IMPORTANCE SAMPLING IN PATH SPACE FOR DIFFUSION PROCESSES

WEI ZHANG[2] , CARSTEN HARTMANN[1,*] , MARCUS WEBER[2] , AND CHRISTOF SCHÜTTE[1,2]

Abstract. Importance sampling is a widely used technique to reduce the variance of the Monte Carlo method. It uses the idea of change of measure to design efficient Monte Carlo estimators. In this work, we study the importance sampling method in the framework of diffusion process and consider the change of measures which can be realized by adding a control force to the original dynamics. For certain exponential-type expectations, the corresponding control force of the optimal change of measure leads to a zero-variance estimator that is related to the solution of a Hamilton-Jacobi-Bellmann equation. We first show that, for a general suboptimal control force, the variance of the resulting estimator is bounded by the L^∞ distance between the suboptimal and the optimal control force. We consider three situations in which we can approximate this optimal control force, thus obtaining efficient estimators with small variance. Numerical examples show the effectiveness of these approximation strategies. The asymptotic optimality of the change of measure approach is proved.

Key words. importance sampling, Hamilton-Jacobi-Bellmann equation, Monte Carlo method, change of measure, rare event, diffusion process.

AMS subject classifications.

1 Introduction Monte Carlo methods are widely established numerical methods among scientists from different disciplines, such as biology, chemistry, physics or engineering (e.g., see [22] and the references therein). They have quite a long history and receives wide applications since the invention of the computer. Nowadays, both the developments and applications of varieties of the standard Monte Carlo method are still attractive to researchers who are confronted with solving high-dimensional practical problems for which deterministic or direct methods become infeasibly expensive. These varieties includes MCMC methods [20, 7], Hybrid Monte Carlo methods [12, 30], Sequential Monte Carlo methods [26, 11], etc. Most of them have been successfully applied to solving different high-dimensional problems [25].

For standard Monte Carlo methods, variance reduction is the key issue to obtain efficient estimations. Although the variances of N samples have the same $O(N^{-\frac{1}{2}})$ scaling order for all different unbiased estimators, the prefactor constants are related to the variances of the estimators and play an important role in the performance of any Monte Carlo methods. Several variance reduction techniques exist in order to decrease the prefactor and thus increase the accuracy or efficiency of the estimators. In this paper, we will focus on one kind of such techniques, the importance sampling method, which is widely used in applications. The basic idea of this method is that, instead of sampling from the original probability distribution, samples are generated from another probability distribution, under which the "important" regions in state space are more frequently sampled. To further illustrate it, consider the situation when a certain region in state space is much more "important" than other parts for computing an expectation value, while its probability measure is very small. In this case, due to the rareness of this region's measure, standard Monte Carlo method will probably fail to sample it sufficiently, which therefore indicates large variance. The importance sampling method, on the other hand, can provide smaller variance and more accurate estimations by sampling the state space using a

[1]Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[2]Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany
[*]Corresponding author. Email: chartman@mi.fu-berlin.de

different probability distribution under which the probability measure of the "important" region is enlarged. Obviously, the selection of this alternative probability distribution is crucial in order to design efficient Monte Carlo estimators and how to choose this probability distribution is the main topic for researchers and practitioners who study or use the importance sampling method. See [6, 17, 2, 14, 13, 34] and references therein.

In this paper, we will focus on the path sampling problem in diffusion processes. More specifically, given certain diffusion process which are described by stochastic differential equations (SDEs), it is known that it defines a probability measure over all the path ensembles, and our aim is to compute the average of some path functional with respect to this induced measure. In this setting, we want to apply the importance sampling method and thus modify the original diffusion process by adding a control force term to it and generate path ensembles from this modified dynamics. This will lead to a change of measure to which the Girsanov theorem [28] of SDE theory applies. We further confine ourselves to certain exponential type of path functionals which will be explicitly given below. For this type of path functional, the optimal change of measure exists and, when importance sampling is used, the estimator's variance becomes zero. Furthermore, the path average can be connected to certain optimal control problem for the diffusion process by adopting logarithmic transformation, and the optimal change of measure is related to a Hamilton-Jacobi-Bellmann (HJB) equation, which, however, is hard to solve when the state space is high-dimensional.

While generally it is impractical to find the optimal control force efficiently, there is hope that approximations to it, to which we will refer as "desired" control forces can be computed that will result in small variance estimators. The main purpose of this paper is to consider situations when such desired control forces can be efficiently designed without solving the high-dimensional HJB equation. To achieve this, we first study the estimator's variance when an arbitrary control force is applied and conclude that, roughly said, the closer this control force to the optimal one, the smaller the variance is. This indicates that designing small variance estimators is equivalent to approximating the HJB equation. From this point of view, we study the asymptotic equations obtained by taking the zero temperature limit and/or time-scale separation limit of the original HJB equation. If these resulting equations can be solved efficiently, we obtain desired control forces and can design small variance estimators based on them. Although, ideally, solving this asymptotic equations already gives us the asymptotic expectation value we want to compute, we emphasize the necessity of our approach because of the following reasons. First, it is not easy to verify whether the original dynamics is in the zero-temperature or time-separation limit regime. Secondly, even if it is, we don't know how close it is to the true expectation value by solving the asymptotic equations alone. On the contrary, our approach provides reliable results and has good performance even when the system is not in the limit regime (see Section 4). Furthermore, since sampling will be performed using the estimators, the asymptotic equations can be solved on a much coarser grid and thus the computation of the control forces is cheap.

Our work is inspired by the works [13, 34], in which rare event simulation was studied, using approximations of the optimal bias on the basis of large deviation arguments. In [13], the authors considered multiscale diffusions with a specific form and the desired control force was found by studying the lower bound of the rate functional in the context of large deviation theory. In [34], the authors proposed to compute the desired control force by solving a deterministic optimal control problem. Comparing to these works, we consider a more general situation and provide a unified point of view in designing efficient importance sampling strategies.

This paper is organized as follows. In Section 2, after describing the exponential expectations studied in this work, we give a general introduction of the importance sampling method in the diffusion process setting. We study the variance of Monte Carlo estimators corresponding to general control forces and give an upper bound for it. In Section 3, we consider three situations in which we can find such desired control forces by approximating the HJB equation. In Section 4, numerical examples are studied to demonstrate the performance of these approximation strategies. The asymptotic optimality of these control forces are proved in Appendix A.

2 Setup and main result We consider the conditional expectation [13, 34]

$$I = \mathbf{E}\Big(\exp\Big(-\beta \int_t^T h(z_s)\, ds\Big) \;\Big|\; z_t = z\Big) \tag{2.1}$$

where $\beta > 0$, time interval $[t, T]$ is fixed. $z_s \in \mathbb{R}^n$ follows the dynamics

$$\begin{aligned} dz_s &= b(z_s)ds + \beta^{-1/2}\sigma(z_s)dw_s, \quad t \le s \le T \\ z_t &= z \end{aligned} \tag{2.2}$$

w_s is a standard $n-$dimensional Wiener process. It is known that (2.2) induces a probability measure $\mathbf{P}$ over the path ensembles $z_s, t \le s \le T$ starting from z. To apply the importance sampling method, we introduce

$$d\bar{w}_s = \beta^{1/2} u_s\, ds + dw_s, \tag{2.3}$$

where u_s will be referred to as the *control force*. Then it follows from Girsanov theorem that $\bar{w}_s$ is a standard Wiener process under $\bar{\mathbf{P}}$, where the Radon-Nikodym derivative is

$$\frac{d\bar{\mathbf{P}}}{d\mathbf{P}} = Z_t = \exp\Big(-\beta^{1/2} \int_t^T u_s\, dw_s - \frac{\beta}{2}\int_t^T |u_s|^2 ds\Big). \tag{2.4}$$

In the following, we will omit the conditioning on the initial value at time t . Let $\bar{\mathbf{E}}$ denote the expectation under probability measure $\bar{\mathbf{P}}$, then we have

$$I = \mathbf{E}\Big(\exp\Big(-\beta \int_t^T h(z_s)\, ds\Big)\Big) = \bar{\mathbf{E}}\Big(\exp\Big(-\beta \int_t^T h(z_s^u)\, ds\Big) Z_t^{-1}\Big), \tag{2.5}$$

with variance

$$\mathrm{Var}_u I = \Big[\bar{\mathbf{E}}\Big(\exp\Big(-2\beta \int_t^T h(z_s^u)\, ds\Big) (Z_t)^{-2}\Big) - I^2\Big]. \tag{2.6}$$

Now (2.2) has the representation

$$\begin{aligned} dz_s^u &= b(z_s^u)ds - \sigma(z_s^u)u_s\, ds + \beta^{-1/2}\sigma(z_s^u)d\bar{w}_s\,, \quad t \le s \le T \\ z_t^u &= z. \end{aligned} \tag{2.7}$$

The *relative standard deviation*

$$\rho_u(I) = \frac{\sqrt{\mathrm{Var}_u I}}{I} \tag{2.8}$$

is often used to quantify the efficiency of the Monte Carlo method. Now consider the calculation of (2.5) by a Monte Carlo sampling in path space. Suppose N trajectories $\{z_s^{u,i}, t \leq s \leq T\}$ of (2.7) have been generated, where $i = 1, 2, \cdots, N$. Using Monte Carlo method, we obtain the unbiased estimator

$$I_N = \frac{1}{N}\sum_{i=1}^{N}\Big(\exp\Big(-\beta\int_t^T h(z_s^{u,i})\,ds\Big)(Z_t^{u,i})^{-1}\Big)\,, \tag{2.9}$$

and its asymptotic variance (for large N) is

$$\mathrm{Var}_u I_N \simeq \frac{\mathrm{Var}_u I}{N} = \frac{1}{N}\Big[\bar{\mathbf{E}}\Big(\exp\Big(-2\beta\int_t^T h(z_s^u)\,ds\Big)(Z_t)^{-2}\Big) - I^2\Big]. \tag{2.10}$$

Notice that $Z_t = 1$ when $u_s \equiv 0$, and we recover the standard Monte Carlo method. The advantage of introducing the control force u_s is that we may choose u_s to decrease the variance of the estimator (2.9). From (2.6) (2.10), this means to choose u_s such that the expectation

$$\bar{\mathbf{E}}\Big(\exp\Big(-2\beta\int_t^T h(z_s^u)\,ds\Big)(Z_t)^{-2}\Big) \tag{2.11}$$

is as close as possible to I^2.

