SIAM J. SCI. COMPUT.
Vol. 40, No. 2, pp. A653–A670

AN AUTOMATIC ADAPTIVE IMPORTANCE SAMPLING ALGORITHM FOR MOLECULAR DYNAMICS IN REACTION COORDINATES*

J. QUER†, L. DONATI‡, B. G. KELLER‡, AND M. WEBER†

Abstract. In this article, we propose an adaptive importance sampling scheme for dynamical quantities of complex systems, which are metastable. The main idea of this article is to combine metadynamics, an algorithm from molecular dynamics simulation, with Girsanov's theorem, a result from stochastic analysis. With an assimilated version of the metadynamics algorithm we build a bias to reduce the metastability in the dynamical system. To correct the sampling of the modified system we apply a reweighting strategy based on Girsanov's theorem. The proposed algorithm has two advantages compared to a standard estimator of dynamic quantities: first, it is possible to produce estimators with a lower variance, and second, we speed up the generation of a typical sample.

Key words. adaptive importance sampling, molecular dynamics, metastability, variance reduction, nonequilibrium sampling, metadynamics, Girsanov

AMS subject classifications. 60G99, 68U20, 65C35, 65C60

DOI. 10.1137/17M1124772

1. Introduction and motivation.

In molecular dynamics (MD) simulation, dynamical quantities like exit times or transition probabilities play an important role. These quantities of interest can be used to characterize the stability of a certain conformation, which is of importance in, for instance, computational drug design [38]. Often these observables are estimated by averaging over long-term trajectories of some continuous space-time model, which describes the molecular movement, e.g., an overdamped Langevin equation. The sampling of dynamical quantities by MD simulations is often difficult since high barriers impact the system, and thus the sampling of ergodic trajectories is very time consuming. Thus, a molecular system is characterized by metastability, and the transitions between metastable states are rare events. Metastability arises from the multimodality of the stationary distribution of the dynamical system; cf. [23]. For this reason it is often modeled by a Markov jump process, which describes the jump from one metastable state to another; cf. [24, 31]. The estimation of dynamical quantities in such systems can be performed by ensemble averages. We will call these estimators Monte Carlo (MC) estimators. The MC approach for sampling these dynamical quantities is computationally costly and is affected by high variance of the estimator. This problem is known in the computational chemistry community as "the sampling problem."

In this article, we propose a new method to reduce the variance of dynamical

*Submitted to the journal's Methods and Algorithms for Scientific Computing section April 7, 2017; accepted for publication (in revised form) November 17, 2017; published electronically March 1, 2018.

http://www.siam.org/journals/sisc/40-2/M112477.html

Funding: This research has been funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems," A05 ("Probing Scales in Equilibrated Systems by Optimal Nonequilibrium Forcing"), and B05 ("Origin of the Scaling Cascades in Protein Dynamics").

†Zuse Institute Berlin, Berlin, 14195, Germany (quer@zib.de, weber@zib.de, www.zib.de/members/quer).

‡Freie Universität Berlin Department of Biology, Chemistry, Pharmacy, Berlin, 14195, Germany (ldonati@zedat.fu-berlin.de, bettina.keller@fu-berlin.de).

quantities estimated by MD simulations. To achieve this, we combine two strategies: In the first step, we use metadynamics to bias the initial potential energy function, and in the second step, we reweight the estimator of the quantity of interest using Girsanov's theorem.

Several methods have been suggested in the literature in order to solve the sampling problem. Many of these methods are based on ideas from importance sampling [29] or use other enhanced sampling approaches [3]. To sample stationary distributions or free energy profiles of dynamical systems that are characterized by metastability, different methods like metadynamics [21], umbrella sampling [35], adaptive biasing force [9], and Boltzmann reweighting [4] have been proposed. These methods are not designed for the sampling of dynamical quantities, and extending them for this purpose, therefore, is not straightforward.

The methods that have been designed for the sampling of dynamical quantities can be divided into two classes. The first class comprises splitting methods like forward flux sampling [2], adaptive multilevel splitting [6], or milestoning [15]. The second class comprises reweighting methods, sometimes called biased Brownian dynamics, like [10, 40]. Since our approach is a reweighting method we will not discuss splitting methods but refer the reader to [1] and the references therein for a detailed discussion.

Due to the interdisciplinary character of MD, several reweighting methods motivated by physical or mathematical arguments have been proposed. For instance, a physically motivated reweighting technique for the reconstruction of time scales has been proposed in [34]. The authors suggest a method to correct times scales from a metadynamics simulation. For the correction of the time scale, transition path theory is used and the derivation is based on a heuristic argument, which requires that the saddle point is not perturbed.

A mathematical approach to the variance reduction of MC methods was proposed by Milstein in [27]. The author proposes to use Girsanov's theorem and add an additional drift to the stochastic differential equation (SDE). The existence of an optimal drift is derived. This optimal bias satisfies a Bellman equation. The variance of this estimator is zero. Likewise, this work shows the difficulty of constructing a proper sampling scheme. Since the optimal bias is a solution to a nonlinear partial differential equation (PDE), the calculation is very costly or even impossible in high dimensions.

In general the reweighting strategy given by Girsanov's theorem always offers an unbiased estimator for any change of drift. Using this property one can build a suboptimal bias with lower computational cost to achieve the variance reduction.

Motivated by the connection between optimal control and importance sampling, different strategies for a general setting have been suggested by [11, 13, 14]. The authors derive importance sampling schemes for different assumptions based on the Hamilton–Jacobi–Bellman (HJB) equation (continuous version of Bellman's equation) in the large deviation context. These ideas have been developed for the application in MD as well; see [37]. In their work, the authors proposed a technique based on the solution of a deterministic control problem associated with the sampling problem.

Another strategy for situations occurring in MD simulations has been studied in [12]. The authors develop an importance sampling scheme for applications with so-called resting points. The numerical examples of the article show that the importance sampling scheme achieves better results when the resting point is taken into account. In order to build such importance sampling schemes, a lot of knowledge about the dynamical system is necessary, but finally this results in a better variance reduction. In [16], an optimization strategy for building the optimal bias was proposed. To derive

the optimization problem, the solution of the HJB equation is projected to a space of parametrized ansatz functions. Then a stochastic optimization problem needs to be solved to find the best approximation. The main challenge in this approach is to place the ansatz functions such that the algorithm converges quickly. In [33], the author develops a performance measure for an importance sampler related to small noise diffusion processes. This offers the possibility of comparing the different importance sampling schemes analytically.

In [40], an algorithm was proposed to overcome the problem of high dimensions. Motivated by the work of Milstein, the authors developed a variance reduction scheme restricted to reaction coordinates. They used an artificial drift to reduce the variance. The correction terms to reweight the results from the biased sampling, arising due to the artificial drift, are interpreted as weights. Hence, these weights are used to control the influence of different trajectories and are updated according to a stochastic rule, which is essentially the same as the Girsanov weight; cf. [28]. Based on these weights, the authors introduce a split-and-kill strategy of trajectories in order to have a proper sampling of the state space. The bias is placed in some reaction coordinates, which project the dynamics into some manifold of interest. In order to find the optimal bias in these reaction coordinates, they minimize a variational Smoluchowski equation. This expression is written in integral form and uses the stationary density distribution. Thus, the stationary distribution needs to be known a priori or has to be sampled by numerical simulation.

