. Torrie and J. P. Valleau, *Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling*, Journal of Computational Physics, 23 (1977), pp. 187–199.
[45] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Plumed 2: New feathers for an old bird*, Computer physics communications, 185 (2014), pp. 604–613.
[46] O. Valsson and M. Parrinello, *Variational approach to enhanced sampling and free energy calculations*, Phys. Rev. Lett., 113 (2014), p. 090601.
[47] C. van der Oord, M. Sachs, D. P. Kovács, C. Ortner, and G. Csányi, *Hyperactive learning for data-driven interatomic potentials*, npj Computational Materials, 9 (2023), p. 168.
[48] H. Wang, T. Zariphopoulou, and X. Y. Zhou, *Reinforcement learning in continuous time and space: A stochastic control approach*, The Journal of Machine Learning Research, 21 (2020), pp. 8145–8178.
[49] E. Weinan, W. Ren, and E. Vanden-Eijnden, *String method for the study of rare events*, Physical Review B, 66 (2002), p. 052301.
[50] L. J. Weiser and E. E. Santiso, *A cgenff-based force field for simulations of peptoids with both cis and trans peptide bonds*, Journal of computational chemistry, 40 (2019), pp. 1946–1956.
[51] J. Yu, L. Lu, X. Meng, and G. E. Karniadakis, *Gradient-enhanced physics-informed neural networks for forward and inverse pde problems*, Computer Methods in Applied Mechanics and Engineering, 393 (2022), p. 114823.
[52] Y. Zang, G. Bao, X. Ye, and H. Zhou, *Weak adversarial networks for high-dimensional partial differential equations*, Journal of Computational Physics, 411 (2020), p. 109409.
[53] Q. Zeng, S. H. Bryngelson, and F. Schäfer, *Competitive physics informed networks*, arXiv preprint arXiv:2204.11144, (2022).
[54] L. Zhang, H. Wang, and W. E, *Reinforced dynamics for enhanced sampling in large atomic and molecular systems*, The Journal of Chemical Physics, 148 (2018), p. 124113.