2.1 Dual optimal control problem To proceed, we make use of the important dual relation [8, 9]

$$\log\mathbf{E}\Big(\exp\Big(-\beta\int_t^T h(z_s)\,ds\Big)\Big) = -\beta\inf_{u\in\mathcal{A}}\bar{\mathbf{E}}\Big\{\int_t^T h(z_s^u)\,ds + \frac{1}{2}\int_t^T |u_s|^2 ds\Big\}. \tag{2.12}$$

The infimum on the right-hand side of (2.12) is taken over a space $\mathcal{A}$ of admissible Markovian feedback controls of the form $u_s = c(s, z_s)$, with a suitable function $c\colon [t,T]\times\mathbb{R}^n \to \mathbb{R}^n$. We call $\hat{u}_s$ the minimizer of (2.12), the *optimal control force*. Let $\hat{w}_s, \hat{Z}_t, \hat{\mathbf{P}}$ be defined as in (2.3) (2.4) by substituting u_s with $\hat{u}_s$, and let $\hat{z}_s$ satisfy (2.7) when $\hat{u}_s$ is applied. Using Jensen's inequality one can show that (2.12) implies

$$\exp\Big(-\beta\int_t^T h(\hat{z}_s)\,ds\Big)\hat{Z}_t^{-1} = I, \qquad \hat{\mathbf{P}} - a.s. \tag{2.13}$$

Combining this with (2.10), it means that, with $\hat{u}_s$, the change of measure is optimal in the sense that it gives zero-variance Monte Carlo estimator.

On the other hand, notice that the right-hand side of (2.12) can be interpreted within the optimal control theory of diffusion process. Let

$$U(t,z) = \inf_{u\in\mathcal{A}}\bar{\mathbf{E}}\Big\{\int_t^T h(z_s^u)\,ds + \frac{1}{2}\int_t^T |u_s|^2 ds \;\Big|\; z_t = z\Big\}. \tag{2.14}$$

By the dynamic programming principle, $U(t,z)$ satisfies the following HJB equation,

$$\begin{aligned}&\frac{\partial U}{\partial t} + \inf_{c\in\mathbb{R}^n}\Big\{h + \frac{1}{2}|c|^2 + (b - \sigma c)\cdot\nabla U + \frac{\sigma\sigma^T}{2\beta} : \nabla^2 U\Big\} = 0,\\ &U(T,z) = 0\end{aligned} \tag{2.15}$$

with the optimal control force $\hat{u}_s$ given by

$$\hat{u}_s = c(s, z_s) = \sigma^T(z_s)\nabla U(s, z_s). \tag{2.16}$$

Variance estimates and statement of the main result Now we can estimate (2.11), thus also the variance (2.6) for an arbitrary u_s. First we suppose that the probability measure $\bar{\mathbf{P}}$ and $\hat{\mathbf{P}}$ are equivalent. From (2.13), we conclude

$$\exp\Big(-\beta\int_t^T h(\hat{z}_s)\,ds\Big)\hat{Z}_t^{-1} = I, \qquad \bar{\mathbf{P}} - a.s. \tag{2.17}$$

From this, we get

$$\begin{aligned}\bar{\mathbf{E}}\Big(\exp\Big(-2\beta\int_t^T h(z_s^u)ds\Big)(Z_t)^{-2}\Big) &= \bar{\mathbf{E}}\Big(\exp\Big(-2\beta\int_t^T h(z_s^u)ds\Big)(\hat{Z}_t)^{-2}\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big)\\ &= I^2\bar{\mathbf{E}}\Big(\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big).\end{aligned} \tag{2.18}$$

Using (2.4) we get

$$\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2 = \exp\Big(-2\beta^{1/2}\int_t^T(\hat{u}_s - u_s)dw_s - \beta\int_t^T(|\hat{u}_s|^2 - |u_s|^2)ds\Big). \tag{2.19}$$

In order to simplify (2.18), we introduce another control force $\widetilde{u}_s$ and change the measure again. Specifically, we choose $\widetilde{u}_s = 2\hat{u}_s - u_s$ and define $\widetilde{w}_t, \widetilde{\mathbf{P}}_t, \widetilde{Z}_t$ as in (2.3) (2.4) by replacing u_s with $\widetilde{u}_s$. Let $\widetilde{\mathbf{E}}$ denote the expectation w.r.t $\widetilde{\mathbf{P}}$. From (2.18) and (2.19), we then get

$$\bar{\mathbf{E}}\Big(\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big) = \widetilde{\mathbf{E}}\Big(\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\widetilde{Z}_t^{-1}Z_t\Big) = \widetilde{\mathbf{E}}\Big(\exp\Big(\beta\int_t^T|\hat{u}_s - u_s|^2ds\Big)\Big). \tag{2.20}$$

Applying the dual relation (2.12) once again, we obtain

$$\begin{aligned}\log\bar{\mathbf{E}}\Big(\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big) &= \log\widetilde{\mathbf{E}}\Big(\exp\Big(\beta\int_t^T|\hat{u}_s - u_s|^2ds\Big)\Big)\\ &= \sup_v\widetilde{\mathbf{E}}^v\Big(\beta\int_t^T|\hat{u}_s - u_s|^2ds - \frac{\beta}{2}\int_t^T|v_s|^2ds\Big)\\ &\le \sup_v\widetilde{\mathbf{E}}^v\Big(\beta\int_t^T|\hat{u}_s - u_s|^2ds\Big)\end{aligned} \tag{2.21}$$

where

$$\frac{d\widetilde{\mathbf{P}}^v}{d\widetilde{\mathbf{P}}} = \exp\Big(-\beta^{\frac{1}{2}}\int_t^T v_s\,d\widetilde{w}_s - \frac{\beta}{2}\int_t^T|v_s|^2ds\Big) \tag{2.22}$$

and $\widetilde{\mathbf{E}}^v$ denotes the expectation with respect to the measure $\widetilde{\mathbf{P}}^v$.

From (2.21), we may conclude that, in order to keep the variance small, we should choose a control u which is uniformly close to the optimal one, $\hat{u}$. Furthermore, if we have an upper bound on $||u - \hat{u}||_{L^\infty}$, then it follows that

$$\log\bar{\mathbf{E}}\Big(\Big(\frac{\hat{Z}_t}{Z_t}\Big)^2\Big) \le \beta||u - \hat{u}||^2_{L^\infty}(T-t), \tag{2.23}$$

where by $||u-v||_{L^\infty} \triangleq \sup\{|u(s,z) - v(s,z)|; z \in \mathbb{R}^n, s \in [t,T]\}$ we denote the norm on the space $C([t,T]\times\mathbb{R}^n, \mathbb{R}^n)$ of continuous functions from $[t,T]\times\mathbb{R}^n$ to $\mathbb{R}^n$.

Combining (2.6) with (2.18), we have obtained the following theorem on the upper bound of the variance for an arbitrary control force u_s.

THEOREM 2.1. *Consider the expectation I given in (2.5) computed by importance sampling with some control force u. Suppose that $\hat{u}$ is the optimal control force obtained from (2.15) and (2.16) and $||u-\hat{u}||_{L^\infty}$ is finite. Then the following upper bound for the variance of I holds*

$$\mathrm{Var}_u I \leq I^2\Big(\exp\Big(\beta||u-\hat{u}||^2_{L^\infty}(T-t)\Big)-1\Big).$$

For $\delta = \beta||u-\hat{u}||^2_{L^\infty}(T-t) \leq 1/2$ we thus have

$$\rho_u(I) \leq \sqrt{2\delta}.$$

The main conclusion of the above derivations is that, while the optimal control force leads to an optimal estimator of zero-variance, a suboptimal one will lead to an estimator of small variance. In the next section, we will discuss several situations when the suboptimal control force can be approximated based on simplifications of the dynamics.

But before closing this section, we introduce the *relative standard deviation* $\rho_s(I)$ for the *sample variance*

$$\mathrm{Var}_s I = \frac{1}{N}\sum_{i=1}^{N}\bigg(\Big(\exp\Big(-\beta\int_t^T h(z_s^{u,i})\,ds\Big)(Z_t^{u,i})^{-1}\Big)-I_N\bigg)^2, \tag{2.24}$$

by

$$\rho_s(I) = \frac{\sqrt{\mathrm{Var}_s I}}{I_N}, \tag{2.25}$$

which will be used in the numerical examples in Section 4. Note that these quantities also depend on the control u, but this dependence is omitted to simplify the notations.

3 Suboptimal control forces The purpose of this section is to discuss strategies to find desired control forces which give efficient Monte Carlo estimators (2.9). Our starting point is based on the observations that the optimal control force $\hat{u}$ satisfies (2.16), which is related to the solution of the HJB equation (2.15), as well as the upper bound (2.23) of the variance. Generally, it is difficult to find the exact solution $\hat{u}$ by solving (2.15), which is a high-dimensional PDE. In the following, we consider situations when $\hat{u}$ can be approximated by certain u, which is easily computable from simpler equations related to (2.15).

Some notation: We introduce some function spaces and norms. Let M_n and S_n denote the set of general and symmetric definite $n\times n$ matrices. We call $C_b^{k,l}(\Omega_1,\Omega_2)$, with $\Omega_1 \subseteq [0,t)\times\mathbb{R}^n$ and $\Omega_2 = \mathbb{R}^q, M_q$ or S_q, the set of functions $f:\Omega_1\to\Omega_2$ with bounded continuous time derivatives up to order k and bounded continuous spatial derivatives up to order l. For $\Omega_1\subseteq\mathbb{R}^n$, $C_b(\Omega_1,\Omega_2)$ denotes the set of bounded continuous functions $f:\Omega_1\to\Omega_2$, while $C_b^m(\Omega_1,\Omega_2)$ denotes the set of functions with bounded continuous spatial derivatives up to order m. For $f\in C_b(\Omega_1,\Omega_2)$, $||f||_{L^\infty} = \sup\{|f(x)|, x\in\Omega_1\}$ denotes the L^∞ norm on $C_b(\Omega_1,\Omega_2)$. When $\Omega_2=\mathbb{R}^1$, we simply write the above function spaces as $C_b^{k,l}(\Omega_1)$, $C_b(\Omega_1)$, $C_b^m(\Omega_1)$.