Our approach is distinguished from the above approaches in the following way. We do not try to find the optimal bias because it is computationally too expensive. Instead, we use the fact that the reweighted estimator is unbiased and construct a suboptimal bias, which will lead to a reasonable variance reduction. We use the a priori knowledge that the considered dynamic system is metastable, and this is why the metadynamics algorithm is used to construct the bias. For this, we need an additional sampling but do not have to know the stationary distribution in advance. Moreover, while the authors of [40] considered a very general approach of variance reduction for solving PDEs by path integrals of SDEs, we focus on the special case of metastable dynamical systems with a high energy barrier.

The article is structured as follows: First, we will give a short introduction into importance sampling and the main result from stochastic analysis based on Girsanov's theorem. Then we are going to review the metadynamics algorithm and show how we assimilate the algorithm for our needs. We will prove that Girsanov's theorem can be applied in the considered strategy. In the end we are going to apply our method to different numerical examples followed by a summary and an outlook.

2. Theory. In this section we present the two main ingredients of our algorithm. In the first part, we briefly review the main idea behind importance sampling, supplemented with some necessary tools from stochastic analysis. The second part is a short introduction into metadynamics and how the algorithm is assimilated.

2.1. Importance sampling. In this article we consider a diffusion process $X_t \in \mathbb{R}^n$ governed by the SDE

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t, \quad X_0 = x, \tag{1}$$

where X_t is the state of the system at time $t \geq 0$, $V : \mathbb{R}^n \to \mathbb{R}$ is a sufficiently smooth (e.g., $\mathcal{C}^\infty$) potential energy function, $\beta > 0$ is an arbitrary scaling factor for the noise, often called the inverse temperature, and B_t is a standard n-dimensional Brownian

motion with respect to the probability measure $\mathbb{P}$ on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$. Moreover we assume that the process is trapped in a metastable region $\mathcal{S} \subset \mathbb{R}^n$, which is an open and bounded set with a smooth boundary. Further, we define a target set $\mathcal{T}$ that is an open and bounded set with a smooth boundary as well. Finally, we define the stopping time $\tau = \inf\{t > 0 : X_t \in \mathcal{T}\}$ to be the first time that the process (1) hits the target set $\mathcal{T}$, e.g., when a dihedral angle of a biomolecule reaches a certain value. The presented theory here can be generalized to a state-dependent function prior to the Brownian motion; cf. [28]. But here we will consider the constant case only.

We are interested in expectations of the form

$$\mathbb{E}[e^{-\beta g(X_{0:T})}], \tag{2}$$

where $X_{0:T}$ is a trajectory of (1) until some finite time T and g is some functional on $\mathcal{C}([0,T] : \mathbb{R}^n)$. We consider these types of quantities because they give us information in terms of the temperature of the system. However, a generalization to other quantities expressed as expectations is possible. As pointed out by [37], an interesting case of this quantity arises when $g = 0$ for $X_{0:T} \in \mathcal{A} \subset \mathcal{C}([0,T], \mathbb{R}^n)$ and $g = \infty$ otherwise. Then (2) becomes

$$\mathbb{P}[X_{0:T} \in \mathcal{A}]. \tag{3}$$

Expectations like (2) are integrals over the entire state space and cannot be calculated analytically. But, given an ensemble of paths, they can be approximated by an unbiased MC estimator

$$I = \mathbb{E}[e^{-\beta g(X_{0:T})}] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} e^{-\beta g(X^i_{0:T})}, \tag{4}$$

where $X^i_{0:T}$, $i \in [1, \dots, N]$, are independent paths of length T, all starting at the same point $X_0 = x \in \mathbb{R}^n$, produced, for example, by numerical integration of (1). This estimator is unbiased. Its variance is given by

$$\mathrm{Var}(I) = \frac{1}{N}(\mathbb{E}[e^{-2\beta g(X_{0:T})}] - \mathbb{E}[e^{-\beta g(X_{0:T})}]^2). \tag{5}$$

The relative error is defined by

$$r(I) = \frac{\sqrt{(\mathrm{Var}(I))}}{\mathbb{E}[I]} = \frac{1}{\sqrt{N}} \sqrt{\frac{\mathbb{E}[e^{-2\beta g(X_{0:T})}]}{\mathbb{E}[e^{-\beta g(X_{0:T})}]^2} - 1}; \tag{6}$$

cf. [37].

To build an importance sampling scheme for a metastable diffusion process, one has to decrease the depth of the minima, which cause the metastable behavior. Since the time evolution of the SDE (1) with a low temperature (i.e., large β) is a negative gradient descent perturbed by some Brownian motion, the process X_t will stay in the region around the minimum of V. By filling the metastable region in $V(\cdot)$, we change the metastable behavior, and thus the sampling of the desired quantity of interest gets easier. But this perturbation changes the underlying stationary distribution as well. To compensate for this perturbation, we use Girsanov's theorem to reweight (or correct) the estimators. Another interpretation of this theorem is that it offers

a way to sample equilibrium quantities of some dynamics by sampling the dynamics out of equilibrium. We can construct a bias, which influences the mutimodality of the stationary distribution in such a way that low probability regions are more probable or high barriers are easier to cross. The main advantage of Girsanov's theorem is that it is not necessary to know the stationary distribution a priori.

Next we state the proposition that gives us the correction term for the quantity of interest sampled in a perturbed system. The proposition follows directly from Girsanov's theorem (cf. [28, p. 155]).

PROPOSITION 1. *Let $X_t \in \mathbb{R}^n$ and $Y_t \in \mathbb{R}^n$ be an Itô diffusion and an Itô process of the form*

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \leq T, \quad X_0 = x, \tag{7}$$

$$dY_t = (u(Y_t) + b(Y_t))dt + \sigma(Y_t)dB_t, \quad t \leq T, \quad Y_0 = x, \tag{8}$$

where $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy some Lipschitz condition such that we can guarantee uniqueness and existence of the solution and the time $T < \infty$. Furthermore, we define for an adapted measurable process $a : \mathbb{R}^n \to \mathbb{R}$ the stochastic process

$$M_t = \exp\left(-\int_0^t a(Y_s)dB_s - \frac{1}{2}\int_0^t a(Y_s)^2 ds\right) \tag{9}$$

for all $t \in [0, T]$ and $\sigma(Y_s)a(Y_s) = u(Y_s)$. Then, given that Novikov's condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |a(Y_t)|^2 dt\right)\right] < \infty \tag{10}$$

holds, for any function $f \in \mathcal{C}_0(\mathbb{R}^n)$ and any stopping time τ adapted to the filtration $\mathcal{F}_T$ (the filtration associated to the Brownian motion B in (7) and (8)) we have

$$\mathbb{E}^x_{\mathbb{P}}[f(X_{0:\tau})] = \mathbb{E}^x_{\mathbb{P}}[M_\tau f(Y_{0:\tau})]. \tag{11}$$

Proof. We are going to give a short sketch of the proof here for completeness.

We define a new probability measure

$$d\mathbb{Q} := M_T d\mathbb{P} \quad \text{on } \mathcal{F}_T.$$

Then

$$\hat{B}_t := \int_0^t a(Y_s)ds + B_t, \quad t \leq T,$$

is a Brownian motion with respect to $\mathbb{Q}$, and in terms of $\hat{B}_t$ the process Y_t can be represented by

$$dY_t = b(Y_t)dt + \sigma(Y_t)\hat{B}_t, \quad Y_0 = x, \ t \leq T.$$

Therefore, the $\mathbb{Q}$-law of Y_t with $Y_0 = x$ is the same as the $\mathbb{P}$-law of X_t with $X_0 = x$ for $t \leq T$. This follows directly from the weak uniqueness of solutions of SDEs (see [28, p. 71, Lemma 5.3.1]). Due to the absolute continuity of the two probability measures $\mathbb{Q}$ and $\mathbb{P}$, we can use a change of measure to rewrite the expectation. Thus, for any function $f \in \mathcal{C}_0(\mathbb{R}^n)$ and any stopping time $\tau \leq T$ which is adapted to the filtration $\mathcal{F}_T$ we can write

$$\mathbb{E}_{\mathbb{P}}[f(X_{0:\tau})] = \mathbb{E}_{\mathbb{Q}}[f(Y_{0:\tau})] = \mathbb{E}_{\mathbb{P}}[M_\tau f(Y_{0:\tau})],$$

which gives us the desired result. ∎

By setting $b(\cdot) = -\nabla V(\cdot)$ and $\sigma = \sqrt{2\beta^{-1}}$ we have the metastable SDE model (1) and can use Proposition 1 to reduce the metastability in the dynamical system.

