Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

JANNES QUER[1] AND LUCA DONATI[2] AND BETTINA G. KELLER [2] AND MARCUS WEBER [1]

An automatic adaptive importance sampling algorithm for molecular dynamics in reaction coordinates

[1]Zuse Institute Berlin
[2]Freie Universität Berlin, Institute of Chemistry and Biochemistry, Takustr. 3, D-14195 Berlin

ZIB Report 17-09 (2017)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

An automatic adaptive importance sampling algorithm for molecular dynamics in reaction coordinates

Jannes Quer [*1], Luca Donati [†2], Bettina G. Keller [‡2] and Marcus Weber[§1]

[1]Zuse Institute Berlin
[2]Freie Universität Berlin, Institute of Chemistry and Biochemistry, Takustr. 3, D-14195 Berlin

March 1, 2017

Abstract

In this article we propose an adaptive importance sampling scheme for dynamical quantities of high dimensional complex systems which are metastable. The main idea of this article is to combine a method coming from Molecular Dynamics Simulation, Metadynamics, with a theorem from stochastic analysis, Girsanov's theorem. The proposed algorithm has two advantages compared to a standard estimator of dynamic quantities: firstly, it is possible to produce estimators with a lower variance and, secondly, we can speed up the sampling. One of the main problems for building importance sampling schemes for metastable systems is to find the metastable region in order to manipulate the potential accordingly. Our method circumvents this problem by using an assimilated version of the Metadynamics algorithm and thus creates a non-equilibrium dynamics which is used to sample the equilibrium quantities.

[*]quer@zib.de
[†]ldonati@zedat.fu-berlin.de
[‡]bettina.keller@fu-berlin.de
[§]weber@zib.de

1 Introduction and Motivation

In Molecular Dynamics simulation (MD) observables like exit times or transition probabilities are often calculated by averages over long trajectories of some continuous space-time-model which describes the molecular movement e.g. an (overdamped) Langevin equation. For these sampling approaches metastability is one of the main difficulties. Metastability means that the trajectory is trapped for a long time in so called metastable state, until it changes very rapidly into another metastable state. The transition between these metastable region happen rarely and thus it is hard to sample them. Mathematically speaking, metastability means that behind the continuous space model there is a Markov Jump Process which describes the jump from one metastable state into another c.f. [1]. In order to sample dynamic quantities between metastable regions the regions have to be known. This is in general also a very hard question but shall not be of a concern in this article. There are many ideas in the literature how this can be achieved see for example [2, 3] and the references therein. Here we assume that the metastable states are known for the considered dynamics.

Thus the computation of dynamics quantities in metastable systems by ensemble averages (we will call these Monte Carlo estimators) is computationally extreme costly. Furthermore the estimators suffer from a high variance. This is why in many applications importance sampling is used. In general one can consider Markov Chain Monte Carlo (MCMC) as some kind of importance sampling because it allows us to perform simulations which only sample in a realistic energetic environment. There are extensions for the problem metastability and MCMC methods like Metadynamics [4], Umbrella Sampling [5] and many more. But all these methods are not designed for the sampling of dynamic quantities and it is not straight forward how to extend them for this purpose. The problem which arise by the calculation of dynamics quantities is sometimes called the sampling problem.

On the other hand there are several ideas in the mathematical literature how an importance sampling strategy for dynamics quantities of complex systems can be designed. In a series of papers different strategies were suggested motivated by a special form of the Hamilton Jacobi Bellman (HJB) equation arising from the large deviation context [6, 7, 8]. These ideas were developed further in the direction of molecular dynamics by [9]. In their paper the authors proposed a sampling scheme which is based on the solution of a deterministic control problem which is associated to the sampling problem. A more sophisticated situation of which could also arise in Molecular Dynamics Simulation was studied in [10]. Here an importance scheme for the situation with resting points was developed. The numerical examples of the paper showed that the for this situation constructed importance sampling scheme is better than the scheme which does not take the resting point into account. [11] proposed a similar strategy also based on the solution of a control problem. In

order to solve this problem the solution of the HJB equation is projected to some space of ansatz functions and then a stochastic optimization problem needs to be solved to find the best approximation. The main difficulty in this approach is to determine the metastable region to place the ansatz functions correctly since the ansatz functions should change the metastability of the dynamics. [12] was able to develop a performance measure for importance sampler related to small noise diffusion processes which give the possibility to compare the different importance sampling schemes analytically.