3.1 Small temperature limit In this subsection, we suppose dynamics (2.2) and consider the situation when $\beta \gg 1$. We also make the following assumptions.

Assumption 1: $b \in C_b^3(\mathbb{R}^n, \mathbb{R}^n)$, $\sigma \in C_b^3(\mathbb{R}^n, M_{n\times n})$, $\sigma\sigma^T \in C_b^3(\mathbb{R}^n, S_n)$. $h \in C_b(\mathbb{R}^n)$.

Instead of (2.15), we first study the equation

$$\begin{aligned} &\frac{\partial \widetilde{U}}{\partial t} + \inf_{c\in\mathbb{R}^n} \Big\{ h + \frac{1}{2}|c|^2 + (b - \sigma c)\cdot \nabla \widetilde{U} \Big\} = 0 \\ &\widetilde{U}(T, z) = 0 \end{aligned} \tag{3.1}$$

which may be simpler to solve, and approximate $\hat{u}$, as given by (2.16), by

$$\hat{u}^0 = \sigma^T \nabla \widetilde{U}. \tag{3.2}$$

From (3.1) and the dynamic programming principle, we can represent $\widetilde{U}$ as the value function of a deterministic optimal control problem, i.e.

$$\widetilde{U}(t, z) = \inf_{u\in\mathcal{A}} \Big\{ \int_t^T h(\xi_s^u)\, ds + \frac{1}{2}\int_t^T |u_s|^2 ds \Big\} \tag{3.3}$$

under the deterministic dynamics

$$\begin{aligned} d\xi_s^u &= (b(\xi_s^u) - \sigma u_s)\, ds, \qquad t \le s \le T, \\ \xi_t^u &= z. \end{aligned} \tag{3.4}$$

This is actually the idea used in [34] to find the desired control force and readers are referred there for more algorithmic details and numerical examples. In this paper, we would like to explain the effectiveness of this approximation strategy with the following theorem (the proof is given in Appendix A)

Theorem 3.1. *Under Assumption 1 and Assumption 3 (see Appendix A), we have*

$$||\nabla U - \nabla \widetilde{U}||_{L^\infty} \le C\beta^{-1}. \tag{3.5}$$

where C is a constant that depends on functions b, σ, h and also t, T. From (2.16) and (3.2), it then follows that

$$||\hat{u} - \hat{u}^0||_{L^\infty} \le C\beta^{-1}. \tag{3.6}$$

Theorem 2.1 then implies the following bound on the relative standard deviation:

$$\rho_u(I) \le C\beta^{-\frac{1}{2}}. \tag{3.7}$$

3.2 Multiscale diffusions In this subsection, we consider the case when the state variable $z \in \mathbb{R}^n$ can be split into a slow variable $x \in \mathbb{R}^k$ and a fast variable $y \in \mathbb{R}^l$, i.e. $z = (x, y)$, $k + l = n$. Specifically, we assume that the dynamics (2.2) is of the form

$$\begin{aligned} dx_s &= \frac{1}{\epsilon} f_0(x, y)ds + f_1(x, y)ds + \alpha_1(x, y)dw_s^1 \\ dy_s &= \frac{1}{\epsilon^2} g_0(x, y)ds + \frac{1}{\epsilon} g_1(x, y)ds + \frac{1}{\epsilon}\alpha_2(x, y)dw_s^2 \end{aligned} \tag{3.8}$$

where $t \le s \le T$ and $\epsilon > 0$ describing the time scale separation. We will make the following assumption.

Assumption 2. $\epsilon \ll 1$, $\epsilon\beta \leq C$, where C is some constant.

We denote the solution of (2.15) as U^ϵ to emphasize its dependence on ϵ. The idea is to use asymptotic analysis method [5, 29] to derive an approximation of U^ϵ, thus obtaining a suboptimal control force close to (2.16).

Let $\phi^\epsilon(t,x,y) = \exp(-\beta U^\epsilon)$. From the dual relation (2.12), we conclude that ϕ^ϵ is just the expectation (2.1) we want to compute. Using the Feynman-Kac formula, we obtain

$$\begin{aligned}
&\frac{\partial \phi^\epsilon}{\partial t} + \mathcal{L}\phi^\epsilon - \beta h \phi^\epsilon = 0 \\
&\phi^\epsilon(T,x,y) = 1
\end{aligned} \tag{3.9}$$

where

$$\mathcal{L} = \frac{1}{\epsilon^2}\mathcal{L}_0 + \frac{1}{\epsilon}\mathcal{L}_1 + \mathcal{L}_2 \tag{3.10}$$

is the infinitesimal generator of (3.8), with

$$\begin{aligned}
\mathcal{L}_0 &= g_0 \cdot \nabla_y + \frac{\alpha_2\alpha_2^T}{2\beta} : \nabla_y^2 \\
\mathcal{L}_1 &= f_0 \cdot \nabla_x + g_1 \cdot \nabla_y \\
\mathcal{L}_2 &= f_1 \cdot \nabla_x + \frac{\alpha_1\alpha_1^T}{2\beta} : \nabla_x^2 .
\end{aligned} \tag{3.11}$$

To find an approximation of ϕ^ϵ, we consider the perturbative expansion $\phi^\epsilon = \phi_0 + \epsilon\phi_1 + \epsilon^2\phi_2 + \cdots$ of (3.9). The derivation of the lowest order expansion is standard (see, e.g., [29]), but we give it for the readers' convenience. Equating different powers of ϵ, we then obtain

$$\frac{\partial \phi_0}{\partial t} + \mathcal{L}_0\phi_2 + \mathcal{L}_1\phi_1 + \mathcal{L}_2\phi_0 - \beta h \phi_0 = 0, \tag{3.12}$$

$$\mathcal{L}_0\phi_0 = 0, \tag{3.13}$$

$$\mathcal{L}_0\phi_1 + \mathcal{L}_1\phi_0 = 0. \tag{3.14}$$

We make the standing assumption that for fixed x, the operator $\mathcal{L}_0$ is non-degenerate, i.e., there exists a unique smooth function $\rho_x(y) \geq 0$ satisfying $\mathcal{L}_0^*\rho_x = 0$, $\int \rho_x(y)dy = 1$, where $\mathcal{L}_0^*$ is the formal adjoint of the operator of $\mathcal{L}_0$ in L^2. This is equivalent to assuming that if we freeze x, the dynamics of y is ergodic with unique invariant distribution $\rho_x(y)$.

With this assumption, from (3.13), we conclude that $\phi_0 = \phi_0(t,x)$ is independent of y. Further suppose that $f_0(x,y)$ satisfies the "centering condition"

$$\int f_0(x,y)\rho_x(y)dy = 0, \quad \forall x \in \mathbb{R}^k$$

and that $\Phi(x,y)$ is a classical solution of the the "cell problem"

$$\mathcal{L}_0\Phi = -f_0, \qquad \int \Phi(x,y)\rho_x(y)dy = 0\,. \tag{3.15}$$

It can be readily shown that now $\phi_1 = \Phi \cdot \nabla_x\phi_0$. Furthermore, by multiplying $\rho_x(y)$ to both sides of (3.12) and integrate with respect to y, we obtain a closed equation for ϕ_0:

$$\frac{\partial \phi_0}{\partial t} + \widetilde{\mathcal{L}}\phi_0 - \beta\widetilde{h}\phi_0 = 0\,. \tag{3.16}$$

Here

$$
\begin{aligned}
&\widetilde{\mathcal{L}}_1 = \widetilde{f}(x)\cdot\nabla_x + \frac{\widetilde{\alpha}(x)\widetilde{\alpha}(x)^T}{2\beta} : \nabla_x^2, \quad \widetilde{h}(x) = \int h(x,y)\rho_x(y)\,dy,\\
&\widetilde{f}(x) = \int \Big(\nabla_x\Phi(x,y)f_0(x,y) + \nabla_y\Phi(x,y)g_1(x,y) + f_1(x,y)\Big)\rho_x(y)\,dy,\\
&\widetilde{\alpha}(x)\widetilde{\alpha}(x)^T = \int \Big[\beta\Big(\Phi(x,y)f_0(x,y)^T + f_0(x,y)\Phi(x,y)^T\Big) + \alpha_1(x,y)\alpha_1(x,y)^T\Big]\rho_x(y)\,dy.
\end{aligned}
\tag{3.17}
$$

which corresponds to the reduced dynamics

$$
\begin{aligned}
dx_s &= \widetilde{f}(x_s)ds + \beta^{-1/2}\widetilde{\alpha}(x_s)dw_s, \quad t\le s\le T\\
x_t &= x.
\end{aligned}
\tag{3.18}
$$

From the Feynman-Kac formula and (3.16), we know that

$$
\phi_0(t,x) = \mathbf{E}\Big(\exp\Big(-\beta\int_t^T \widetilde{h}(x_s)\,ds\Big)\ \Big|\ x_t = x\Big). \tag{3.19}
$$

Apply the expansion of ϕ^ϵ to $U^\epsilon = -\beta^{-1}\log\phi = U_0 + U_1\epsilon + o(\epsilon)$, it follows

$$
U^\epsilon = -\beta^{-1}\log(\phi_0 + \epsilon\phi_1 + o(\epsilon)) = -\beta^{-1}\log\phi_0 - \beta^{-1}\frac{\phi_1}{\phi_0}\epsilon + o(\epsilon). \tag{3.20}
$$