Furthermore, it is possible to derive another reweighting formula (9) if $u(\cdot)$ is of gradient form ($u(\cdot) = \nabla v(\cdot)$). Then one can use Itô's formula and calculate another expression for the stochastic integral term in (9), as it done in [24, p. 838]. Applying Itô's formula to v, we get

$$\begin{aligned} v(Y_T) - v(Y_0) = &\int_0^T \frac{1}{\beta}\nabla^2 v(Y_s) - \nabla V(Y_s)\cdot\nabla v(Y_s) + |\nabla v(Y_s)|^2 ds \\ &+ \sqrt{2\beta^{-1}}\int_0^T \nabla v(Y_s) dB_s, \end{aligned} \tag{12}$$

where ∇^2 is the Laplacian. Now rearranging terms, we get a new expression for the stochastic integral, which can be used in (9) to derive

$$\begin{aligned} M_T = &\exp\left(\frac{1}{2\beta^{-1}}(v(Y_T) - v(Y_0))\right. \\ &\left. + \frac{1}{2\beta^{-1}}\int_0^T \left(\nabla V(Y_s)\cdot\nabla v(Y_s) + \frac{1}{2}|\nabla v(Y_s)|^2 - \beta^{-1}\nabla^2 v(Y_s) ds\right)\right). \end{aligned} \tag{13}$$

This expression is still stochastic because Y_s is a stochastic process. From the first point of view it seems that this term could be more easily treated numerically compared to the stochastic integral. We will investigate this in the example section.

Focusing on importance sampling, we can use Proposition 1 to rewrite the estimator (4) as

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} e^{-\beta g(Y^i_{0:\tau})} M^i_{0:\tau}, \tag{14}$$

where $Y^i_{0:\tau}$ and $M^i_{0:\tau}$ are independent samples from (8) and (9). For $a(Y_t)$ satisfying Novikov's condition and a bounded stopping time, M_τ is a continuous bounded local martingale which yields $\mathbb{E}[M_t] = 1,\ t \in [0, \tau]$. Then the importance sampling estimator is an unbiased estimator with expectation

$$\mathbb{E}[\hat{I}] = \mathbb{E}[e^{-\beta g(X_{0:\tau})}]; \tag{15}$$

cf. [24]. Following [37], we know that the relative error of this estimator is

$$r(\hat{I}) := \frac{1}{\sqrt{N}}\sqrt{\frac{\mathbb{E}[e^{-2\beta g(Y_{0:\tau})}(M_{0:\tau})^2]}{\mathbb{E}[e^{-\beta g(X_{0:\tau})}]^2} - 1}. \tag{16}$$

In order to control the relative error, we have to control the ratio

$$R(\hat{I}) := \sqrt{\frac{\mathbb{E}[e^{-2\beta g(Y_{0:\tau})}(M_{0:\tau})^2]}{\mathbb{E}[e^{-\beta g(X_{0:\tau})}]^2}}. \tag{17}$$

To apply Proposition 1 to dynamic quantities like exit times, we have to guarantee the fulfillment of Novikov's condition. This can be achieved by making different assumptions on the stopping time. We will present two different approaches known from the literature that state how these assumptions can be formulated.

Condition 1. Let $W = \mathcal{C}([0,\infty],\mathbb{R}^n)$ be the space of continuous paths of arbitrary length equipped with the Borel σ-algebra $\sigma(W)$. This σ-algebra is generated by all cylinder sets of the form $\{f \in W : f(t_1) \in E_1, f(t_2) \in E_2, \dots, f(t_k) \in E_k\}$, where $k \in \mathbb{N}$, $E_i \in \mathcal{B}(\mathbb{R}^n)$, and $0 \le t_1 \le t_2 \le \cdots \le t_k < \infty$. Further, let $\mathcal{F}_t = \sigma(\{w_s : s \le t\})$ denote the σ-algebra generated by the Brownian motion (w_s) up to time $t < \infty$. Then Girsanov's theorem holds on the measurable space $(W, \sigma(W))$ as long as the family $(M_t)_{t\ge 0}$ of random variables

$$M_T = \exp\left(\frac{-1}{\sqrt{2\beta^{-1}}}\int_0^T a(Y_s)dB_s - \frac{1}{2}\int_0^T \frac{1}{2\beta^{-1}}|a(Y_s)|^2 ds\right) \tag{18}$$

is a uniform integrable martingale. By Itô's formula, $(M_t)_{t\ge 0}$ is a nonnegative local martingale, which is uniformly integrable if Novikov's condition (10) holds [28]. This could be achieved by taking an admissible control which is defined up to a random stopping time τ, as done in [17].

Condition 2. Another way to guarantee the applicability of Proposition 1 is to assume that the stopping time is bounded for the specific problem. This assumption is by far nontrivial and can only be shown analytically in a very few situations. Anyhow, from a numerical viewpoint it is impossible to simulate trajectories which have infinite length. One has to stop the simulation after a finite number of steps. The quantity of interest can be approximated by the quantity of interest conditioned on the event happening in a finite simulation time. This assumption can be formalized by considering the stopping time $\hat{\tau} = \min(\tau, T_N)$, where T_N is the length of the numerical simulation. Then Novikov's condition follows for a reasonable function $u(\cdot)$. This treatment has been suggested in [24].

Condition 1 implies that one uses a bias function only up to a certain finite time. Thus, one can guarantee that the integral in the Girsanov weight (9) is bounded. The simulation could continue without the bias function acting on the trajectory. Condition 2 means that the sampling of the quantity of interest has to be finite in time. If the sampling is too long ($t > T_N$), the simulation is stopped.

In conclusion, Proposition 1 gives us an option to sample the dynamic quantity of interest from a different dynamical system without knowing the stationary distribution a priori. The different dynamical system can be changed in such a way that the quantity of interest is observed more often. The main difficulty of applying this strategy to a metastable system is to determine the metastable regions to change it accordingly. For this realization, we are going to use metadynamics. This algorithm is used in MD to sample the free energy surface and can be seen as an adaptive biasing method. In order to use this algorithm for our purposes, we are going to assimilate the algorithm slightly.

2.2. Metadynamics. The method metadynamics was first proposed in [19] under the name *local elevation*. It was reintroduced as metadynamics in [21]. It is an adaptive method for sampling the free energy surface (FES) of high-dimensional molecular systems. Thus, the main interest of this method is to find the stationary distribution. The method combines dynamics in reaction coordinates with adaptive bias potentials. The idea of this approach is to perturb the energy landscape when the simulation is trapped in a metastable region. This is done by locally adding Gaussian functions along a reaction coordinate, which fill up the minima in which the simulation is trapped. In this way, it is possible to explore the energy landscape in a rather short time compared to the plain sampling approach. The convergence of

some certain variants of metadynamics was proven in [8].