In order to design an importance sampling scheme one needs a lot of a priori knowledge about the dynamics. Especially the placing of the bias function can be very challenging. Our method circumvents this problem by placing the bias functions automatically assimilated to the dynamics. In order to do so we are going to use the Metadynamics algorithm which is quite famous in MD. This algorithm was first proposed by [13] called Local Elevation and was reintroduced by the name Metadynamics by [4]. The algorithm is designed for sampling the free energy surface. This is done by adding a bias to the states which have been already visited by the trajectory and so fills up the minimum which causes the metastability. In the literature there are many ideas which share the main idea of using an adaptive biasing force to speed up the sampling for example [14, 15] or [16]. These methods belong to a broader class of algorithms which are called adaptive biased forcing (ABF) methods. An overview with rigorous mathematical analysis of the different methods can be found in [17]. But as stated before all of these methods are not designed for the sampling of dynamics quantities. Also general non equilibrium methods in the sense of [18] or [19] which can be related to the methods above are not applicable in this situation because they are not designed for the sampling of dynamics quantities. This is why we are going to assimilate the Metadynamics algorithm and connect it to the standard importance sampling by Girsanov's theorem. In this way it is possible to generate an automatic importance sampling scheme for dynamics quantities for metastable systems.

The paper is structured as follows. We will first give a short introduction into the stochastic analysis which is needed for our method namely Girsanov's theorem. Then we are going to review the Metadynamics algorithm. In the second part we are going show how these two things are combined and how the Metadynamics algorithm is tailored to our needs. We will proof that Girsanov's theorem can be applied in the considered setting. In the end we are going to apply our method to different numerical examples.

2 Theory

In this part we present the two main ingredients of our algorithm. The first part is a short repetition of the main idea behind importance sampling supplemented with stochastic analysis which is needed. The second part is a short introduction into the Metadynamics algorithm.

2.1 Importance Sampling

In this article we consider a diffusion process given by the stochastic differential equation (SDE) as a model of the atomistic movement of a molecule. This SDE satisfies

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

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 some probability space $(\Omega, \mathbb{P}, \mathcal{F})$. The method which is presented in this context can be generalized to the situation in which the term in front of the Brownian motion is state dependent, but we will only consider this simple case here. We moreover assume that the process is trapped in a metastable region $\mathcal{S} \subset \mathbb{R}^n$ which is an open and bounded set with smooth boundary. We further define a target set $\mathcal{T}$ also a open and bounded set with smooth boundary. Furthermore 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 e.g. when an dihedral angle reaches a certain value.

We are interested in dynamics quantities of the from

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

where $X_{0:\tau}$ is a trajectory of (1) until time τ and g is some function on $\mathcal{C}([0,\tau] \times \mathbb{R}^n)$. As pointed out by [9] a very interesting case of this quantity arises if $g = 0$ if $X_t \in \mathcal{S}$ and $g = \infty$ otherwise. Then the quantity of interest (2) becomes

$$\mathbb{P}[X_t \in \mathcal{S}]. \tag{3}$$

The main problem with these expectations is that they cannot be reweighted with the standard techniques like WHAM etc. c.f. [20] or
[19].

A Monte Carlo estimator of (2) looks like this

$$I = \frac{1}{N}\sum_{i=1}^{N} e^{-\beta g(X^i_{0:\tau})} \tag{4}$$

where the $X^i_{0:\tau}$ is a independent sample of the path given by (1). This estimator is unbiased because of the independence of the samples. The variance of the estimator is given by

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

The relative error is given by

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

To build now an importance sampling scheme for this diffusion process (1) one has to increase the depth of the minima which cause the metastable behaviour. The deterministic part of the SDE given in (1), called drift in the SDE literature, is a negative gradient descent. So the process X_t will always stay in the region around the minimum of V when the drift dominates the equation. If the stochastic part, called diffusion, of the equation will dominate the deterministic part the process will leave the minimum and go into the next minimum. To overcome this problem we can change the deterministic part of the SDE. But by changing the deterministic part we are also going to change the quantity of interests. To compensate the change we can use Girsanov's theorem which we are going to state next.

A more physical interpretation of this theorem is that it gives a way to sample equilibrium quantities of some dynamics by sampling the dynamics out of equilibrium.

We now state Girsanov's theorem following [21].

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

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

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

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. Suppose that there exists a process

$$\sigma(Y_t)a_t = u_t \tag{9}$$

which satisfies the Novikov condition

$$\mathbb{E}[\exp(\frac{1}{2}\int_0^T a_t^2 dt)] < \infty. \tag{10}$$

Then we can define a measure $d\mathbb{Q}$ on the probability space for $t \leq T$

$$d\mathbb{Q} = M_T d\mathbb{P} \quad on\ \mathcal{F}_T \tag{11}$$

with

$$M_t = \exp(-\int_0^T a_t dB_t - \frac{1}{2}\int_0^T a_t^2 dt). \tag{12}$$

Proof. [21] p.158 □

In terms of the sampling we are interested in creating different dynamics which is not metastable any more and will give us the same statistics as the original dynamics. To achieve this we can sample the dynamics (13) then reweight the quantity of interest by the weight (15). Decreasing the metastability can be achieved by changing the drift of the dynamics.