Therefore $U_0 = -\beta^{-1}\log\phi_0$. Combining (3.19) and the dual relation (2.12), we have

$$
U_0(t,x) = \inf_{u\in\mathcal{A}}\bar{\mathbf{E}}\Big\{\int_t^T \widetilde{h}(x_s^u)\,ds + \frac{1}{2}\int_t^T |u_s|^2 ds\Big\}, \tag{3.21}
$$

where x_s^u satisfies

$$
\begin{aligned}
dx_s^u &= \widetilde{f}(x_s^u)ds - \widetilde{\alpha}(x_s^u)u_s ds + \beta^{-1/2}\widetilde{\alpha}(x_s^u)dw_s, \quad t\le s\le T\\
x_t^u &= x
\end{aligned}
\tag{3.22}
$$

Applying the expansion to (2.16), we find the expansion of the optimal control force $\hat{u} = (\hat{u}_1, \hat{u}_2)$ with

$$
\begin{aligned}
\hat{u}_1 &= \alpha_1^T\nabla_x U_0 + O(\epsilon) = -\beta^{-1}\frac{\alpha_1^T\nabla_x\phi_0}{\phi_0} + O(\epsilon),\\
\hat{u}_2 &= \alpha_2^T\nabla_y U_1 + O(\epsilon) = -\beta^{-1}\frac{\alpha_2^T\nabla\phi_1}{\phi_0} + O(\epsilon).
\end{aligned}
\tag{3.23}
$$

Thus, in the case of multiscale diffusions, one possible strategy for finding the desired control force is to first compute U_0 from (3.21) or (3.19), which corresponds to a low-dimensional stochastic optimal control problem, and approximate $\hat{u}$ by

$$
\hat{u}^0 = (\alpha_1^T\nabla_x U_0, \alpha_2^T\nabla_y U_1). \tag{3.24}
$$

Here we would like to comment on the asymptotic formula (3.23), for which the rigorous proof is beyond the scope the current paper. Roughly speaking (3.23) is the asymptotic expansion of the gradient of the solution to the above homogenization problem. Adopting the proof of Theorem 3.1 as in the Appendix, using the maximum principle, will give a uniform upper bound with a constant C that depends on ϵ, which defeats the purpose of Theorem 2.1. A rigorous convergence proof of (3.23) requires the use of what is known in the literature as "corrector results" or "gradient estimates", involving technical PDE analysis. The interested readers is referred to [3, 4, 21, 27, 5] for further references. Here we simply assume that (3.23) holds with an error that is uniformly of order ϵ and summarize the above results as follows.

THEOREM 3.2. *Under mild assumptions (see, e.g., , [21]),*

$$||\nabla U^\epsilon - \nabla U_0||_{L^\infty} \leq C\epsilon, \tag{3.25}$$

where C is some constant independent of ϵ. From (3.23), (2.16), we have

$$||\hat{u} - \hat{u}^0||_{L^\infty} \leq C\epsilon \tag{3.26}$$

Then, using Assumption 2 and Theorem 2.1, it follows that

$$\rho_u(I) \leq C\epsilon^{\frac{1}{2}}. \tag{3.27}$$

REMARK 1. *The case when $f_0 \equiv g_1 \equiv 0$ in (3.8) corresponds to the "averaging" problem [29]. In this case, we can assume $\phi^\epsilon = \phi_0 + \epsilon\phi_1 + o(\epsilon)$ and proceed similarly as above. The equation of ϕ_0 is still given by (3.16), with*

$$\begin{aligned} \widetilde{f}(x) &= \int f_1(x,y)\rho_x(y)\,dy, \\ \widetilde{\alpha}(x)\widetilde{\alpha}(x)^T &= \int \alpha_1(x,y)\alpha_1(x,y)^T\rho_x(y)\,dy. \end{aligned} \tag{3.28}$$

(3.21) is unchanged and the optimal control force $\hat{u}$ can be approximated by

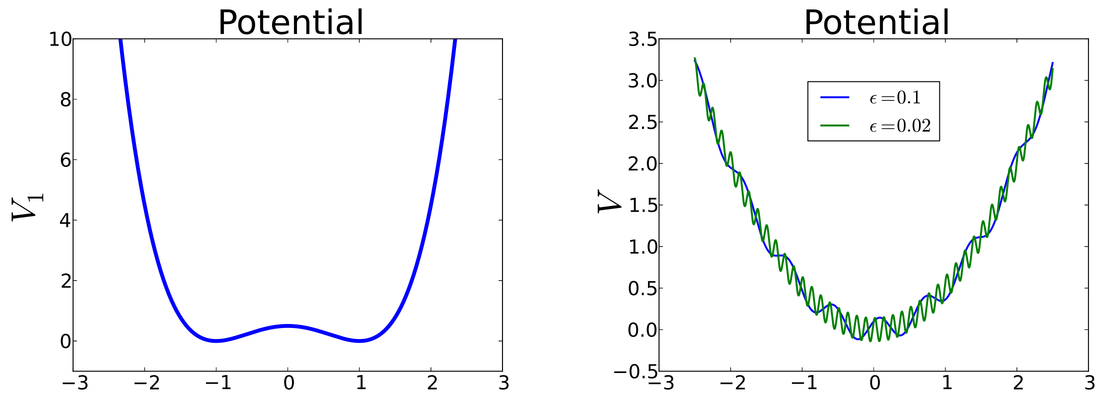
$$\hat{u}^0 = (\alpha_1^T \nabla_x U_0, 0). \tag{3.29}$$

A corrector estimate is not needed in this case, neither the solution of a cell problem, as the zeroth-order expansion already yields the desired uniform approximation (cf. [29, Sec. 20.4]).

3.3 Multiscale diffusions in zero temperature limit We focus on the "averaging" problem in Remark 1 and use the same notations there. We further assume that $\beta \gg 1$. More precisely we consider an approximation, in which we first send $\epsilon \to 0$ and then go to the zero-temperature limit, i.e. send $\beta \to \infty$. The order in which the limits are taken is motivated by the observation that in the zero-temperature limit the fast dynamics may not be ergodic anymore (cf. Assumption 2), which would exclude many systems of practical relevance and would require completely different mathematical techniques. In many applications, e.g. molecular dynamics or climate modelling in which simplified models are sought that preserve the large deviation properties of the system this order is quite natural [31, 1, 18].

Proceeding in this way and utilizing the dynamic programming principle and (3.21), we know that the leading term of U^ϵ satisfies

$$\begin{aligned} &\frac{\partial U_0}{\partial t} + \inf_{c\in\mathbb{R}^n}\Big\{\widetilde{h} + \frac{1}{2}|c|^2 + (\widetilde{f} - \widetilde{\alpha}c)\cdot\nabla U_0 + \frac{\widetilde{\alpha}\widetilde{\alpha}^T}{2\beta} : \nabla^2 U_0\Big\} = 0, \\ &U_0(T,x) = 0. \end{aligned} \tag{3.30}$$



(a) $V_1(x)$ (b) $V^\epsilon(x)$

Fig. 4.1: Potentials. (a) $V_1(x)$ in Example 4.1. (b) $V^\epsilon(x)$ in Example 4.2.

where $\widetilde{f}, \widetilde{\alpha}$ are given in (3.28). Using the idea of Subsection 3.1, we may further approximate (3.30) by considering

$$\begin{aligned} &\frac{\partial \widetilde{U}_0}{\partial t} + \inf_{c\in\mathbb{R}^n} \Big\{ \widetilde{h} + \frac{1}{2}|c|^2 + (\widetilde{f} - \widetilde{\alpha}c)\cdot \nabla \widetilde{U}_0 \Big\} = 0, \\ &\widetilde{U}_0(T,x) = 0. \end{aligned} \tag{3.31}$$

Correspondingly, we can interpret $\widetilde{U}_0$ as the following low-dimensional deterministic optimal control problem

$$\widetilde{U}_0(t,x) = \inf_{u\in\mathcal{A}} \Big\{ \int_t^T \widetilde{h}(\xi_s^u)\, ds + \frac{1}{2}\int_t^T |u_s|^2 ds \Big\} \tag{3.32}$$

with the deterministic dynamics

$$\begin{aligned} d\xi_s^u &= (\widetilde{f}(\xi_s^u) - \widetilde{\alpha} u_s)\, ds, \quad t \le s \le T, \\ \xi_t^u &= x. \end{aligned} \tag{3.33}$$

The optimal control force $\hat{u}$ can now be approximated by

$$\hat{u}^0 = (\alpha_1^T \nabla_x \widetilde{U}_0, 0). \tag{3.34}$$

In this situation, both parameters ϵ, β are involved in the approximation. Here we will only focus on proposing the above algorithm and the L^∞ convergence of the control force as in previous subsections will not be given. Numerical results using the above algorithm will be shown in the following section.

4 Numerical examples

We present two simple numerical examples that illustrate how suboptimal controls can be used to do importance sampling with nearly optimal efficiency.

4.1 Two-dimensional diffusion with stiff potential

Consider the potential $V = V_1 + V_2$, with

$$V_1(x) = \frac{1}{2}(x^2-1)^2, \quad V_2(x,y) = \frac{1}{2}(x-y)^2 \tag{4.1}$$

and the two-dimensional SDE

$$\begin{aligned}
dx_s &= -\frac{\partial V(x_s, y_s)}{\partial x} dt + \beta^{-1/2} dw_s^1 \\
dy_s &= -\frac{1}{\epsilon}\frac{\partial V(x_s, y_s)}{\partial y} dt + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}} dw_s^2 \\
x_0 &= -1, \quad y_0 = 0.
\end{aligned} \tag{4.2}$$

A similar dynamics has been studied in [23]. Our purpose is to calculate the expectation

$$I = \mathbf{E}\Big(\exp\Big(-\beta\int_0^T h(x_s, y_s) ds\Big) \;\Big|\; x_0 = -1,\, y_0 = 0\Big) \tag{4.3}$$

with $h(x, y) = (x-1)^2$.

Before going further, it is worthy to briefly illustrate the difficulties to compute (4.3) using standard Monte Carlo method, especially when β is large. On one hand, the exponential integrand in (4.3) is peaked where the trajectories spend large portion of time at the minimum point of h, i.e. $x = 1$. On the other hand, in order to get close to $x = 1$, trajectories, starting from $x_0 = -1$, need to cross the energy barrier of V_1 (Fig. 4.1(a)). The probability of these barrier-crossing trajectories is small when β is large. Combining these two facets, we can see that these rare trajectories play an important role when computing (4.3), and the standard Monte Carlo method, due to insufficient sampling of these events, will be inefficient (large relative error) in this case. In the following, we will apply the importance sampling method introduced in the previous section to this example and show some numerical results.