In order to apply the method, it is assumed that the high-dimensional system can be projected onto a few relevant collective coordinates. One possible way to find these collective variables for stochastic dynamics is to average out the fast degrees of freedom; see [22] or [39], for example. A more general overview can be found in [7] and the references therein. In general this projection can be written as $s : \mathbb{R}^n \to \mathbb{R}^d$ with $d \ll n$. Only the dependence of these parameters on the free energy $\mathcal{F}(s(x))$ is considered. The exploration of the FES is guided by the forces $F_i^t = -\partial \mathcal{F}(s_i(x))/\partial s_i^t$. The exploration of the FES gets stuck in a local minimum of the FES. In order to sample the FES more efficiently, a bias force is added to the system whenever the simulation is stuck in such a minimum. With metadynamics, one constructs a bias potential $V_{bias} : \mathbb{R}^d \to \mathbb{R}$, which is composed of $K \in \mathbb{N}$ Gaussian functions. The complete bias potential is

$$V_{bias}(x) = \sum_{i=1}^{K} \frac{w_i}{\sqrt{2\pi\lambda_i^2}} \exp\left(- \frac{(x - c_i)^2}{2\lambda_i^2} \right), \tag{19}$$

where $w_i \in \mathbb{R}$ is a weight, $c_i \in \mathbb{R}^d$ is the center of the Gaussian, and $\lambda \in \mathbb{R}$ is the width. These functions are placed along the trajectory to allow for an easy escape from this region, using the derivatives as an artificial force. The method can be parallelized easily since the bias force depends on the history of the individual trajectory only. This makes the method extremely efficient. Additionally, the bias also prevents the trajectory from going back to the visited states.

For simplicity we assume that all considered functions and variables are in the low-dimensional collective variable space and stick with the old notation. Of course, Girsanov's theorem is not restricted to the collective variable space.

2.3. Assimilation of metadynamics. We are going to assimilate the metadynamics algorithm to the sampling of dynamical quantities of interest. For our framework we do not have to calculate the complete FES. We only need a bias, which makes sure the trajectory does not get trapped in the metastable region. This is the reason why we add an additional sampling before we start sampling the quantity of interest to build a bias. In order to build a bias, which decreases the metastability, we use metadynamics in the metastable region only.

The bias is built in the following way: When the trajectory is trapped in a metastable region, we start a metadynamics simulation until the trajectory has left the metastable region. In every kth step, we add a Gaussian function to the potential such that the metastability is reduced. The force is then changed with the gradient of this Gaussian. When the trajectory hits the target set $\mathcal{T}$ for the first time, we save the bias and stop the metadynamics simulation. The bias consists of $\#steps\ needed/k$ bias functions. The choice of k is a compromise between adding as few bias functions as necessary, getting a small hitting time τ, and not perturbing the potential too much. Depending on the choice of the parameters w and λ, a certain number of bias functions is needed. It is obvious that the simulation of metadynamics gets more expensive the more bias functions are added due to the increasing number of function evaluations. That is why all parameters should be adapted to the problem such that the computation does not get too costly.

After having built the bias potential, the sampling of the original trajectory is continued with the bias potential. To correct the quantity of interest at the end of the calculation, we must sample the weights (9) as well. This can be done on the fly.

Remark 1. The construction of the bias potential depends on the history of the trajectory. Since the simulation to get the bias function is done in an additional step, the potential is not time dependent. Furthermore, the discretization of (1) always gives a discrete time Markov process because of the time independence of the Brownian motion. The construction of the bias potential itself is not Markovian because it depends on the history of the trajectory. Since the construction of the bias function and the sampling of the quantity of interest are done independently of each other, the bias does not have any influence on the Markovianity of the perturbed SDE (8). In general, an extension of the proposed method for non-Markovian dynamics should be possible. In this regard, one could use the metadynamics methods proposed in [5] and the general reweighting formula given in [28].

Remark 2. The method is not restricted to the use of metadynamics. Any stochastic approximation algorithm (e.g., adaptive biasing force) or even a deterministic algorithm could be used, provided the bias satisfies Novikov's condition. The method is also not restricted to the case in which the drift term ($b(\cdot)$) of the SDE is of gradient form. Since the Girsanov formula does not use the stationary distribution, all calculations are still valid. However, if the bias is not of gradient form, the alternative Girsanov formula cannot be applied.

3. The algorithm. Now we present the algorithm in pseudocode. We will use the metadynamics algorithm to build a bias in the metastable regions of the potential. Then we sample the quantity of interest in this biased potential N times and reweight the sampling with the weight given by (9).

Algorithm 1. Adaptive importance sampling.

dynamics X_t, Y_t, starting set $\mathcal{S}$, target set $\mathcal{T}$
initialization: $X_0 = Y_0 = x$; w_i, λ_i
Step 1: Build bias potential
while transition has not occurred **do**
  sample the dynamics X_t given in eq. (8);
  every kth steps: add a new bias function to $u(\cdot)$;
end while
save the bias potential;
Step 2: Sample the quantity of interest
for N **do**
  sample the quantity of interest with the additional bias according to eq. (8);
  sample the weights according to eq. (9);
end for
reweight according to eq. (11);
return estimators for quantities of interest

Proof of Novikov's condition. To apply Girsanov's theorem one has to make sure that Novikov's condition is satisfied.

LEMMA 1. *Let $\hat{\tau}$ be the stopping time as given in Condition* 2. *Further, let the bias potential consist of $K < \infty$ bias functions. We choose the weights of the bias function $V_{bias}(\cdot)$ such that the sum is bounded $\sum_{i=1}^{K} w_i < C_w$. If the perturbation potential is uniformly continuous, then the Novikov condition holds and we can use Proposition* 1 *to calculate path-dependent quantities from nonequilibrium sampling for the equilibrium dynamics.*

Proof. Since the bias function is added to the potential V, the resulting SDE is given by

$$dY_t = -\nabla V(Y_t) + \nabla V_{bias}(Y_t; c, w, \lambda)dt + \sqrt{2\beta^{-1}}dB_t, \quad Y_0 = x. \tag{20}$$

We have to verify Novikov's condition for $\nabla V_{bias}(\cdot)$, which states that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^{\hat{\tau}} \Big|\frac{\nabla V_{bias}(Y_t; c, w, \lambda)}{\sqrt{2\beta^{-1}}}\Big|^2 dt\right)\right] < \infty. \tag{21}$$

We are going to show that the time integral is bounded, from which we can then conclude that Novikov's condition holds:

$$\begin{aligned}\frac{1}{2}\int_0^{\hat{\tau}} \Big|\frac{\nabla V_{bias}(Y_t; c, w, \lambda)}{\sqrt{2\beta^{-1}}}\Big|^2 dt &= \frac{\beta}{4}\int_0^{\hat{\tau}} \Big\|\nabla V_{bias}(Y_t; c, w, \lambda)\Big\|_2^2 dt \\ &\leq \frac{\beta}{4}\int_0^{\hat{\tau}} \Big\|\nabla V_{bias}(Y_t; c, w, \lambda)\Big\|_\infty^2 dt \\ &\leq \frac{\beta}{4}\hat{\tau}\Big\|\nabla V_{bias}(Y_t; c, w, \lambda)\Big\|_\infty^2 < \infty.\end{aligned}$$

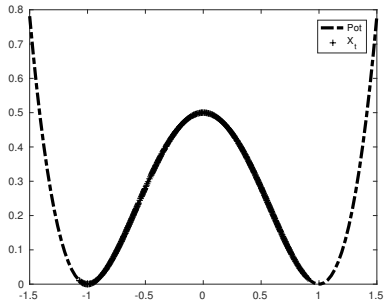
The $\|\cdot\|_\infty$ is bounded because the ansatz functions for the gradient are gradients of Gaussian functions, which are bounded. Furthermore, the stopping time is bounded due to Condition 2. It follows that the whole expression is bounded, and from this we conclude that (21) is satisfied. Therefore Novikov's condition holds. ∎

4. Examples. In the following, we study different numerical examples of the method presented above. We assume that the reaction coordinates are given such that we have a low-dimensional representation of the high-dimensional dynamics. The first 1D example shows the construction of the bias potential with fixed parameters. In the second 1D example we are going to use the alternative reweighting formula (13) to correct the statistics from the nonequilibrium sampling. In the last part of this section we will show a 2D example for alanine-dipeptide in reaction coordinates.