To illustrate the idea behind Girsanov's theorem we will give an example in the following. The example is motivated by [21] and extended to our situation.

Example:

The perturbed dynamics then satisfy

$$dY_t = (\sqrt{2}u_t - \nabla V(Y_t))dt + \sqrt{2\beta^{-1}}dB_t, \quad Y_0 = x \tag{13}$$

where $\sqrt{2}u_t$ is a function which changes the gradient of the potential in the metastable regions. To get the right statistics from the perturbed dynamics we construct a new probability measure by using the theorem from Girsanov. For a fixed time $T < \infty$ we have

$$d\mathbb{Q} = M_T d\mathbb{P} \quad \text{on } \mathcal{F}_T \tag{14}$$

with

$$M_t = \exp\Big(-\int_0^T a_t dB_t - \frac{1}{2}\int_0^T |a_t|^2 dt\Big) \tag{15}$$

and a_t is the change of the drift divided by the noise prefractor. In the proposed algorithm we will consider a change of drift which will be independent of the time t, but it is easy to generalize this and to consider a change which is both state and time dependent [1]. In the further presentation we will drop the time dependence and denote the change of drift by a_0

$$a_0 = \frac{1}{\sqrt{2\beta^{-1}}}\sqrt{2}u_t. \tag{16}$$

Then

$$\hat{B}_t := \int_0^t a_0 ds + B_t \tag{17}$$

is a Brownian motion with respect to the measure $\mathbb{Q}$ for $t \leq T$ and we get

$$dY_t = (-\nabla V(Y_t))dt + \sqrt{2\beta^{-1}}d\hat{B}_t. \tag{18}$$

If we know set $Y_0 = x$ we know that the pair $(Y_t, \hat{B}_t)$ is a weak solution of (1) for $t \leq T$, which means that the solution agree on average but not necessarily pathwise. By weak uniqueness the $\mathbb{Q}$-law of Y_t coincides with the $\mathbb{P}$-law of X_t so that we can use

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[f_1(X_{t_1}) \dots f_k(X_{t_k})] = \mathbb{E}_{\mathbb{Q}}[f_1(Y_{t_1}) \dots f_k(Y_{t_k})] = \\ = \mathbb{E}_{\mathbb{P}}[M_T f_1(Y_{t_1}) \dots f_k(Y_{t_k})]\end{aligned}$$

for $f_1 \dots f_k \in \mathcal{C}_0(\mathbb{R}^n)$ and $t_1 \dots t_k \leq T$. In this way it is possible to get statistics which belong to the dynamics at equilibrium (1) from a non-equilibrium simulation (13).

Coming back to the importance sampling. Due to Girsanov's theorem we can now write the importance sampling estimator as

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} e^{-\beta g(Y^i_{0:\tau})} M^i_{0:\tau} \tag{19}$$

where $Y^i_{0:\tau}$ and $M^i_{0:\tau}$ are independent samples from (13) and (15). For a u_t which satisfies Novikov condition and a stopping time which is bounded one can guarantee that M_t is a continuous bounded local martingale with $\mathbb{E}[M_t] = 1$. Then the importance sampling estimator is an unbiased estimator with expectation

$$\mathbb{E}[\hat{I}] = \mathbb{E}[e^{-\beta g(X_{0:\tau})}] \tag{20}$$

cf. [1]. The relative error of this estimator is

$$r(\hat{I}) := \frac{1}{\sqrt{N}} \sqrt{\frac{\mathbb{E}[e^{-2\beta g(Y_{0:\tau})}(M_{0:\tau})^2]}{\mathbb{E}[e^{-\beta g(X_{0:\tau})}]^2} - 1}. \tag{21}$$

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

$$R(\hat{I}) := \sqrt{\frac{\mathbb{E}[e^{-2\beta g(Y_{0:\tau})}(M_{0:\tau})^2]}{\mathbb{E}[e^{-\beta g(X_{0:\tau})}]^2}}. \tag{22}$$

In order to apply Girsanov's theorem for dynamic quantities like exit time in which a stopping time is used one has to make sure that the Novikov condition holds. This can be achieved by different assumptions on the stopping time. We will present two different ways how the assumptions can be formulated in the next two conditions.