Suboptimal controls using averaged equations of motion Keeping the slow variable x in (4.2) fixed, the invariant measure of the fast variables y reads

$$\rho_x(y) \propto e^{-2\beta V_2} = e^{-\beta(x-y)^2}. \tag{4.4}$$

The reduced dynamics is simply

$$d\widetilde{x}_s = -V_1'(\widetilde{x}_s) ds + \beta^{-1/2} dw_s\,, \quad \widetilde{x}_0 = -1, \tag{4.5}$$

i.e. a one-dimensional diffusion in the double well potential.

We first consider the method in Subsection 3.2. As in (3.16), the equation for ϕ_0

$$\begin{aligned}
&\frac{\partial \phi_0}{\partial t} + \widetilde{\mathcal{L}}\phi_0 - \beta h \phi_0 = 0 \\
&\phi_0(T, x) = 1
\end{aligned} \tag{4.6}$$

has to be solved, where

$$\widetilde{\mathcal{L}} = -V_1'\frac{\partial}{\partial x} + \frac{1}{2\beta}\frac{\partial}{\partial^2 x}, \quad \widetilde{h}(x) = h(x) = (x-1)^2. \tag{4.7}$$

This is be done by first discretizing (4.6) with respect to time t,

$$\Big(\frac{1}{\Delta t} - \widetilde{\mathcal{L}}\Big)\phi_0^j = \Big(\frac{1}{\Delta t} - \beta h\Big)\phi_0^{j+1}\,, \quad j = 0, 1, \cdots, n-1\,, \tag{4.8}$$

where ϕ_0^j denotes the approximation of ϕ_0 at time $t_j = j\Delta t$, $j = 0, 1, \cdots, n$, $\Delta t = \frac{T}{n}$, and subsequent discretization in space using a finite-volume scheme [24]. The equations must then be solved backward in time which then yields the optimal control in accordance with (3.24):

$$\hat{u}^0 = (-\beta^{-1}\frac{\phi_0'}{\phi_0}, 0). \tag{4.9}$$

After applying the control force (4.9), the dynamics (4.2) becomes

$$\begin{aligned} dx_s^u &= -\frac{\partial V(x_s^u, y_s^u)}{\partial x}dt + \beta^{-1}\frac{\phi_0'(x_s^u)}{\phi_0(x_s^u)}dt + \beta^{-1/2}d\bar{w}_s^1 \\ dy_s^u &= -\frac{1}{\epsilon}\frac{\partial V(x_s^u, y_s^u)}{\partial y}dt + \beta^{-1/2}\frac{1}{\sqrt{\epsilon}}d\bar{w}_s^2 \\ x_0^u &= -1, \quad y_0^u = 0. \end{aligned} \tag{4.10}$$

Table 4.1: $T = 1.0$, $N = 10^4$. Monte Carlo method with importance sampling. Column I, I_N are the mean values computed with $N = 10^5, 10^4$ respectively. $\mathrm{Var}_s I, \rho_s(I)$ are the sample variance and the relative standard deviation defined in (2.25) (2.24). R_c is the ratio of the trajectories those have crossed the potential barrier.

β	ϵ	n_x	Δt	I	I_N	$\mathrm{Var}_s I$	$\rho_s(I)$	R_c
1.0	0.1	2000	1.0×10^{-7}	5.36×10^{-2}	5.38×10^{-2}	2.9×10^{-4}	0.32	6.9×10^{-1}
	0.01		1.0×10^{-8}	4.88×10^{-2}	4.88×10^{-2}	3.1×10^{-5}	0.11	6.7×10^{-1}
	0.001		1.0×10^{-8}	4.84×10^{-2}	4.83×10^{-2}	3.0×10^{-6}	0.04	6.8×10^{-1}
5.0	0.1	5000	1.0×10^{-7}	3.70×10^{-7}	3.65×10^{-7}	2.2×10^{-13}	1.26	7.6×10^{-1}
	0.01		1.0×10^{-7}	1.84×10^{-7}	1.85×10^{-7}	4.7×10^{-15}	0.37	7.3×10^{-1}
	0.001		1.0×10^{-8}	1.71×10^{-7}	1.70×10^{-7}	3.8×10^{-16}	0.11	7.3×10^{-1}
10.0	0.1	8000	1.0×10^{-7}	1.65×10^{-13}	1.70×10^{-13}	1.9×10^{-25}	2.56	8.6×10^{-1}
	0.01		5.0×10^{-8}	3.76×10^{-14}	3.80×10^{-14}	5.1×10^{-28}	0.59	8.4×10^{-1}
	0.001		1.0×10^{-8}	3.24×10^{-14}	3.24×10^{-14}	3.3×10^{-29}	0.18	8.4×10^{-1}

Results We will use the suboptimal dynamics (4.10) to compute the expectation (4.3) following (2.9). Table 4.1 shows the numerical results of the Monte Carlo method with the above importance sampling strategies, while Table 4.2 are the results using standard Monte Carlo method. We set $T = 1$ and $N = 10^4$ trajectories are generated in each case. To avoid the discretization bias, very small time step-sizes (Δt) are used; n_x denotes the mesh size used when solving (4.8). In Table 4.1, the column with label "I" presents results using large $N = 10^5$ and are considered as the "true mean value". Among the trajectories, we monitor the one for which $x_s > 0$ for certain $0 \leq s \leq T$ and R_c denotes the ratio of these trajectories, which can be considered as an indicator of the ratio of barrier-crossing events sampled. Now we make some comparisons of the results of Table 4.1 and Table 4.2. For $\beta = 1$, both methods give acceptable mean values, while the sample variance is smaller when importance sampling is used. In the case of standard Monte Carlo method, only about 0.2% of trajectories cross the barrier when $\beta = 5$, and this becomes worse when $\beta = 10$, i.e. no barrier-crossing trajectories are sampled. This is expected as these trajectories become more and more rare as β increases. In fact, the mean

Table 4.2: $T = 1.0$, $N = 10^4$. Standard Monte Carlo method. The column labels have the same meaning as in Table 4.1.

β	ϵ	Δt	I_N	$\mathrm{Var}_s I$	$\rho_s(I)$	R_c
	0.1	1.0×10^{-7}	5.31×10^{-2}	6.7×10^{-3}	1.54	2.6×10^{-1}
1.0	0.01	1.0×10^{-8}	4.76×10^{-2}	5.5×10^{-3}	1.56	2.6×10^{-1}
	0.001	1.0×10^{-8}	4.86×10^{-2}	5.7×10^{-3}	1.55	2.6×10^{-1}
	0.1	1.0×10^{-7}	2.62×10^{-7}	2.6×10^{-11}	19.46	2.4×10^{-3}
5.0	0.01	1.0×10^{-7}	2.90×10^{-7}	2.8×10^{-10}	57.70	1.7×10^{-3}
	0.001	1.0×10^{-8}	2.80×10^{-7}	2.1×10^{-10}	51.75	1.7×10^{-3}
	0.1	1.0×10^{-7}	8.14×10^{-15}	6.6×10^{-26}	31.56	0
10.0	0.01	5.0×10^{-8}	1.31×10^{-15}	2.7×10^{-27}	39.66	0
	0.001	1.0×10^{-8}	6.20×10^{-16}	6.0×10^{-28}	39.51	0

Table 4.3: $T = 1.0$, $N = 10^4$. Monte Carlo method with importance sampling by solving the deterministic optimal control problem (4.11)(4.12), with a coarse grid 2000×1000.

β	ϵ	Δt	I_N	$\mathrm{Var}_s I$	$\rho_s(I)$	R_c
	0.1	1.0×10^{-7}	5.41×10^{-2}	1.4×10^{-3}	0.69	8.3×10^{-1}
1.0	0.01	1.0×10^{-8}	4.84×10^{-2}	7.1×10^{-4}	0.55	8.0×10^{-1}
	0.001	1.0×10^{-8}	4.83×10^{-2}	6.5×10^{-4}	0.53	8.1×10^{-1}
	0.1	1.0×10^{-7}	3.25×10^{-7}	2.1×10^{-13}	1.41	8.5×10^{-1}
5.0	0.01	1.0×10^{-7}	1.92×10^{-7}	1.7×10^{-14}	0.68	8.3×10^{-1}
	0.001	1.0×10^{-8}	1.70×10^{-7}	8.0×10^{-15}	0.53	8.7×10^{-1}
	0.1	1.0×10^{-7}	1.54×10^{-13}	1.5×10^{-25}	2.51	9.0×10^{-1}
10.0	0.01	5.0×10^{-8}	3.82×10^{-14}	7.9×10^{-28}	0.74	9.0×10^{-1}
	0.001	1.0×10^{-8}	3.23×10^{-14}	1.6×10^{-28}	0.39	9.0×10^{-1}

values when $\beta = 10$ are not correct even in magnitude (comparing to the "true mean value"). On the contrary, in the case of importance sampling, the ratio of barrier-crossing trajectories are $70\% \sim 80\%$, thus are well sampled. The mean values remain stable when several runs are carried out, indicating that the estimate is indeed unbiased and converged.

We should mention that the importance sampling estimators perform well even when $\epsilon = 0.1$, i.e. when the time-scale separation of the dynamics is not at all in the limiting regime. In this case, the asymptotic of expectation value does not provide us the correct results (especially when $\beta = 5, 10$) and the Monte Carlo method is obvious needed. Comparing the relative standard deviation $\rho_s(I)$ in Table 4.1 with Table 4.2, we see the importance sampling estimator is still much more efficient than its standard counterpart. For each β, the controls used to apply the importance sampling methods are shown in Fig. 4.2. The blue regions in these figures shows that the control force will help the state to cross the energy barrier. We observe that, when β becomes larger ($\beta = 5, 10$), the solutions become closer to the limiting case ($\beta = +\infty$).