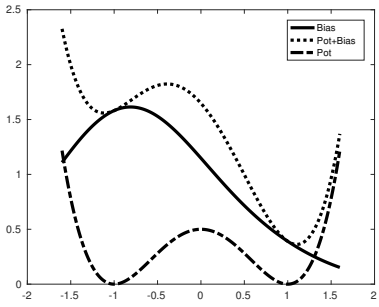
For the first examples we consider the dynamics given by (1) and the potential given by

$$V(x) = \frac{1}{2}(x^2 - 1)^2. \tag{22}$$

This potential has two minima at $x = \pm 1$ and a local maximum at $x = 0$. We are going to calculate two different quantities of interest. The first quantity of interest is the probability of all continuous paths which start at a point x in the metastable region $\mathcal{S}$ and reach the target set in time $\hat{\tau}$. This can be written mathematically as $\mathbb{P}(\mathcal{A})$, where $\mathcal{A} = \{X_{0:\hat{\tau}} \in \mathcal{C}([0, \hat{\tau}], \mathbb{R}^n) | X_0 = x(x \in \mathcal{S}), X_{\hat{\tau}} \in \mathcal{T}\}$. For this quantity of interest, we choose $g(Y_t) = 0$ for $Y_t \in \mathcal{S},\ t \in [0, \hat{\tau}]$, and choose $g(Y_t) = 1$ for $Y_t \in \mathcal{T},\ t \in [0, \hat{\tau}]$. The second quantity of interest is the moment generating function of the stopping $\hat{\tau}$. To sample this, we set $g(Y_{0:\hat{\tau}}) = \hat{\tau}$. The trajectories $Y_{0:\hat{\tau}}$ are realizations of (8) with $b(\cdot) = -\nabla V(\cdot)$, $u(\cdot)$ is the bias constructed by the metadynamics simulation, and $\sigma = \sqrt{2\beta^{-1}}$. We compare our method with the results of a standard MC estimator for the different quantities. We will see that in the first example, our method achieves the variance reduction for both reweighting formulas given in this article. Furthermore, the average transition time was decreased in the biased simulation. For this, we



(a) *The dashed-dotted line is the potential function* (22)*, and the black crosses show a realization of* (1) *performing the desired transition we want to sample.*



(b) *The dashed-dotted curve shows the original potential* (22)*. The solid black curve shows the bias V_{bias} produced by the metadynamics algorithm, and the dotted curve shows the corresponding biased potential* $(V + V_{bias})$.

FIG. 1. (a) *Trajectory of interest in the unperturbed potential.* (b) *Potential, bias, and perturbed potential.*

estimate the mean first hitting time (MFHT) for our experiments. The MFHT is the average time trajectories needed to reach the target set. If the trajectory did not hit the target set, the MFHT is set to T_N.

For the performed 1D experiments, we define the metastable region $\mathcal{S} = [-1.5, 0]$. We choose the starting point of the SDE (1) in the metastable region $X_0 = -1$ and fix $\beta = 3.0$ for all experiments. The stopping time is defined as the first hitting time of the target set $\mathcal{T} = [0.9, 1.1]$. In the first two examples we sampled 100,000 trajectories by using a standard Euler–Maruyama discretization with a time step $\Delta t = 10^{-4}$ in MATLAB; cf. [18]. Our aim was to investigate the variance within the different trajectories. At maximum, we calculated $T_N = 15{,}000$ time steps. The random number generator was fixed at `rng(1,'twister')` in order to have a better comparison within the different examples. Other random number generators have been tested showing similar results.

4.1. Diffusion in a double-well potential with fixed parameters. For this computation we chose $w_i = 0.05,\ \lambda_i = 0.8$ for all bias functions. The c_i of every bias function is chosen as the current value of the trajectory when the new bias function is added. The weights are calculated by (9).

In this example, 79 bias functions have been used; see Figure 1(b) for the calculated bias and the resulting potential energy. The estimators of the MC and the importance sampling are in good agreement for both cases; cf. Table 1. Our results show that the variance of the biased estimator is reduced for both quantities of interest using the reweighting approach. The variance for the transition probability is reduced by 65% and for the moment generating function by 76%. Hence, the automatically generated bias potential by the adjusted metadynamics algorithm is actually a good potential in the sense of importance sampling. Additionally, the MFHT is faster compared to the plain MC approach. This experiment shows that our method achieves the desired goals of variance reduction and computational speedup.

4.2. 1D with alternative reweighting formula. In this example, we use the alternative reweighting formula as shown in (13). In order to calculate the bias, we

TABLE 1

Comparison of the importance sampling estimators and the MC estimators for the simulation with fixed parameters of the biased potential.

	MC	GIR
$P(\mathcal{A})$	4.8470×10^{-2}	4.8323×10^{-2}
Var	4.6121×10^{-2}	1.6404×10^{-2}
$R(I)$	4.4307	2.6504
$\mathbb{E}[e^{-\beta\tau}]$	2.569×10^{-3}	2.4885×10^{-3}
Var	2.5850×10^{-4}	6.9180×10^{-5}
$R(I)$	6.2561	3.3463
MFHT	1.4804	1.4425

TABLE 2

Comparison of the importance sampling estimators and the MC estimators for the simulation with the alternative Girsanov formula.

	MC	AGIR
$P(\mathcal{A})$	4.8470×10^{-2}	4.8329×10^{-2}
Var	4.6121×10^{-2}	1.6407×10^{-2}
$R(I)$	4.4307	2.6504
$\mathbb{E}[e^{-\beta\tau}]$	2.5690×10^{-3}	2.4856×10^{-3}
Var	2.5850×10^{-4}	6.9170×10^{-5}
$R(I)$	6.2561	3.3459
MFHT	1.4804	1.4425

used the same parameters as in the first example. Since the random number generator was fixed, the bias is exactly the same as in the first example; see Figure 1(b).

In order to calculate the Girsanov weights, we need V_{bias}, ∇V_{bias}, and $\nabla^2 V_{bias}$. For V_{bias} as given in (19), the derivatives can be calculated easily.

In this case, the MC estimator and the importance sampling estimator agree very well; cf. Table 2. The variance reduction is very similar to the other reweighting formula. The variance for the transition probability is reduced by 64%, and the variance for the moment generating function is reduced by 73%. These experiments show that the alternative Girsanov formula can be applied as well to correct the biased estimators. The reduction of the MFHT is the same as in the first example since we used the same seed for the random number generator.