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* $0 \leq t_1 \leq t_2 \dots t_k < \infty$. *Let further* $\mathcal{F}_t = \sigma(\{w_s : s \leq t\})$ *denote the* σ*-algebra generated by the Brownian motion* (w_s) *up to time* $t < \infty$. *Then, Girsnaov's theorem holds on the measurable space* $(W, \sigma(W))$ *as long as the family* $(M_t)_{t\geq 0}$ *of random variables*

$$M_t = \exp\Big(\frac{-1}{\sqrt{2\beta^{-1}}}\int_0^t a_s dB_s - \frac{1}{2}\int_0^t \frac{1}{2\beta^{-1}}|a_s|^2 ds\Big) \tag{23}$$

is a uniform integrable martingale. By Itô's formula [21], $(M_t)_{t\geq 0}$ *is a nonnegative local martingale, which is uniformly integrable if Novikov's condition holds* (10). *This could be achieved by taking a admissible control which is defined up to a random stopping time* τ *as it is done in [22].*

Condition 2. *Another way to make sure that Girsanov's theorem is applicable is to assume that the stopping time is bounded for the specific problem. The second assumption is by far non trivial and can only be shown analytically in very few situations. On the other hand from a numerical viewpoint it is impossible to simulate trajectories which have length infinity. So one has to stop the simulation after a finite number of steps. So one can approximate the quantity of interest by the quantity of interest conditioned on the event happens in a finite simulation time. The trajectories which do not simulate the quantity of interest within this simulation time are then excluded from the further calculation or the simulation is repeated. This procedure corresponds exactly to the second argument which guarantees the stopping time to be bounded from which Novikov's condition follow for a reasonable function* u_t. *The now considered stopping time is* $\hat{\tau} = \min(\tau, T_N)$ *where* T_N *is the length of the numerical simulation which is a similar treatment as suggested in [1].*

Condition 1 would mean that one uses a bias function only up to a certain time which is finite. Thus one can guarantee that the integral in the Girsanov weight (15) is bounded. The simulation could go on without that the bias function is 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 ($t > T_N$) one has to repeat the sampling and exclude the trajectory which has not reached the target set yet from the further calculations.

Summarizing Girsanov's theorem gives us an opportunity to sample a dynamic with is less metastable but will give us the same estimators for the quantities of interest by reweigthing the expectation. The main difficulty of applying Girsanov's theorem for this is to determine the metastable region for changing it accordingly. To do this for different systems 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 force method. In order to use this algorithm for our purpose we are going to assimilate the algorithm.

2.2 Metadynamics

Metadynamics, proposed by [4], is an adaptive method for sampling the free energy surface (FES) of high dimensional molecular systems. The method combines dynamics in reaction coordinates with adaptive bias potentials. The idea of the method is to perturb the energy landscape when the simulation is trapped in a metastable region. This is done by adding locally 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 method is originally designed to calculate free energy differences. It is possible to prove that the method converges exactly to the free energy surface when the sampling is done sufficiently long.

The method assumes that the high dimensional system can be projected down on a few relevant collective coordinates s_i $i = 1, \ldots, n$. Only the dependence of these parameters on the free energy $\mathcal{F}(s)$ is considered. One possible way to find these collective variables for stochastic dynamics can be found in [23]. But there are more ideas how this can be achieved in the literature see [20] and the reference therein. The exploration of the free energy surface (FES) is guided by the forces $F_i^t = -\partial\mathcal{F}(s)/\partial s_i^t$. The exploration of the free energy surface gets stuck most often 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 minima. With Metadynamics one constructs a bias potential $V_{bias} : \mathbb{R}^n \to \mathbb{R}$ which is composed of $K \in \mathbb{N}$ bias functions b_i. We define $b_i's$ in the next paragraph. The complete bias potential is then

$$V_{bias}(x) = \sum_{i=1}^{K} w_i \exp\Big(- \frac{(x - c_i)^2}{2\lambda_i^2} \Big) \tag{24}$$

where $w_i \in \mathbb{R}$ is a weight $c_i \in \mathbb{R}$ is the center of the Gaussian and $\lambda \in \mathbb{R}$ is the width. These functions are placed along the trajectory so that it can escape form this region easily. The method can be parallelized easily since the bias force only depends on the history of the individual trajectory. This make the method extremely efficient. The bias also prevents the trajectory from going back to the visited states. When the sampling is done for a sufficient long time the bias will converge to the negative free energy surface plus a constant see e.g. [17].

2.3 Assimilation of Metadynamics

We are going to assimilate the idea of bias forcing to the sampling of dynamics quantities. For our framework we do not have to calculate the complete FES. We only need a bias which makes sure that we do not get trapped in the metastable region. Before we start sampling the quantity of interest we are going to build a

basing function. The major difference of Metadynamics and our method is that instead of changing the potential energy function V we change the gradient of the potential energy ∇V. Thus we generate a dynamics which satisfies

$$dY_t = (\nabla V_{bias}(Y_t) - \nabla V(Y_t))dt + \sqrt{2\beta^{-1}}dB_t \tag{25}$$

where now the gradient of the system is biased with a sum of Gaussian functions. We denote the bias of the gradient as bias function which is defined by

$$\nabla V_{bias}(Y_t) := \sum_{i=1}^{K} b_i(Y_t, c, w, \lambda). \tag{26}$$

The bias function is build in the same way as the bias potential in Metadynamics. But for our purpose it is enough to generate a bias potential only in the metastable region. From a physical point of view the bias function could be interpreted as an additional force which is applied on the system to steer the system in a direction.