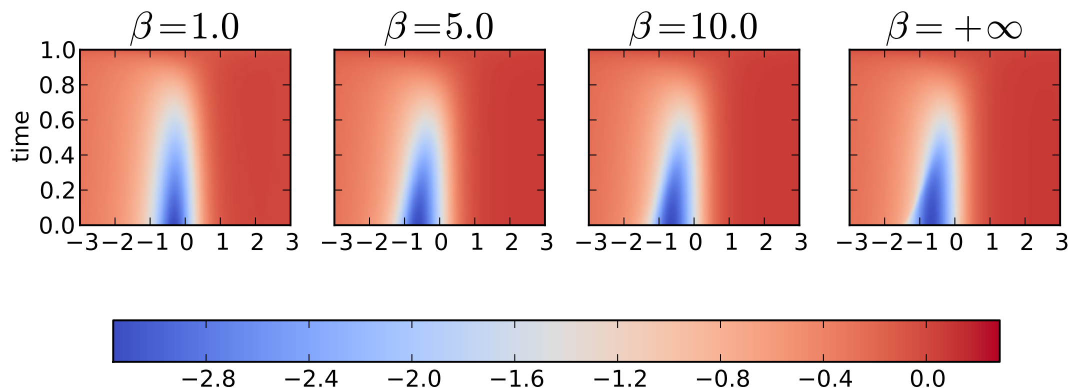


Fig. 4.2: The first three figures show the control forces computed from (4.9) for different β, using the method of Subsection 3.2. The rightmost figure ($\beta = +\infty$) shows the control force in (4.13), which is computed by solving the deterministic optimal control problem (4.11) (4.12).

Suboptimal controls using averaged equation and zero-temperature limit We also carried out the importance sampling method illustrated in Subsection 3.3. Combining (4.5) and (3.32) (3.33), we solve the deterministic optimal control problem

$$\widetilde{U}_0(t,x) = \inf_{u \in \mathcal{A}} \Big\{ \int_t^T h(\xi_s^u) ds + \frac{1}{2} \int_t^T |u_s|^2 ds \Big\} \tag{4.11}$$

with the deterministic dynamics

$$\begin{aligned} d\xi_s^u &= (-V_1'(\xi_s^u) - u_s) ds, \quad t \le s \le T, \\ \xi_t^u &= x. \end{aligned} \tag{4.12}$$

and use the control

$$\hat{u}^0 = (\nabla_x \widetilde{U}_0, 0) \tag{4.13}$$

to apply the importance sampling method. The optimal control problem can be solved using the optimization method of [34]. But instead of solving it "on the fly", here it is solved for each x, t on a coarse grid (2000×1000) using parallel computing before sampling, and then is used to compute the control force $\hat{u}^0$ (see the rightmost figure in Fig. 4.2). The results are shown in Table 4.3. We see that comparable results are obtained as in the Table 4.1, especially when $\beta = 5, 10$. We should mention that our numerical method for solve (4.11) is far from efficient, as the computational effort spent was more than it would have been for solving the the original sampling problem. However, we do not see this as a criterion for exclusion, because fast methods for solving deterministic control problems, such as (4.11), are available, e.g. direct methods [16] or methods based on the Pontryagin Maximum Principle [35].

4.2 Motion in a multiscale potential In this subsection, we consider a one-dimensional diffusion with periodic coefficients [13, 29]:

$$\begin{aligned} dx_s &= -\nabla V^\epsilon(x_s)ds + \beta^{-1/2}dw_s, \quad t \le s \le T, \\ x_t &= x\,, \end{aligned} \tag{4.14}$$

where $\epsilon > 0$, $V^\epsilon(x) = V(x) + p(x/\epsilon)$, $p(x)$ is a λ-periodic function. When $\epsilon \ll 1$, the potential function V^ϵ is the sum of a smooth potential function V and a highly oscillating perturbation $p(x/\epsilon)$, that is superimposed on the smooth potential landscape; see Figure 4.1(b).

Derivation of the corrector We start by condering the general situation and consider the expectation

$$\phi(t,x) = \mathbf{E}\Big(\exp\Big(-\beta\int_t^T h(x_s, x_s/\epsilon)\,ds\Big)\;\Big|\; x_t = x\Big).$$

Set $U = -\beta^{-1}\log\phi$, by dual relation (2.12), U is associated with the optimal control problem

$$U(t,x) = \inf_{u\in\mathcal{A}}\mathbf{E}\Big(\int_t^T \Big[h(x_s^u, x_s^u/\epsilon) + \frac{1}{2}|u_s|^2\Big]ds \;\Big|\; x_t^u = x\Big),$$

under dynamics

$$dx_s^u = -\frac{1}{\epsilon}\nabla p(x_s^u/\epsilon)ds - \nabla V(x_s^u)ds - u_s ds + \beta^{-1/2}dw_s. \tag{4.15}$$

To connect it with the homogenization problem studied in Subsection 3.2, we introduce the auxiliary variable $y = x/\epsilon$ and set $z = (x, y)$, by which (4.14) can be written as the redunant system of equations

$$\begin{aligned} dx_s &= -\frac{1}{\epsilon}\nabla p(y_s)ds - \nabla V(x_s)ds + \beta^{-1/2}dw_s, \\ dy_s &= -\frac{1}{\epsilon^2}\nabla p(y_s)ds - \frac{1}{\epsilon}\nabla V(x_s)ds + \frac{\beta^{-1/2}}{\epsilon}dw_s, \\ x_t &= x, \quad y_t = \frac{x}{\epsilon}. \end{aligned} \tag{4.16}$$

Notice that the same noise and the same control are applied to both variables.

Now consider the expectation

$$\widetilde{\phi}(t,x,y) = \mathbf{E}\Big(\exp\Big(-\beta\int_t^T h(x_s, y_s)\,ds \;\Big|\; x_t = x, y_t = y\Big).$$

and the associated optimal control problem

$$\widetilde{U}(t,x,y) = \inf_{u\in\mathcal{A}}\mathbf{E}\Big(\int_t^T \Big[h(x_s^u, y_s^u) + \frac{1}{2}|u_s|^2\Big]ds \;\Big|\; x_t^u = x, y_t^u = y\Big),$$

with dynamics

$$\begin{aligned} dx_s^u &= -\frac{1}{\epsilon}\nabla p(y_s^u)ds - \nabla V(x_s^u)ds - u_s ds + \beta^{-1/2}dw_s \\ dy_s^u &= -\frac{1}{\epsilon^2}\nabla p(y_s^u)ds - \frac{1}{\epsilon}\nabla V(x_s^u)ds - \frac{1}{\epsilon}u_s ds + \frac{\beta^{-1/2}}{\epsilon}dw_s. \end{aligned}$$

We can again take advantage of the dual relation $\widetilde{U}(t,x,y) = -\beta \log \widetilde{\phi}(t,x,y)$, with the identification $U(t,x) = \widetilde{U}(t,x,x/\epsilon)$. Applying the results of the previous section, we can compute the leading term of $\widetilde{U}(t,x,y)$ which satisfies the optimal control problem

$$U_0(t,x) = \inf_{u\in\mathcal{A}} \mathbf{E}\Big(\int_t^T \Big[\widetilde{h}(x_s^u) + \frac{1}{2}|u_s|^2\Big] ds \;\Big|\; x_t^u = x\Big)$$

under the dynamics

$$dx_s^u = -K\nabla V(x_s^u)ds - \sqrt{K}u_s ds + \sqrt{K}\beta^{-1/2} dw_s, \tag{4.17}$$

where

$$K = \int (I + \nabla_y \Phi(y))(I + \nabla_y \Phi(y))^T \rho(y) dy, \tag{4.18}$$

$\widetilde{h}$ is given in (3.17), $\rho(y)$ is the invariant measure of fast variable y and $\Phi(y)$ solves the cell equation. By the specific form of the infinitesimal generator (3.10)–(3.11), we find that $\rho(y) \propto \exp(-2\beta p(y))$. Moreover $\phi_1(t,x,y)$ is λ-periodic in y and, by (3.14), satisfies

$$\mathcal{L}_0 \phi_1 = -\mathcal{L}_1 \phi_0 = \nabla p \cdot \nabla_x \phi_0.$$

Solution of the cell problem In one dimension, the cell problem $\mathcal{L}_0\Phi(y) = p'$ can be solved analytically,

$$\Phi(y) = -y + \frac{\lambda}{\int_0^\lambda e^{2\beta p(z)} dz} \int_0^y e^{2\beta p(z)} dz, \tag{4.19}$$

which allows us to compute $\phi_1(t,x,y) = \Phi(y)\frac{\partial \phi_0}{\partial x}$. If we call

$$L = \int_0^\lambda e^{2\beta p(z)} dz, \quad \widetilde{L} = \int_0^\lambda e^{-2\beta p(z)} dz \tag{4.20}$$

then (4.18) gives $K = \lambda^2/(L\widetilde{L})$. From (2.16) and the expansion $\widetilde{\phi}(t,x,x/\epsilon) = \phi_0(t,x) + \epsilon\phi_1(t,x,x/\epsilon) + o(\epsilon)$, we have

$$\begin{aligned}
\hat{u} &= \frac{\partial}{\partial x} U(t,x) = -\beta^{-1} \frac{\partial_x \widetilde{\phi}(t,x,x/\epsilon)}{\widetilde{\phi}(t,x,x/\epsilon)} \\
&= -\beta^{-1} \frac{\partial_x \phi_0(t,x) + \partial_y \phi_1(t,x,x/\epsilon)}{\phi_0(t,x)} + O(\epsilon) \\
&= -\beta^{-1} \frac{\lambda e^{2\beta p(x/\epsilon)}}{\int_0^\lambda e^{2\beta p(z)} dz} \frac{\partial_x \phi_0}{\phi_0} + O(\epsilon) \\
&= \hat{u}^0(t,x) + O(\epsilon).
\end{aligned} \tag{4.21}$$

Results For the numerical test, we choose

$$\begin{aligned}
&p(x) = 0.1(\cos x + \sin x), \quad V(x) = x^2/2, \quad h = h(x) = \sin^2 x, \quad x_0 = -0.5 \\
&t = 0,\, T = 1.
\end{aligned}$$