4.3. 2D diffusion process. In this example, we consider the dynamics of the molecule alanine-dipeptide (Ac-A-NHMe) in the limit of high friction. We reduced the problem to the motion of the backbone torsion angles ϕ and ψ, modeled by the overdamped Langevin equation:

$$\begin{aligned} \mathrm{d}\phi_t &= -\frac{\nabla V(\phi_t, \psi_t)}{\gamma}\mathrm{d}t + \frac{\sigma}{\gamma}\mathrm{d}B_t, \\ \mathrm{d}\psi_t &= -\frac{\nabla V(\phi_t, \psi_t)}{\gamma}\mathrm{d}t + \frac{\sigma}{\gamma}\mathrm{d}B_t, \end{aligned} \tag{23}$$

where $V(\phi, \psi)$ is the potential energy function of the system, γ is the friction, and B_t is a Brownian motion. The SDE in (23) has been solved using the Euler–Maruyama scheme with the Boltzmann constant $k_B = 0.00831451\,\mathrm{kJ/(mol \cdot K)}$, at temperature $T = 300\,\mathrm{K}$, with friction $\gamma = 1\,\mathrm{ps}^{-1}$ and $\Delta t = 0.002\,\mathrm{ps}$ as integrator time steps.

The constant σ represents the volatility of the Brownian motion and depends on the temperature of the system:

$$\sigma = \sqrt{2k_B T\gamma}. \tag{24}$$

The weights are calculated by (9). The reaction coordinate function $V(\phi, \psi)$ is not known and is therefore estimated from a full atomic MD simulation at high temperature as the FES of the relevant coordinates (ϕ, ψ). We performed an all-atom MD simulation of acetyl-alanine-methylamide (Ac-A-NHMe, alanine-dipeptide) in explicit water at 900 K, with the GROMACS 5.0.2 simulation package [36], the force field AMBER ff-99SB-ildn [25], and the TIP3P water model [20].

If we assume that the distribution of the relevant coordinates is proportional to the true stationary distribution given by the Boltzmann distribution $h(\phi, \psi) \approx \pi(\mathbf{x}) = \exp(-\beta V(\mathbf{x}))/Z$, with Z being the partition function, then the FES can be estimated from the histogram $h(\phi, \psi)$ as

$$V(\phi, \psi) = -\frac{1}{\beta} \log h(\phi, \psi) + \text{const.} \tag{25}$$

Since we need the analytical value of the potential gradient, we have estimated a bicubic spline of $V(\phi, \psi)$ using the ALGLIB [32] package.

In molecular dynamics simulation, the location of the minimum is usually not known. However, it is easy to bring the system into a configuration in a local minimum by optimization algorithms like gradient descent. Thus, instead of estimating the probability of jumping from one minimum to another, we have reformulated the problem and try to estimate the probability of leaving the starting minimum after a certain time. In order to approximate this probability we are going to calculate the probability of all continuous trajectories, which have reached a certain distance after some time. This can be formalized by calculating the probability of the set $\mathcal{A} = \{X_{0:\hat{\tau}} \in \mathcal{C}([0, \hat{\tau}] : \mathbb{R}^n) | X_0 = x(x \in \mathcal{S}), X_{\hat{\tau}} \in \mathcal{T}\}$, where $\mathcal{T} = \{z : |X_0 - z| = \hat{d}\}$.

We have chosen to start all simulations at point $\{\phi_0 = -1.5; \psi_0 = -0.2\}$, i.e., at the bottom of the right α-region of the Ramachandran space [30] and at distance $\hat{d} =$ 0.63, 0.83, 1.0. We have repeated our experiment 100 times, while each experiment is a collection of 10,000 trajectories and the maximum number of time steps for a single trajectory is 50. For each experiment, we estimated the bias by first running a metadynamics simulation and stopping when the trajectory reaches the desired distance $\hat{d}$. The bias is built adaptively with our method. Then we performed the simulations adding the bias to $V(\phi, \psi)$ in the metastable regions.

Figure 2 shows the free energy profile of the system estimated from the full atomic simulation (A), one example of metadynamics potential (B), and the biased potential (C). The effect of the metadynamics potential is to fill the local minimum and accelerate the exit of the trajectory from it.

Figure 3 shows the free energy profile from a different point of view together with two trajectories. While the trajectory spends a lot of time without any dominant direction in the unbiased potential (A), it is forced to leave the local minimum in a northerly direction in the second case (B).

In Table 3, we have collected the results for different experiments. The parameter Δk denotes the frequency at which the metadynamics potential was updated. Since in this kind of experiment the trajectories are quite short, we added a new Gaussian function to the bias potential at every time step, with a new center c_i equal to the last state visited.

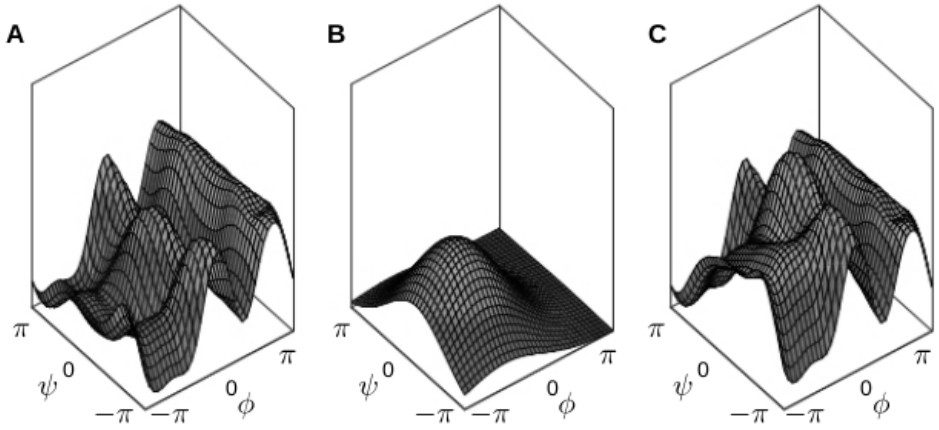


FIG. 2. *Free energy profile* (A), *metadynamics potential* (B), *and free energy profile and metadynamics potential* (C). *To estimate the metadynamics potential, we have used* $w_i = 0.9$, $\lambda_i = 1.6$, *and* $\Delta k = 1$.

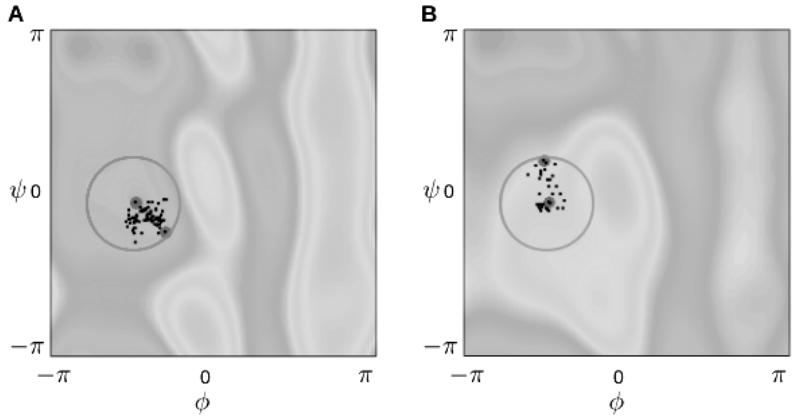


FIG. 3. *FES of alanine-dipeptide* (A), *and biased FES* (B). *The circle* ($\hat{d} = 0.83$) *represents the target that the trajectory has to hit.*

TABLE 3
Table for the numerical examples with alanine-dipeptide.