Suppose the trajectory is trapped in the metastable set $\mathcal{S}$. The sampling is stopped and the bias function is built. For this we start a trajectory in the metastable region. In every k-th step $k \in \mathbb{N}$ we add a bias function $b_i : (\mathbb{R}^n, \mathbb{R}^n, \mathbb{R}, \mathbb{R}) \to \mathbb{R}$ to the bias function of the form

$$b_i(x; c, w, \lambda) = \frac{w_i}{\sqrt{2\pi\lambda_i^2}} \exp\Big(- \frac{(x - c_i)^2}{2\lambda_i^2}\Big), \quad i \in [1, \dots, K] \tag{27}$$

where $c_i \in \mathbb{R}^n \ \forall i \in [1, \dots, N]$ is the center of the bias function, $w_i \in \mathbb{R} \ \forall i \in [1, \dots, K]$ is a weight and $\lambda_i \in \mathbb{R} \ \forall i \in [1, \dots, N]$ is the width of the bias function. We do this to decrease the metastable region such that the trajectory leaves this region faster. We choose the center c_i of the Gaussian functions to be the current position of the trajectory when the bias function is added. The other parameters w_i and λ_i are constants which must be assimilated to the simulation. Especially the sign of the w_i has to be chosen carefully. The proposed algorithm does not generate a flat energy landscape since the gradient of the system is perturbed. The resulting dynamics is a Brownian motion with drift. This is why one needs to know in which direction the dynamics has to be steered. The signs of w_i's have to be chosen accordingly. Other ways of getting the parameters for the bias function are possible. We introduce another idea in 1.

When the trajectory hits the target set we save the bias potential. The bias potential consists of *number of steps needed*$/k$ bias functions. The choice of k is a compromise between adding as less bias functions as necessary, getting a small hitting time τ and not perturbing the potential too much. It depends on the parameters w_i and λ_i how many bias functions are 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. This is why all parameters should be adapted to the problem such that the computation does not get to costly.

After the bias potential is built the sampling of the original trajectory is continued with the biased potential. To resume the original values of the quantity of interest the calculated values have to be corrected by the formula given in (15). So while sampling the perturbed dynamics the Girsanov weights M_t has to be sampled as well, but these values can be computed on the fly.

Remark 1. *In order to create a good bias potential within a reasonable computational cost one can use the history of the trajectory to estimate the parameters of the bias functions. The midpoint c_i can be chosen to be the mean of the average of the last k steps and the λ_i can be chosen to be the maximal distance form the starting point of the last k steps times a constant, $C(\lambda_i = \max(|X_{i*(1:k)} - c_i|)),\ C \in \mathbb{R}$. This can be more efficient as we will see in the example shown above. In the literature there are a lot of extensions of Metadynamics which could be used as well e.g. [24]*

Remark 2. *The construction of the bias potential depends on the history of the trajectory. But since the simulation to get the bias function is done in a additional step the potential is not getting time dependent. So the simulated dynamic does not have to be changed only the potential changes. Furthermore the discretization of* (1) *always gives a discrete time Markov process because of the independence of the Brownian motion. The construction of the bias potential itself is not Markovian because it depends on the history of the trajectory. But since the construction of the bias function and the sampling of the quantity of interest are done independent of each other the bias does not have an influence of the Markovianity of the perturbed SDE* (13)*. In general an extension of the proposed method for Non-markovian dynamics should be possible. For this one could use the Metadynamics methods proposed in [16].*

3 The algorithm

We now present the algorithm in pseudo code. We will first use the Metadynamics algorithm to build an bias potential such that the dynamics is not metastable any more. We then sample the transition in this biased potential N times and weight the sampling with the weight given by the Girsanov's theorem.

3.1 Pseudo Code

Data: dynamics Y_t, starting set $\mathcal{S}$, traget set $\mathcal{T}$
Result: estimators for quantities of interest
initialisation: $Y_0 = x$; a_i, λ_i
Step 1: Build bias potential
while *transition has not occurred* **do**
  sample the dynamic Y_t;
  every k steps add a new bias function;
end
save the bias potential;
Step 2: Sample the quantity of interest
for N **do**
  sample the transition in the biased potential;
  sample the Girsanov weights for the biased potential;
end
reweight according to Girsanov;