The periodicity of $p(x)$ is $\lambda = 2\pi$. ϕ_0 is solved similarly as in the previous example, with

$$\widetilde{\mathcal{L}} = -K\nabla V \cdot \nabla + \frac{K}{2\beta}\Delta \tag{4.22}$$

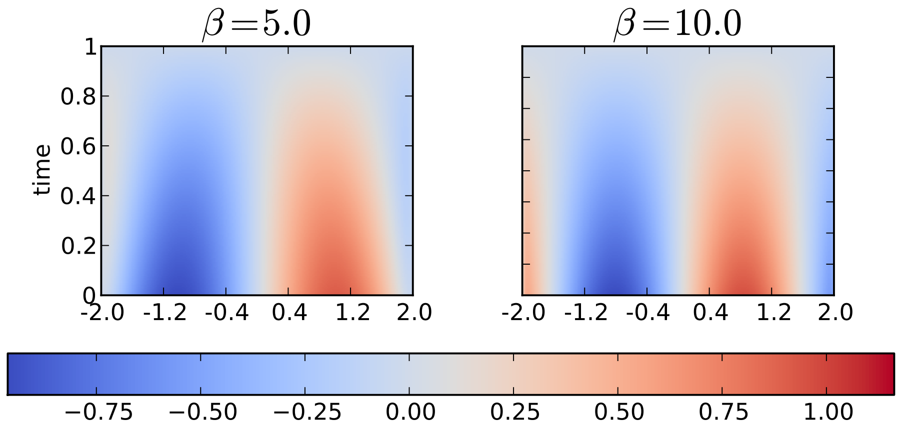


Fig. 4.3: the non-oscillatory part of the control forces $\hat{u}^0$ in (4.21) for $\beta = 5, 10$.

The control $\hat{u}^0$ is used to change the measure and the dynamics (4.14) becomes

$$\begin{aligned} dx_s^u &= -V'(x_s^u)ds - \frac{1}{\epsilon}p'(x_s^u/\epsilon)ds - \hat{u}^0 ds + \beta^{-1/2}d\widetilde{w}_s, \quad 0 \le s \le T, \\ x_0 &= -0.5. \end{aligned} \tag{4.23}$$

which is used to generate trajectories. Table 4.4 records the simulation results for $\beta = 5.0$ and 10.0 with different values of ϵ. As before, n_x is the mesh size used to compute ϕ_0. $N = 10^4$ trajectories are sampled in each case. The column with label "I" presents results using large $N = 10^5$ and are considered as the "true mean value". When generating trajectories, small Δt is chosen to avoid discretization bias. From the results, we can see that in each case, the variance is reduced when importance sampling is used. Moreover, for fixed β, as ϵ decreases, the sample variance of the importance sampling method also decreases, which is accordance with Theorem 3.2. Fig. 4.3 shows the non-oscillatory part of the control forces, which has an effect pushing the state x close to 0, which is where h attains its minimum.

5 Discussion and conclusions Importance sampling is a widely used variance reduction technique when designing efficient Monte Carlo estimators. To successfully apply this method, clever and careful selection of the change of measure is a key point. In the diffusion process setting, the change of measure is realized by adding a control force to the original system and the optimal control force is related to HJB equation. Our starting point is that, although it is not easy to find the optimal control force, it is possible to approximate it and the resulting estimators may also be efficient intuitively.

Based on the relation between the optimal control force and the HJB equation, we have studied approximations of the HJB equation for which three different situations—timescale separation, small temperature and the combination of the former two—where each approximation

Table 4.4: Monte Carlo method with and without importance sampling. $T = 1.0$, $N = 10^4$. $K = 0.407728$ when $\beta = 5.0$, while $K = 0.055302$ when $\beta = 10.0$. The column with label "I" shows the mean values sampled with large $N = 10^5$.

β	ϵ	n_x	Δt	I	IP			Standard		
					I_N	$\mathrm{Var}_s I$	$\rho_s(I)$	I_N	$\mathrm{Var}_s I$	$\rho_s(I)$
5.0	0.05	5000	1×10^{-7}	0.496	0.497	2.4×10^{-3}	9.8×10^{-2}	0.496	3.0×10^{-2}	0.04
	0.02	5000	1×10^{-7}	0.426	0.426	4.3×10^{-4}	4.9×10^{-2}	0.427	3.4×10^{-2}	0.43
	0.01	5000	1×10^{-7}	0.436	0.436	1.0×10^{-4}	2.3×10^{-2}	0.435	3.3×10^{-2}	0.42
	0.008	5000	5×10^{-8}	0.438	0.438	6.4×10^{-5}	1.8×10^{-2}	0.438	3.3×10^{-2}	0.41
	0.005	5000	5×10^{-8}	0.442	0.441	3.0×10^{-5}	1.2×10^{-2}	0.443	3.4×10^{-2}	0.42
	0.002	5000	5×10^{-8}	0.446	0.446	4.0×10^{-5}	1.4×10^{-2}	0.448	3.3×10^{-2}	0.41
10.0	0.05	2000	1×10^{-5}	0.198	0.198	1.3×10^{-3}	1.8×10^{-1}	0.198	5.8×10^{-3}	0.38
	0.02	2000	5×10^{-6}	0.104	0.104	3.0×10^{-4}	1.7×10^{-1}	0.104	4.4×10^{-3}	0.64
	0.01	2000	5×10^{-6}	0.109	0.109	5.8×10^{-5}	7.0×10^{-2}	0.109	2.3×10^{-3}	0.44
	0.008	8000	5×10^{-7}	0.111	0.111	3.0×10^{-5}	4.9×10^{-2}	0.111	2.0×10^{-3}	0.40
	0.006	8000	5×10^{-7}	0.114	0.114	1.7×10^{-5}	3.6×10^{-2}	0.114	1.6×10^{-3}	0.35
	0.004	8000	1×10^{-7}	0.117	0.117	7.7×10^{-6}	2.4×10^{-2}	0.117	1.8×10^{-3}	0.36
	0.002	8000	5×10^{-8}	0.120	0.120	4.1×10^{-6}	1.7×10^{-2}	0.120	1.7×10^{-3}	0.34

gave rise to an efficient Monte Carlo estimator. Under additional smoothness assumptions, the asymptotic optimality of these approximations has been proved. This demonstrates that, when temperature is small or timescales in the system are separated, the resulting approximate control force are suitable for importance sampling of the full system, in that they yield efficient Monte Carlo estimators with small variance and small relative error. The fact our results are based on uniform approximations in the L^∞ norm is partly due to type of sampling problem—deterministic initial condition and finite time horizon—and partly because it makes the stability analysis of relative error (or standard deviation) easier. By a more careful analysis of (2.20) and (2.21) it may be possible to weaken the assumptions. Weaker assumptions may be also possible when considering ergodic limits as in [10] or for problems that involve random stopping times [19]. We should mention that expected values of exponentials have been also studied in the context of the numerical solution of optimal control problems for diffusions (e.g. in [32]) and may be combined with the ideas and strategies to find approximations of the optimal force in the optimal control problem proposed in this article.

Even though we have studied only simple model systems, the result in this paper are promising. Future work should address adaptive importance sampling strategies for more general expectations such as the mean first passage times, or small probabilities of functionals of paths (e.g. committor probabilities [15]). In another interesting research direction is to study systems in which the time scale separation may not be explicit, but some idea of what the relevant coarse-grained variables are is available. This situation is typical for many problems arising in biology, climate modelling, physics, etc. and to generalize the ideas developed in this paper to such general *complex systems*, design fast and efficient importance samping strategies, is an important and interesting research topic.

Acknowledgements The authors acknowledge financial support by the DFG Research Center MATHEON.

Appendix A. Proof of Theorem 3.1.

In this section, we proof Theorem 3.1 and assume $t = 0$ without loss of generality. The method relies on the maximum principle for parabolic PDEs, and we use some analysis in [33] and [29]. We make the further assumption

Assumption 3: U, $\widetilde{U} \in C_b^{1,3}([0,T] \times \mathbb{R}^n)$ and the bounds are independent of β.

The following results from [33] will be useful.

THEOREM A.1. *If $f \in C^{1,2}([0,T) \times \mathbb{R}^n) \cup C_b([0,T] \times \mathbb{R}^n)$ and $c, g \in C_b([0,T))$ satisfy :*

$$\frac{\partial f}{\partial s} + \mathcal{L}_s f + c(s) f \geq -g(s), \quad 0 \leq s < T, \tag{A.1}$$

then

$$f(s,x) \leq ||f(T,\cdot)|| \exp\Big(\int_s^T c(u)\, du\Big) + \int_s^T g(t) \exp\Big(\int_s^t c(u)\, du\Big) dt \tag{A.2}$$

LEMMA A.2. *Let $a : \mathbb{R}^1 \to S_n$ be a function having two continuous derivatives. Suppose that*

$$\lambda_0 = \sup\Big\{\frac{|\langle\theta, a(x)''\theta\rangle|}{|\theta|^2} : x \in \mathbb{R}^1 \text{ and } \theta \in \mathbb{R}^n \backslash \{0\}\Big\} < \infty, \tag{A.3}$$

then for and symmetric $n \times n$-matrix u,

$$(tr(a(x)'u))^2 \leq 4n^2 \lambda_0 tr(ua(x)u), \quad x \in \mathbb{R}^n. \tag{A.4}$$

Appendix B. Proof of Theorem 3.1. The HJB equation (2.15) can be written as

$$\frac{\partial U}{\partial t} + b \cdot \nabla U + \frac{\sigma\sigma^T}{2\beta} : \nabla^2 U - \frac{|\sigma^T \nabla U|^2}{2} + h = 0 \tag{B.1}$$

With the notations in Subsection 3.1, we also have the equation for $\widetilde{U}$ as

$$\frac{\partial \widetilde{U}}{\partial t} + b \cdot \nabla \widetilde{U} - \frac{|\sigma^T \nabla \widetilde{U}|^2}{2} + h = 0 \tag{B.2}$$

Let $R = U - \widetilde{U}$, subtracting (B.2) from (B.1), we obtain

$$\frac{\partial R}{\partial t} + \Big(b - \frac{\sigma\sigma^T \nabla(\widetilde{U} + U)}{2}\Big) \cdot \nabla R + \frac{\sigma\sigma^T}{2\beta} : \nabla^2 R = -\frac{\sigma\sigma^T}{2\beta} : \nabla^2 \widetilde{U} \tag{B.3}$$