	MC1	GIR1	MC2	GIR2	GIR3	GIR4	MC5	GIR5
$\hat{d}$	0.63	0.63	0.83	0.83	0.83	0.83	1.0	1.0
w_i	-	0.9	-	0.15	0.9	0.9	-	0.9
λ_i	-	1.6	-	0.5	0.5	1.6	-	2.0
Δk	-	1	-	1	1	1	-	1
$P(\mathcal{A})$	6.46×10^{-1}	6.46×10^{-1}	3.04×10^{-1}	3.04×10^{-1}	2.96×10^{-1}	3.04×10^{-1}	1.53×10^{-1}	1.54×10^{-1}
Var	2.46×10^{-5}	1.97×10^{-5}	2.09×10^{-5}	1.11×10^{-5}	1.36×10^{-3}	1.14×10^{-5}	1.34×10^{-5}	5.21×10^{-6}
$\mathbb{E}[e^{-\beta\tau}]$	6.33×10^{-1}	6.33×10^{-1}	2.97×10^{-1}	2.97×10^{-1}	2.89×10^{-1}	2.97×10^{-1}	1.49×10^{-1}	1.50×10^{-1}
Var	2.37×10^{-5}	1.87×10^{-5}	1.98×10^{-5}	1.06×10^{-5}	1.28×10^{-3}	1.39×10^{-5}	1.27×10^{-5}	4.91×10^{-6}
MFHT	25.42	21.26	30.21	26.47	20.49	25.05	32.82	27.73

In the first two columns (MC1, GIR1), we have considered a short distance $\hat{d} = 0.63$, and the results show that the reweighted estimator allows us to predict the MC estimators with a variance reduction approximatively of 20% and a reduction of the MFHT of 17%.

The next four columns (MC2, GIR2, GIR3, GIR4) report the data for the experiments with $\hat{d} = 0.83$. In the first case (GIR2), we have used $w_i = 0.15$ and $\lambda_i = 0.5$, obtaining a variance reduction of almost 50% and MFHT reduced by 13 %. In the next case (GIR3), we tried to increase the height of the Gaussians to $w_i = 0.9$ with a small variance $\lambda_i = 0.5$. The dynamics is strongly accelerated; thus the trajectory leaves the local minimum quickly, but the precision of the estimators is reduced compared to results obtained by MC estimators. In particular, we observe an increase in the variance of two orders of magnitude. By increasing the variance of the Gaussian functions to $\lambda_i = 1.6$, keeping $w = 0.9$ (GIR4), the results of the reweighted estimators agree with the results obtained by MC simulations again.

The last two columns show the results for $\hat{d} = 1$. In this case, we have increased the variance to $\lambda_i = 2$. We observe a variance reduction of almost 60%. However, in this case the dynamics is slower and the MFHT is reduced only by 14%.

To study the dependence of the method on the metadynamics parameters, we have repeated the last experiment with $\hat{d} = 1.0$ for six values of w_i (the height of the Gaussians) and six values of λ_i (the variance of the Gaussians).

Figure 4 shows the error of the reweighted estimators with respect to the estimators obtained by MC simulations, while Figure 5 shows the variance reduction or increase. The two figures give a heuristic argument for the presented choice of parameters. The success of the method depends on the choice of the initial parameters. If the parameter λ of the Gaussian functions is bigger than the height w, we obtain the best results, i.e., a better value of the estimators and a significant variance reduction. This fact has a heuristic explanation. Thus, if the parameter λ is too small compared to the height w, the bias results in high narrow peaks. Adding this bias leads to a much more rugged potential. On the contrary, a large variance guarantees that the minimum is filled properly.

5. Summary and outlook. In this article, we developed an algorithm for automatic assimilated importance sampling for dynamical quantities in metastable systems. For this, we considered metastable stochastic systems. The main idea is to bias the dynamics in the metastable region. This can be achieved by constructing bias functions raising the region around the minimum of the drift term. To build this bias, we use the metadynamics algorithm, which was adjusted to create a bias decreasing the metastability only. In order to correct the sampling in the perturbed system, we proposed a reweighting scheme based on Girsanov's theorem. We proved that the reweighting scheme can be applied in our situation. We also considered an alternative reweighting formula which can be derived by using Itô's lemma. We tested our approach in different numerical examples. In a 1D example, we showed that this algorithm generates estimators for different quantities of interest with a reduced variance compared to an MC estimator. The bias also reduces the MFHT of the rare events, which results in shorter simulation time. We also tested the alternative reweighting formula. The results of this approach are as good as the results of the original formula. Whether one of the formulas has certain advantages compared to the other is the subject to ongoing research. Recent work shows a deeper mathematical problem with the alternative formula in the long-term limit; cf. [26]. In a 2D example, we also showed that our method works for higher-dimensional reaction coordinates. Further-

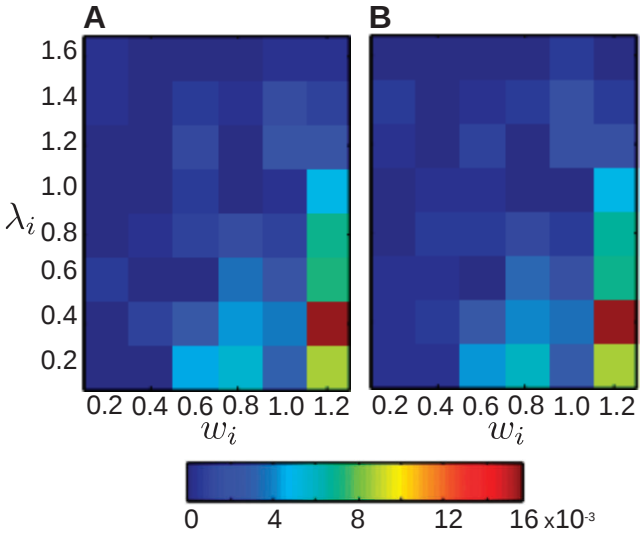


FIG. 4. *Error of* $P(\mathcal{A})$ (A) *and* $\mathbb{E}[e^{-\beta\tau}]$ (B) *with respect to the same quantities estimated by MC simulation as a function of the parameters* w_i *and* λ_i.

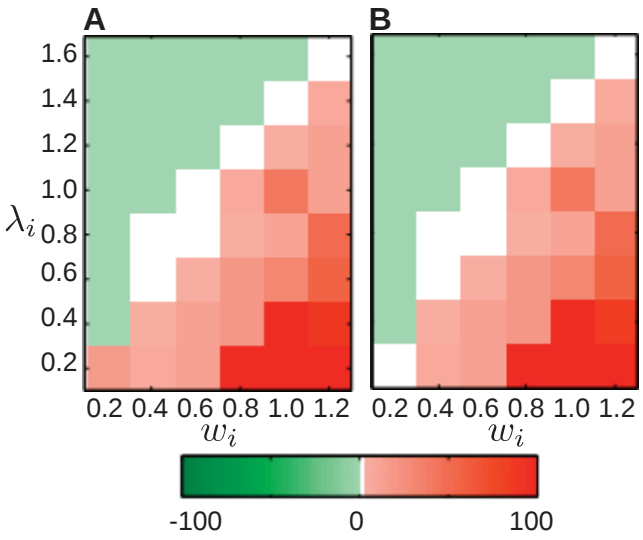


FIG. 5. *Ratio of the variance of* $P(\mathcal{A})$ (A) *and* $\mathbb{E}[e^{-\beta\tau}]$ (B) *with respect to variance of the same quantities estimated by MC simulation as a function of the parameters* w_i *and* λ_i. *The green color denotes a variance reduction.*

more, the 2D example gave us a heuristic understanding of how to choose the bias parameters in order to achieve appropriate variance reduction. The first impression is that a bigger bias potential is always better. But this is in general not the case for two reasons: First, the numerical treatment of Girsanov's formula is quite delicate, and a higher bias makes this even worse. Second, the analysis of [33] showed that

there is an exponential constant in the decay of the variance involved. A higher bias function will also have an impact on this constant, leading to a slower decay of the variance.