3.2 Proof of Novikov condition

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

Lemma 1. *Let $\hat{\tau}$ be the stopping time as given in Condition 2. Further let $\mathcal{S} \subset \mathbb{R}^n$ be a bounded domain and the biased potential consists of $N < \infty$ bias functions. We choose the weights of the bias function such that the sum over the squares is bounded $\sum_{i=1}^{N} a_i^2 < C_1$. When the perturbation potential is uniformly continuous then the Novikov condition hold and we can use the Girsanov theorem to calculate path dependent quantities from non equilibrium sampling for the equilibrium dynamics.*

Proof. Novikov's conditions states that

$$\mathbb{E}[\exp(\frac{1}{2}\int_0^{\hat{\tau}} |V_{bias}(Y_s; c, w, \lambda)|^2 ds)] < \infty. \tag{28}$$

We assume that the metastable region is bounded so we can further assume the the ansatz functions added by the metadynamics have bounded support. We are going to show that the integral $\int_0^{\tau} |f(x; c, \lambda)|^2 ds$ is bounded from which we can then follow that the Novikov condition hold.

$$\int_0^{\hat{\tau}} |V_{bias}(Y_s; c, \lambda)|^2 ds = \int_0^{\hat{\tau}} \Big| \sum_{i=1}^{N} a_i \exp\Big(- \frac{(Y_s - c_i)^2}{2\lambda^2}\Big)\Big|^2 ds$$

$$
\begin{aligned}
&\leq \int_0^{\hat{\tau}} \Big(\sum_{i=1}^{N} \Big| a_i \exp\Big(- \frac{(Y_s - c_i)^2}{2\lambda^2} \Big) \Big| \Big)^2 ds \\
&= \int_0^{\hat{\tau}} \Big(\sum_{i=1}^{N} a_i \exp\Big(- \frac{(Y_s - c_i)^2}{2\lambda^2} \Big) \Big)^2 ds \\
&\leq \int_0^{\hat{\tau}} \sum_{i=1}^{N} a_i^2 \sup_{Y_s \in \mathcal{S}} \sum_{i=1}^{N} \exp\Big(- \frac{(Y_s - c_i)^2}{2\lambda^2} \Big)^2 ds \\
&= \int_0^{\hat{\tau}} \sum_{i=1}^{N} a_i^2 \sum_{i=1}^{N} \sup_{Y_s \in \mathcal{S}} \exp\Big(- \frac{(Y_s - c_i)^2}{2\lambda^2} \Big)^2 ds \\
&\leq \int_0^{\hat{\tau}} C_1 C_2 dt = C\hat{\tau} < \infty
\end{aligned}
$$

We know that the $\sup_{x \in \mathcal{S}} \sum_{i=1}^{N} \exp\Big(- \frac{(Y_s - c_i)^2}{\lambda^2} \Big)^2$ is bounded because we $\exp(x)$ is a continuous differentiable function on a bounded domain and so is its sum. The sup can be pulled inside the sum because all bias functions are independent of each other. Furthermore the stopping time is bounded due to the assumption in Condition 2. If follows that the whole expression is bounded and from this we conclude that (28) is satisfied and thus Novikov's condition holds. □

4 Examples

In the following we study different applications of the method presented above and the presented variants. We will first study a low dimensional example in which we show all variants of our method first the construction of the bias potential with fixed parameters, second the construction of the bias potential with estimating the parameters form the history of the trajectory as proposed in Remark 1.

We consider the dynamics given by (1) and the potential given by

$$
V(x) = \frac{1}{2}(x^2 - 1)^2. \tag{29}
$$

This potential has two minima at $x = \pm 1$ and a local maximum at $x = 0$. We are interested in the transition probability for (1) starting at $X_0 = -1$ with $\beta = 3.0$ for all experiments. The stopping time is defined as the first hitting time of the set $\mathcal{T} = [0.9, 1.1]$. We sample the dynamics with a standard Euler Mayurama discretization with a time step $\Delta t = 10^{-4}$ c.f. [25]. We are going to compute two quantities first the transition probability for which we are going to chose the function $g(Y_s) = 0$ for $Y_s \in \mathcal{S}$ and $g(Y_s) = 1$ for $Y_s \in \mathcal{T}$. Secondly we are going to compute the moment generating function of the stopping time for which we chose

$g(Y_s) = \int_0^{\hat{\tau}} \mathbb{1}_{\mathcal{T}}(Y_s)ds$ where $\mathbb{1}_{\mathcal{T}}$ is the indicator function of the target set $\mathcal{T}$. For both quantities we are interested in getting estimators with a lower variance compared to the Monte Carlo estimator (4) and a speed up in the simulations.

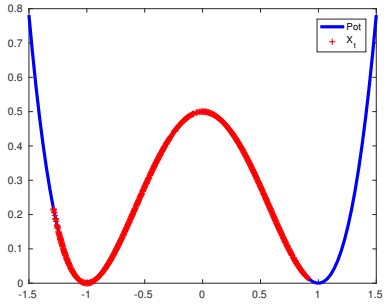

Figure 1: In blue the potential function (29) is shown and red is a realisation of (1) shwoing the desired transition we want to sample.