To simplify the notations, we denote

$$\widetilde{\mathcal{L}} = \Big(b - \frac{\sigma\sigma^T \nabla(\widetilde{U} + U)}{2}\Big) \cdot \nabla + \frac{\sigma\sigma^T}{2\beta} : \nabla^2 = \sum_{i=1}^n b^i(t,z) \frac{\partial}{\partial z_i} + \frac{1}{2} \sum_{i=1}^n a^{ij}(t,z) \frac{\partial^2}{\partial z_i \partial z_j}$$

$$F = -\frac{\sigma\sigma^T}{2} : \nabla^2 \widetilde{U}$$

This yields

$$\frac{\partial R}{\partial t} + \widetilde{\mathcal{L}} R - \beta^{-1} F = 0\,. \tag{B.4}$$

Taking the derivative with respect to z_i in (B.4), we obtain

$$\begin{aligned}
&\frac{\partial}{\partial t}\frac{\partial R}{\partial z_i} + \sum_{j=1}^{n} b^j \frac{\partial}{\partial z_j}\frac{\partial R}{\partial z_i} + \frac{1}{2}\sum_{j,k=1}^{n} a^{jk}\frac{\partial^2}{\partial z_j \partial z_k}\frac{\partial R}{\partial z_i} + \sum_{j=1}^{n}\frac{\partial b^j}{\partial z_i}\frac{\partial R}{\partial z_j} \\
&+ \frac{1}{2}\sum_{j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k} - \beta^{-1}\frac{\partial F}{\partial z_i} = 0.
\end{aligned}$$

Let $w = \sum\limits_{i=1}^{n}(\frac{\partial R}{\partial z_i})^2$. Multiplying the above equation by $2\frac{\partial R}{\partial z_i}$ and summing over i, we get

$$\begin{aligned}
&\frac{\partial w}{\partial t} + \widetilde{\mathcal{L}} w - \sum_{i,j,k=1}^{n} a^{jk}\frac{\partial^2 R}{\partial z_j \partial z_i}\frac{\partial^2 R}{\partial z_k \partial z_i} + \sum_{i,j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k}\frac{\partial R}{\partial z_i} + 2\sum_{i,j=1}^{n}\frac{\partial b^j}{\partial z_i}\frac{\partial R}{\partial z_j}\frac{\partial R}{\partial z_i} \\
&-2\beta^{-1}\sum_{i=1}^{n}\frac{\partial F}{\partial z_i}\frac{\partial R}{\partial z_i} = 0
\end{aligned} \tag{B.5}$$

By Lemma A.2, we know

$$\Big(\sum_{j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k}\Big)^2 = \Big(tr(\frac{\partial a}{\partial z_i}H)\Big)^2 \le 4n^2\lambda_i tr(HaH) \tag{B.6}$$

where $H = (H_{jk}) = \Big(\frac{\partial^2 R}{\partial z_j \partial z_k}\Big)$. Therefore we have

$$\Big(\sum_{i,j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k}\frac{\partial R}{\partial z_i}\Big)^2 \le \Big[\sum_{i=1}^{n}\Big(\sum_{j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k}\Big)^2\Big] w \le 4C_1\gamma w \tag{B.7}$$

with $\gamma = tr(HaH)$. It then follows that

$$-\sum_{i,j,k=1}^{n} a^{jk}\frac{\partial^2 R}{\partial z_j \partial z_i}\frac{\partial^2 R}{\partial z_k \partial z_i} + \sum_{i,j,k=1}^{n}\frac{\partial a^{jk}}{\partial z_i}\frac{\partial^2 R}{\partial z_j \partial z_k}\frac{\partial R}{\partial z_i} \le 2C_1^{1/2}\gamma^{1/2}w^{1/2} - \gamma \le C_1 w\,. \tag{B.8}$$

The other two terms in (B.5) can be estimated by

$$2\sum_{i,j=1}^{n}\frac{\partial b^j}{\partial z_i}\frac{\partial R}{\partial z_j}\frac{\partial R}{\partial z_i} - 2\beta^{-1}\sum_{i=1}^{n}\frac{\partial F}{\partial z_i}\frac{\partial R}{\partial z_i} \le C_2 w + C_3\beta^{-1}w^{1/2} \le C_4 w + \beta^{-2} \tag{B.9}$$

Combining (B.8) (B.9) together, It follows from (B.5) that

$$\frac{\partial w}{\partial t} + \widetilde{\mathcal{L}} w + (C_1 + C_4)w \ge -\beta^{-2} \tag{B.10}$$

Applying Theorem A.1 and notice $w(T,\cdot) = 0$, we obtain

$$||\hat{u} - \hat{u}^0||_{L^\infty} = ||\sigma^T(\nabla U - \nabla\widetilde{U})||_{L^\infty} \le C||\nabla U - \nabla\widetilde{U}||_{L^\infty} \le C\beta^{-1}. \tag{B.11}$$

REFERENCES

[1] E. V.-E. A.J. Majda, I. Timofeyev, *A mathematical framework for stochastic climate models*, Commun. Pure Appl. Math., 54 (2001), pp. 891–974.

[2] S. Asmussen and D. P. Kroese, *Improved algorithms for rare event simulation with heavy tails*, Adv. Appl. Prob., 38 (2006), pp. 545–558.

[3] M. Avellaneda and F.-H. Lin, *Compactness methods in the theory of homogenization*, Comm. Pure Appl. Math., 40 (1987), pp. 803–847.

[4] ———, *Compactness methods in the theory of homogenization II: equations in non-divergence form*, Comm. Pure Appl. Math., 42 (1989), pp. 139–172.

[5] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, Studies in mathematics and its applications, 1978.

[6] J. Blanchet and P. Glynn, *Efficient rare-event simulation for the maximum of heavy-tailed random walks*, Ann. Appl. Probab., 18 (2008), pp. 1351–1378.

[7] S. P. Brooks, *Markov chain Monte Carlo method and its application*, J. R. Stat. Soc. Series D (The Statistician), 47 (1998), pp. 69–100.

[8] A. Budhiraja and P. Dupuis, *A variational representation for positive functionals of infinite dimensional Brownian motion*, Probab. Math. Statist., 20 (2000), pp. 39–61.

[9] P. Dai Pra., L. Meneghini, and W. J. Runggaldier, *Connections between stochastic control and dynamic games*, Math. Control Signals Systems, 9 (1996), pp. 303–326.

[10] M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain markov process expectations for large time, I*, Comm. Pure Appl. Math., 28 (1975), pp. 1–47.

[11] A. Doucet, N. De Freitas, and N. Gordon, eds., *Sequential Monte Carlo methods in practice*, 2001.

[12] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, 195 (1987), pp. 216–222.

[13] P. Dupuis, K. Spiliopoulos, and H. Wang, *Importance sampling for multiscale diffusions*, Multiscale Model. Simul., 10 (2012), pp. 1–27.

[14] P. Dupuis and H. Wang, *Importance sampling, large deviations, and differential games*, Stochastics and Stochastic Rep., 76 (2004), pp. 481–508.

[15] W. E, W. Ren, and E. Vanden-Eijnden, *Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces and transition tubes*, Chem. Phys. Lett., 413 (2005), pp. 242–247.

[16] G. Fraser-Andrews, *A multiple-shooting technique for optimal control*, J. Optimiz. Theory App., 102 (1999), pp. 299–313.

[17] P. Glasserman, P. Heidelberger, and P. Shahabuddin, *Asymptotically optimal importance sampling and stratification for pricing path-dependent options*, Math. Finance, 9 (1999), pp. 117–152.

[18] C. Hartmann, *Model Reduction in Classical Molecular Dynamics*, PhD Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 2007.

[19] C. Hartmann and C. Schütte, *Efficient rare event simulation by optimal nonequilibrium forcing*, J. Stat. Mech. Theor. Exp., 2012 (2012), p. P11004.

[20] W. K. Hastings, *Monte Carlo sampling methods using markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109.

[21] A. Holmbom, *Homogenization of parabolic equations an alternative approach and some corrector-type results*, Appl. Math., 42 (1997), pp. 321–343.

[22] D. Kroese, T. Taimre, and Z. Botev, *Handbook of Monte Carlo Methods*, John Wiley and Sons, New York, 2011.

[23] J. C. Latorre, C. Hartmann, and C. Schütte, *Free energy computation by controlled Langevin dynamics*, Procedia Comput Sci., 1 (2010), pp. 1597 – 1606.

[24] J. C. Latorre, P. Metzner, C. Hartmann, and C. Schütte, *A structure-preserving numerical discretization of reversible diffusions*, Commun. Math. Sci., 9 (2010), pp. 1051–1072.

[25] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, 2nd ed., 2008.

[26] J. S. Liu and R. Chen, *Sequential Monte Carlo methods for dynamic systems*, J. Amer. Statist. Assoc., 93 (1998), pp. 1032–1044.

[27] P. Ming and P. Zhang, *Analysis of the heterogeneous multiscale method for parabolic homogenization problems*, Math. Comput., 76 (2007), pp. 153–177.

[28] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, 6th ed.,

2010.

[29] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.

[30] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *A direct approach to conformational dynamics based on Hybrid Monte Carlo*, J. Comput. Phys., 151 (1999), pp. 146 – 168.

[31] C. Schütte, J. Walter, C. Hartmann, and W. Huisinga, *An averaging principle for fast degrees of freedom exhibiting long-term correlations*, Multiscale Model. Simul., 2 (2004), pp. 501–526.

[32] C. Schütte, S. Winkelmann, and C. Hartmann, *Optimal control of molecular dynamics using Markov state models*.

[33] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Springer-Verlag, 1979.

[34] E. Vanden-Eijnden and J. Weare, *Rare event simulation of small noise diffusions*, Comm. Pure Appl. Math., 65 (2012), pp. 1770–1803.

[35] M. Weiser, *Interior point methods in function space*, SIAM J. Control Optim., 44 (2005), pp. 1766–1786.