In the next step we aim at analyzing our importance sampling scheme in more detail. One possible direction for this is the analysis of our method in terms of subsolutions of the HJB equation as proposed in [33]. Another direction could be to use log-Sobolev inequalities as proposed in [23]. We are especially interested in discovering under which conditions our proposed method produces subsolutions of the corresponding HJB equation.

Directly perturbing the gradient is another methodological path we are currently following. An approach like this could use adaptive biasing force methods for building the bias instead of metadynamics. First simulations have shown that this leads to an even shorter MFHT because the perturbed gradient of the biased dynamical system is smooth compared to the bias generated by the metadynamics approach. How the variance of the estimator is influenced by this is the subject of ongoing research.

We suggest that our method proposed here serves the generation of an initial guess at solving the optimization problem in [16]. The combination of these two methods could lead to an optimized artificial force, which yields the potential to achieve an even better variance reduction.

The split-and-kill strategy introduced by Zou and Skeel in [40] is a very appealing idea to control the growth of the reweighting factor, which frequently causes numerical issues. We are going to consider this strategy in the future and hope to introduce more stability in the calculations of the reweighting factor.

Acknowledgment. We would like to thank the anonymous reviewers for their very helpful advice and their careful reading.

REFERENCES

[1] R. J. Allen, C. Valeriani, and P. R. ten Wolde, *Forward flux sampling for rare event simulations*, J. Phys. Condensed Matter, 21 (2009), 463102.

[2] R. J. Allen, P. B. Warren, and P. R. ten Wolde, *Sampling rare switching events in biochemical networks*, Phys. Rev. Lett., 94 (2005), 018104.

[3] R. Bernardi, M. Melo, and K. Schulten, *Enhanced sampling techniques in molecular dynamics simulations of biological systems*, Biochim. Biophy. Acta (BBA), 1850 (2015), pp. 872–877.

[4] M. Bonomi, A. Barducci, and M. Parrinello, *Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics*, J. Comput. Chem., 30 (2009), pp. 1615–1621.

[5] G. Bussi, A. Laio, and M. Parrinello, *Equilibrium free energies from nonequilibrium metadynamics*, Phys. Rev. Lett., 96 (2006), 090601.

[6] F. Cérou and A. Guyader, *Adaptive multilevel splitting for rare event analysis*, Stoch. Anal. Appl., 25 (2007), pp. 417–443.

[7] C. Chipot and A. Pohorille, *Free Energy Calculations*, Springer Series in Chemical Physics, Springer, Berlin, Heidelberg, 2007.

[8] J. F. Dama, M. Parrinello, and G. A. Voth, *Well-tempered metadynamics converges asymptotically*, Phys. Rev. Lett., 112 (2014), 240602.

[9] E. Darve and A. Porohille, *Calculating free energy using average forces*, J. Chem. Phys., 115 (2001), pp. 9169–9183.

[10] L. Donati, C. Hartmann, and B. Keller, *Girsanov reweighting for path ensembles and Markov state models*, J. Chem. Phys., 146 (2017).

[11] P. Dupuis, A. D. Sezer, and H. Wang, *Dynamic importance sampling for queueing networks*, Ann. Appl. Probab., 17 (2007), pp. 1306–1346.

[12] P. Dupuis, K. Spiliopoulos, and X. Zhou, *Escaping from an attractor: Importance sampling and rest points* I, Ann. Appl. Probab., 25 (2015), pp. 2909–2958.

[13] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stoch. Stoch. Rep., 76 (2004), pp. 481–508.
[14] P. Dupuis and H. Wang, *Subsolutions of an Isaacs equation and efficient schemes for importance sampling*, Math. Oper. Res., 32 (2007), pp. 723–757.
[15] A. K. Faradjian and R. Elber, *Computing time scales from reaction coordinates by milestoning*, J. Chem. Phys., 120 (2004), 10880.
[16] C. Hartmann and C. Schütte, *Efficient rare event simulation by optimal nonequilibrium forcing*, J. Statist. Mech. Theory Exper., 2012 (2012), P11004.
[17] C. Hartmann, C. Schütte, and W. Zhang, *Model reduction algorithms for optimal control and importance sampling of diffusions*, Nonlinearity, 29 (2016), pp. 2298–2326.
[18] D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, SIAM Rev., 43 (2001), pp. 525–546, https://doi.org/10.1137/S0036144500378302.
[19] T. Huber, A. E. Torda, and W. F. van Gunsteren, *Local elevation: A method for improving the searching properties of molecular dynamics simulation*, J. Computer-Aided Molecular Design, 8 (1994), pp. 695–708.
[20] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. Klein, *Comparison of simple potential functions for simulating liquid water*, J. Chem. Phys., 79 (1983), pp. 926–935.
[21] A. Laio and M. Parrinello, *Escaping free-energy minima*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 12562–12566.
[22] F. Legoll, T. Leli'evre, and S. Olla, *Pathwise estimates for an effective dynamics*, Stochastic Process. Appl., 127 (2017), pp. 2841–2863.
[23] T. Leliévre, *Two mathematical tools to analyze metastable stochastic processes*, in Numerical Mathematics and Advanced Applications 2011, Springer, Heidelberg, 2013, pp. 791–810.
[24] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numer., 25 (2016), pp. 681–880.
[25] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Improved side-chain torsion potentials for the amber ff99sb protein force field*, Proteins, 78 (2010), pp. 1950–1958.
[26] P. Malsom and F. Pinski, *Pinned Brownian Bridges in the Continuous-Time Limit*, preprint, https://arxiv.org/abs/1704.01991, 2017.
[27] G. N. Milstein, *Numerical Integration of Stochastic Differential Equations*, Math. Appl. 313, Springer Science & Business Media, 1994.
[28] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Hochschultext/Universitext, Springer, Berlin, 2003.
[29] J. Quer and H. Lie, *Some Connections between Importance Sampling and Enhanced Sampling Methods in Molecular Dynamics*, Tech. report, Zuse Institute Berlin, 2017.
[30] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations*, J. Mol. Biol., 5 (1963), pp. 95–99.
[31] C. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics*, AMS, 2013.
[32] B. Sergey, *ALGLIB*, 1999, www.alglib.net.
[33] K. Spiliopoulos, *Nonasymptotic performance analysis of importance sampling schemes for small noise diffusions*, J. Appl. Probab., 52 (2015), pp. 797–810.
[34] P. Tiwary and M. Parrinello, *From metadynamics to dynamics*, Phys. Rev. Lett., 111 (2013), 230602.
[35] G. Torrie and J. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys., 23 (1977), pp. 187–199.
[36] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and J. Berendsen, H., *GROMACS: Fast, flexible, and free*, J. Comput. Chem., 26 (2005), pp. 1701–1718.
[37] E. Vanden-Eijden and J. Weare, *Rare event simulation of small noise diffusions*, Comm. Pure Appl. Math., 65 (2012), pp. 1770–1803.
[38] M. Weber, C. Zoschke, A. Sedighi, E. Fleige, R. Haag, and M. Schäfer-Korting, *Free energy simulations of cargo-carrier interactions for core-multishell nanotransporters*, J. Nanomedicine Nanotechnology, 5 (2014), 234.
[39] W. Zhang, C. Hartmann, and C. Schütte, *Effective dynamics along given reaction coordinates, and reaction rate theory*, Faraday Discuss., 195 (2016), pp. 365–394.
[40] G. Zou and R. D. Skeel, *Robust variance reduction for random walk methods*, SIAM J. Sci. Comput., 25 (2004), pp. 1964–1981, https://doi.org/10.1137/S1064827503424025.