4.1 One dimensional diffusion in a double well potential with fixed parameters

For this computation we chose $w_i = 0.1,\ \lambda_i = 0.8$ for all bias functions. The sign of the w_i's are chosen to be positive since we want to observe a transition from the left to the right well. The c_i of every bias function is chosen as the current value of the trajectory when the new bias function is added to the bias potential. The maximal simulation time was $T_N = 15000$ time steps. We also fixed the random number generator to have a better comparison within the different experiments. We computed 1000 realization of trajectories and calculated the quantities of interest over the ensemble average (Monte Carlo estimator). To variance of the estimators was calculated by repeating the experiment 10 times. The random number generator was set to rng(i,'twister'), where i is the number of iteration in the variance experiment. The experiments was repeated for different random number generators. The experiments showed similar results as shown in the table above.

In this example 66 bias functions have been used. The estimators of the Monte Carlo and the importance sampling agree in both cases. One also sees that the variance of the estimator is reduced for both quantities of interest. The variance for the transition probability is reduce by a factor 3/2 and for the moment generating function by a factor of 2. So the automatically generated bias potential by the adjusted Metadynamics algorithm is actually a good potential in the sense of importance

[1]This is actually the conditional probability $P(X_t \in \mathcal{T} | t \leq T_N)$. The mathematical misuse is due to space issues.

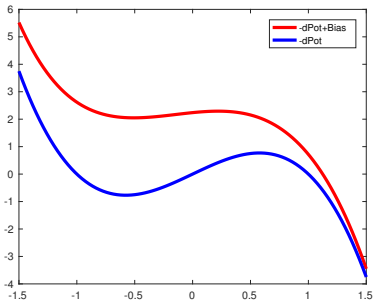

Figure 2: In red the gradient of the potential function (29) is shown and in blue the gradient of the biased potential is shown.

	MC	GIR
$P(X_t \in \mathcal{T})$ [1]	4.87×10^{-2}	5.18×10^{-2}
Var	5.9×10^{-5}	3.9×10^{-5}
$R(I)$	0.1588	0.1287
$\mathbb{E}[e^{-\beta\tau}]$	2.6×10^{-3}	2.5×10^{-3}
Var	1.3×10^{-7}	8×10^{-8}
$R(I)$	0.1412	0.1088
MFHT	1.0901	0.8626
Δt	10^{-4}	10^{-4}

Table 1: Comparison of the importance sampling estimators and the Monte Carlo estimators for the simulation with fixed parameters of the biased potential.

sampling. The reduction of the MFHT means that on average the trajectories of the importance sampling scheme are about 2000 steps shorter. This shows that also a speed up in simulating the interesting events was achieved.

4.2 One dimension with estimated parameters

In this experiment we build the bias potential by estimating the parameters. The parameter λ_i is calculated from the history of the trajectory as described in Remark 1 with the $C = 10$. The center of the bias function was calculated by the average over the last k steps of the trajectory. All other parameters are kept as in the experiment above.

In this example 67 bias functions have been used. Again we see that the Monte Carlo estimators and the importance sampling estimator agree quite good. We have again a variance reduction for both quantities of interest. But the variance for the estimator for the hitting probability is not as good as in the fixed parameter case.

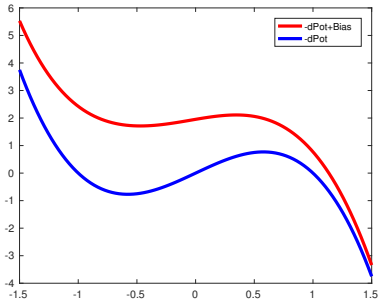

Figure 3: In blue the gradient of the potential function (29) is shown and in red the gradient of the biased potential is shown.

	MC	GIR
$P(X_t \in \mathcal{T})$ [1]	4.87×10^{-2}	5.45×10^{-2}
Var	5.9×10^{-5}	5.3×10^{-5}
$R(I)$	0.1588	0.1364
$\mathbb{E}[e^{-\beta\tau}]$	2.6×10^{-3}	2.8×10^{-3}
Var	1.3×10^{-7}	8×10^{-8}
$R(I)$	0.1412	0.1014
MFHT	1.0901	0.7716
Δt	10^{-4}	10^{-4}

Table 2: Comparison of the importance sampling estimators and the Monte Carlo estimators for the simulation with estimated parameters of the biased potential.

In the case for the moment generating function the desired variance reduction is achieved comparable to the case of the fixed parameters. Furthermore the MFHT is shorter compared to the first calculation.

5 Summary and Outlook

In this article we developed an algorithm for automatic assimilated importance sampling. The main idea is to generate a similar dynamics as the original one in which the metastability is reduced. This can be achieved by constructing bias potentials which raise the region around the minimum of the potential energy function. For this we combined the algorithm of Metadynamics which was adjusted to produce bias potentials with a reweigting strategy based on Girsanov's theorem. We proofed that the reweigting scheme based on the theorem of Girsanov can be applied in our situation. We also invented an algorithm which uses the history of the dynamics to

build bias potentials which are even better adapted to the potential. In a one dimensional example we could show that this algorithm generates estimators for different quantities of interest with a reduced variance compared a Monte Carlo estimator. The bias potential also speeded up the sampling of the event. The variant of the algorithm that used the history of the trajectory to estimate the parameters for the bias potential provided a better bias potential in the sense of variance reduction and speed up. It seems that a bigger bias potential is always better but this is in general not the case because the handling of the Girsanov weights gets numerically quite delicate.

For future work we want to try to analyse our importance sampling scheme. There are a lot of ideas in the literature how this can be done e.g. [26] or [12]. If this is possible and under which condition we can guarantee that the bias potential satisfies the conditions of the theory presented in these articles is ongoing research.

Our proposed method could also be used to generate a first initial guess for the optimization problem proposed in [11]. The bias potential then would have to be approximated by a coarser set of ansatz functions. Then one could solve the optimization problem by a gradient descent in order to find the best weights to approximate the best bias potential in the sense of the approximation. The combination of those two methods is also ongoing research.

Acknowledgement

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).

References

[1] Tony Lelièvre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[2] B. Keller, X. Daura, and W. van Gunsteren. Comparing geometric and kinetic cluster algorithms for molecular simulation data. *The Journal of Chemical Physics*, 132(7):074110, 2010.

[3] O. Lemke and B.G. Keller. Density-based cluster algorithms for the identification of core sets. *The Journal of Chemical Physics*, 145(16):164104, 2016.

[4] A. Laio and M. Parrinello. Escaping free- energy minima. *PNAS*, 20(10):12562–12566, 2002.

[5] G.M Torrie and J.P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[6] Paul Dupuis and Hui Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastic Reports*, 76(6):481–508, 2004.

[7] Paul Dupuis, Ali Devin Sezer, and Hui Wang. Dynamic importance sampling for queueing networks. *The Annals of Applied Probability*, 17(4):1306–1346, 08 2007.

[8] Paul Dupuis and Hui Wang. Subsolutions of an isaacs equation and efficient schemes for importance sampling. *Mathematics of Operations Research*, 32(3):723–757, 2007.

[9] Eric Vanden-Eijden and Jonathan Weare. Rare event simulation of small noise diffusions. *Communications on Pure and Applied Mathematics*, 65:1770 –1803, 2012.

[10] P. Dupuis, K. Spiliopoulos, and Xiang Zhou. X. Escaping from an attractor: Importance sampling and rest points i. *Annals of Applied Probability*, 25(5):2909–2958, 2015.

[11] C. Hartmann and Ch. Schütte. Efficient rare event simulation by optimal nonequilibrium forcing. *Journal Statistical Mechanics Theory and Experiments*, 2012, 2012.

[12] Konstantinos Spiliopoulos. Non-asymptotic performance analysis of importance sampling schemes for small noise diffusions. *Journal of Applied Probability*, 52:1–14, 2015.

[13] T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *Journal of Computer-Aided Molecular Design*, 8:695–708, 1994.

[14] F. Wang and D. P. Landau. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E*, 64:056101, Oct 2001.

[15] J. Hénin and C Chipot. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *The Journal of Chemical Physics*, 121(7):2904–2914, 2004.

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

[17] T. Leliévre, Mathias Rousset, and Gabriel Stoltz. *Free energy computations : a mathematical perspective*. Imperial College Press, London, Hackensack (N.J.), Singapore, 2010.

[18] R. Zwanzig. High temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, 22(8):1420–1426, 1954.

[19] 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(8):1011–1021, 1992.

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

[21] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Hochschultext / Universitext. Springer, Berlin, 2003.

[22] Carsten Hartmann, Christof Schütte, and Wei Zhang. Model reduction algorithms for optimal control and importance sampling of diffusions. *Nonlinearity*, 29(8):2298, 2016.

[23] Wei Zhang, Carsten Hartmann, and Christof Schütte. Effective dynamics along given reaction coordinates, and reaction rate theory. *Faraday Discuss.*, 2016.

[24] A. Barducci, G. Bussi, and M. Parrinello. Well tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100, Jan 2008.

[25] D.J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(03):525–546, 2001.

[26] T. Leliévre. Two mathematical tools to analyze metastable stochastic processes. *Numerical Mathematics and Advanced Applications*, 224(12):791–810, 2011.